



Analisi delle caratteristiche socio-demografiche dei clienti

In questo documento vengono analizzate le caratteristiche socio-demografiche dei clienti e le eventuali relazioni esistenti tra queste e le variabili che possono essere utilizzate per misurare la “bontà” di un cliente per la società.

Indicazioni tecniche


Alcune indicazioni per la lettura di questo documento:

- **breve spiegazione** se una parola è evidenziata in verde, questo simbolo  alla fine o all’inizio della riga in cui si trova la parola contiene una brevissima spiegazione del concetto statistico. Per leggere il contenuto del commento basta toccare il simbolo con il mouse.
- **[Commento o descrizione più dettagliata di una tecnica]** una frase evidenziata in giallo rimanda ad un documento che contiene alcune considerazioni o metodologiche o sulle modalità di utilizzo e di interpretazione degli strumenti statistici cui si fa riferimento. Per aprire il documento, selezionate la  nella barra del menu di Adobe:



Cliccando con il mouse sulla frase, si aprirà il documento.

NB: i richiami metodologici cui si rimanda nel testo (ordinati per argomento) sono tutti contenuti in questo documenti [Richiami metodologici]

- **Output non riportato nel documento** alcuni output non sono riportati nel documento per non appesantirlo eccessivamente; selezionando la  e cliccando con il mouse sulla frase, si aprirà un nuovo documento che contiene l’output.

Paragrafi e strumenti statistici

Per agevolare (ai fini didattici) la lettura del documento riportiamo per ogni paragrafo, le principali tecniche statistiche utilizzate.

1. Caratteristiche “personali” dei clienti

Analisi di distribuzioni univariate:

Distribuzioni delle frequenze; Istogramma

Studio dell’associazione:

Distribuzioni congiunte:

Tablelle a due vie e distribuzioni condizionate.

Tablelle a più vie: la costruzione di “profili”

2. Caratteristiche dei nuclei familiari dei clienti

Analisi di distribuzioni univariate:

Distribuzioni delle frequenze. Il caso dei dati mancanti

Studio dell'associazione:

Distribuzioni congiunte:

Tabelle a due vie e distribuzioni condizionate.

Tabelle a più vie: la costruzione di "profili"

Box plot affiancati

3. Caratteristiche socio demografiche e scelte di investimento

Studio della dipendenza

Analisi della dipendenza in media:

Box plot (analisi stratificata)

Funzione di regressione: Medie condizionate. Valutazione descrittiva della dipendenza in media: indice Eta.

Analisi inferenziale della funzione di regressione: Analisi della varianza a una via; test post hoc. Analisi della varianza nel caso di varianze non omogenee negli strati.

Test di normalità. Analisi non parametrica della varianza

Test T per la verifica dell'uguaglianza tra due medie. Test non parametrici per la verifica dell'uguaglianza tra due popolazioni.

Analisi della varianza a più vie

Analisi della relazione lineare:

diagrammi di dispersione, coefficiente di correlazione lineare. Verifica dell'ipotesi nulla di non correlazione.

Misure di concordanza. Verifica dell'ipotesi di assenza di concordanza/discordanza.

Misure di sintesi condizionate e valutazione dell'importanza relativa di profili

Studio della connessione:

Tabelle a doppia entrata

Misure di associazione e loro valutazione inferenziale.

Analisi delle caratteristiche socio-demografiche dei clienti

Oltre all'analisi diretta degli investimenti dei clienti della società, è utile per l'azienda considerare anche le caratteristiche socio-demografiche della propria clientela.

In primo luogo è infatti importante individuare le caratteristiche tipiche dei clienti, tracciando se possibile anche dei profili tipo e valutando se nel tempo si sono avute delle modifiche sostanziali.

In secondo luogo, è fondamentale cercare di comprendere se le modalità e le scelte di investimento dei clienti sono in qualche modo legate alle caratteristiche socio-demografiche. In questo caso, l'azienda può da una parte diversificare l'offerta a seconda del ritratto "socio-demografico" del cliente, e dall'altra decidere di spingere su nuovi prodotti su settori specifici della clientela.

Ecco quindi, di seguito, l'analisi delle caratteristiche socio-demografiche dei clienti (quali la loro professione, il loro stato civile, la composizione del loro nucleo familiare, ecc.) e la relazione di queste con le scelte di investimento.

1. Caratteristiche "personali" dei clienti

Nella Tabella 1.1. sono riportate le **distribuzioni delle frequenze** dei caratteri socio-demografici relativi ai **clienti della società** (Output di SPSS)¹.



Tab. 1.1. Distribuzioni di frequenza univariate dei caratteri

Sesso	Fr. assoluta	Percentuale
Femmina	336	33.5
Maschio	666	66.5

Professione	Fr. assoluta	Percentuale
Architetto	10	1.0
Avvocato	19	1.9
Commercialista	2	.2
Geometra	4	.4
Giornalista	2	.2
Ingegnere	12	1.2
Ragioniere	9	.9
altra lib.prof	137	13.7
SubTotale: Libero professionista	195	19.5
Imprenditore	42	4.2
Dirigente	25	2.5
SubTotale: imprend/dirig	67	6.7
Insegnante	26	2.6
Quadro	15	1.5
Impiegato	327	32.6
SubTotale: Impiegato	368	36.7
Agricoltore	5	.5
Artigiano	30	3.0
Operaio	31	3.1
SubTotale: Artigiano	66	6.6
Casalinga	49	4.9
Studente	15	1.5
Disoccupato	2	.2
SubTotale: Stud/disocc	17	1.7
Commerciante	77	7.7
Pensionato	163	16.3
Totale per tutte le tabelle	1002	100.0

¹ Output ottenuto con SPSS (Analyze; Descriptive Statistics, Frequencies)

Tab. 1.1. (continua) Distribuzioni di frequenza univariate dei caratteri

Titolo di Studio	Fr. assoluta	Percentuale
Licenza Elementare	26	2.6
Licenza Media	94	9.4
Diploma²	543	54.2
Laurea	339	33.8

Abitazione	Fr. assoluta	Percentuale
Affitto	84	8.4
Proprietà	874	87.2
Altro	2	.2
Presso Parenti	42	4.2

Vive o no in capoluogo di provincia	Fr. assoluta	Percentuale
Non vive in capoluogo di provincia	217	21.7
Vive in capoluogo di provincia	785	78.3

Totale per tutte le tabelle	1002	100.0
-----------------------------	------	-------

A partire dalle misure di sintesi delle distribuzioni (ovvero le **mode** delle distribuzioni e in un caso anche la **mediana**, messe in evidenza in Tab. 1.1), possiamo cominciare a tratteggiare le caratteristiche “tipiche” dei clienti della società. I clienti sono per la maggior parte di sesso maschile (66.5%), hanno un diploma di scuola superiore (54.2%) oppure la laurea (33.8%), sono impiegati (valore modale 32.6% o 36.7% se consideriamo anche quadri e insegnanti), liberi professionisti (19.5%) o pensionati (16.3%) e abitano per la maggior parte in casa di proprietà e in capoluogo di provincia. Nella Figura 1.1. viene rappresentata graficamente la distribuzione dell’età del cliente al momento dell’entrata per mezzo di un **istogramma** (classi di uguale ampiezza)³.

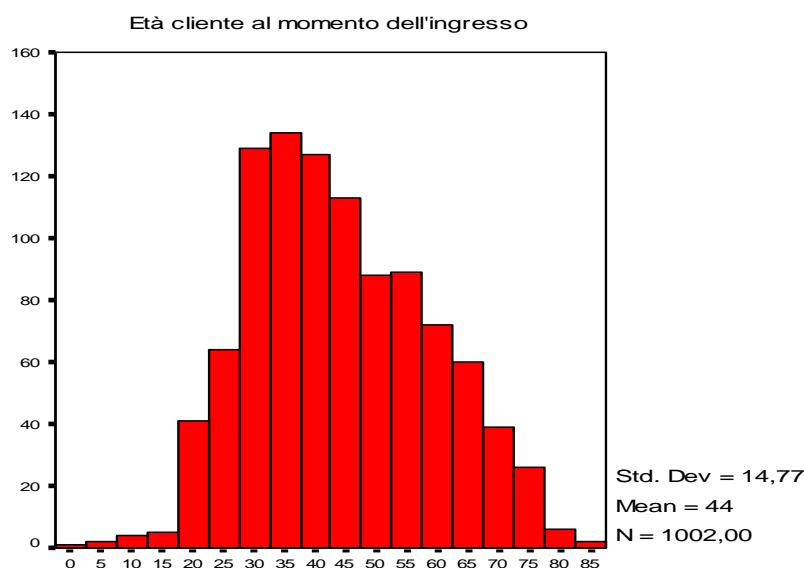


Fig. 1.1. Distribuzione dell’età del cliente al momento dell’ingresso

La maggior parte dei clienti entra in contatto con la società tra i 30 e i 45 anni, anche se si nota che la distribuzione presenta una coda destra “pesante”, cioè caratterizzata da una **densità di frequenza** che decresce piuttosto lentamente. Ciò indica una certa rilevanza anche del segmento di clienti che entrano in età superiore a 45 anni.

Ovviamente, le distribuzioni univariate si concentrano su un unico carattere; quindi da queste non è possibile costruire un “profilo” del cliente rappresentativo, in quanto nell’analisi univariata vengono perse le informazioni relative alle relazioni tra un carattere e gli altri (fino ad ora, abbiamo tracciato

² Questa modalità, oltre a rappresentare la moda della distribuzione, è anche la mediana.

³ Grafico ottenuto con SPSS (Graphs; Histogram).

tanti “profili tipici” quante sono le variabili osservate). Un passo successivo per comprendere meglio le caratteristiche dei clienti della società è perciò quello di considerare le distribuzioni congiunte delle diverse coppie di caratteri (Tabelle 1.2, 1.3 e 1.4) e le opportune distribuzioni subordinate.

[Analisi condizionata: le distribuzioni di frequenza]

Considerando ad esempio la **tabella a doppia entrata** di (Sesso, Età al momento dell'ingresso – in classi) notiamo che gli uomini tendono a diventare clienti relativamente prima delle donne: l'età media in ingresso è infatti circa 45 anni per le donne e inferiore ai 44 per gli uomini.

[Analisi condizionata: le misure di sintesi]

Tabella 1.2. Distribuzioni bivariate: Sesso * Età al momento dell'ingresso in classi. Crosstabulation⁴

Sesso		Età al momento dell'ingresso in classi						Total	Età media
		<25	25-35	35-45	45-55	55-65	>65		
F	Count	40	79	59	55	59	44	336	44,93
	% within Sesso	11,9%	23,5%	17,6%	16,4%	17,6%	13,1%	100,0%	
M	Count	43	167	186	128	87	55	666	43,60
	% within Sesso	6,5%	25,1%	27,9%	19,2%	13,1%	8,3%	100,0%	
Totale	Count	83	246	245	183	146	99	1002	44,05
	% within Sesso	8,3%	24,6%	24,5%	18,3%	14,6%	9,9%	100,0%	

Inoltre, con riferimento al titolo di studio (Tabella 1.3), ci sono più diplomate tra le donne, mentre la percentuale di laureati è più alta tra i clienti maschi.

Tabella 1.3. Distribuzioni bivariate: Sesso * Titolo di studio. Crosstabulation

Sesso		Titolo di Studio				Total
		Licenza Elementare	Licenza media	Diploma	Laurea	
F	Count	9	31	215	81	336
	% within Sesso	2,7%	9,2%	64,0%	24,1%	100,0%
M	Count	17	63	328	258	666
	% within Sesso	2,6%	9,5%	49,2%	38,7%	100,0%
Total	Count	26	94	543	339	1002
	% within Sesso	2,6%	9,4%	54,2%	33,8%	100,0%

Per quanto riguarda la professione (Tabella 1.4), la categoria lavorativa modale è quella di impiegato per entrambi i sessi (come a livello marginale). Tuttavia, mentre tra i clienti maschi le categorie più rappresentate a seguire sono quella dei liberi professionisti e dei pensionati, tra le donne spiccano anche le pensionate e le casalinghe.

Tabella 1.4. Distribuzioni bivariate: Sesso * Professione. Crosstabulation

Sesso		Professione (in classi)							Total	
		Libero profession	Imprendit./Dirigente	Impiegato	Commerc.	Artigiano	Casalinga	Stud./Disoccup.		Pensionato
F	Count	30	5	161	16	11	49	5	59	336
	% within Sesso	8,9%	1,5%	47,9%	4,8%	3,3%	14,6%	1,5%	17,6%	100,0%
M	Count	165	62	207	61	55		12	104	666
	% within Sesso	24,8%	9,3%	31,1%	9,2%	8,3%		1,8%	15,6%	100,0%
Total	Count	195	67	368	77	66	49	17	163	1002
	% within Sesso	19,5%	6,7%	36,7%	7,7%	6,6%	4,9%	1,7%	16,3%	100,0%

Potremmo ora procedere a considerare le distribuzioni congiunte delle diverse coppie di caratteri. Tuttavia, a questo punto, per costruire un profilo della clientela con riferimento alle variabili considerate (sesso, livello di istruzione, titolo di studio e fascia di età al momento dell'ingresso),

⁴ Tabella ottenuta con SPSS (Analyze; Descriptive Statistics, Crosstabs)

conviene considerare una tabella a 4 vie, in modo da individuare quali sono le caratteristiche dei clienti appartenenti ai gruppi più numerosi (Tabella 1.5).

Tabella 1.5. Tabella congiunta per Professione, Età all'ingresso, Sesso e Titolo di Studio.

Professione (in classi) * Età al momento dell'ingresso in classi * Sesso * Titolo di Studio Crosstabulation

Count				Età al momento dell'ingresso in classi						Total	
Titolo di Studio	Sesso	Professione (in classi)	Impiegato	<25	25-35	35-45	45-55	55-65	>65		
Licenza Elementare	F	Professione (in classi)	Impiegato				1			1	
			Artigiano				2			2	
			Casalinga			1	1			1	3
			Pensionato/a					1		2	3
	Total			1	4	1	3	9			
	M	Professione (in classi)	Commerciante						3		3
Artigiano				1	1		2			4	
Pensionato/a							6	4		10	
Total			1	1		11	4		17		
Licenza media	F	Professione (in classi)	Impiegato		4	2					6
			Commerciante	1	1						2
			Artigiano	1		3		1			5
			Casalinga				3	2			5
			Pensionato/a			1		5	7		13
	Total			2	5	6	3	8	7	31	
	M	Professione (in classi)	Imprenditore/Dirigente			1	2				3
			Impiegato		3	3	1		2		9
			Commerciante	1	4	1	2				8
			Artigiano	2	6	7	7	2			24
Studente/Disoccupato			1							1	
Total			4	14	13	17	10	3	5	63	
Diploma	F	Professione (in classi)	Libero professionista	1	5	2					8
			Imprenditore/Dirigente		1		2				3
			Impiegato	14	44	24	13	14	7		116
			Commerciante	2	3		6	2	1		14
			Artigiano	3		1					4
			Casalinga	3	5	6	7	13	5		39
			Studente/Disoccupato	3		1					4
	Pensionato/a	1			5	9	12		27		
	Total			27	58	34	33	38	25	215	
	M	Professione (in classi)	Libero professionista	2	6	16	4	5	1		34
			Imprenditore/Dirigente	3	5	4	8	1			21
			Impiegato	7	44	60	21	6	3		141
			Commerciante	7	9	10	12	4	4		46
			Artigiano	7	8	5	4	3			27
Studente/Disoccupato			6	3	1					10	
Pensionato/a				2	5	17	15	10		49	
Total			32	77	101	66	34	18	328		
Laurea	F	Professione (in classi)	Libero professionista	4	4	4	4	4	2	22	
			Imprenditore/Dirigente		1	1					2
			Impiegato	6	10	12	7	1	2		38
			Casalinga		1			1			2
			Studente/Disoccupato	1							1
			Pensionato/a			1	4	6	5		16
	Total			11	16	18	15	12	9	81	
	M	Professione (in classi)	Libero professionista	5	36	39	24	13	14		131
			Imprenditore/Dirigente	1	8	11	7	7	4		38
			Impiegato	1	29	16	5	4	2		57
			Commerciante		1	1	2				4
Studente/Disoccupato				1						1	
Total			7	75	71	45	32	28	258		

Se a questo punto consideriamo solo gli incroci con frequenza assoluta maggiore di 10 (Tabella 1.6), vediamo che dei 384 casi teoricamente possibili (ovvero $2 \cdot 8 \cdot 6 \cdot 4$ casi, corrispondenti alle possibili combinazioni di sesso, professione, gruppo di età e livello di studio) 26 “profili” rappresentano già il 50% del totale dei clienti della società. In altre parole, ci sono alcuni clienti che hanno caratteristiche comuni per sesso, professione, titolo di studio e età. Nello stesso tempo però non riusciamo a rilevare un unico gruppo di investitori particolarmente utile a rappresentare la clientela dell’azienda (anche il gruppo più numeroso, infatti, raccoglie solo il 6% della clientela).

Tabella 1.6. Profili più comuni tra i clienti della società (frequenze assolute >10).

Titolo di Studio	Sesso	Professione	Età (Ingresso)	Frequenze congiunte	
				Absolute	Relative
Diploma	M	Impiegato	35-45	60	0,0599
Diploma	F	Impiegato	25-35	44	0,0439
Diploma	M	Impiegato	25-35	44	0,0439
Laurea	M	Libero professionista	35-45	39	0,0389
Laurea	M	Libero professionista	25-35	36	0,0359
Laurea	M	Impiegato	25-35	29	0,0289
Diploma	F	Impiegato	35-45	24	0,0240
Laurea	M	Libero professionista	45-55	24	0,0240
Diploma	M	Impiegato	45-55	21	0,0210
Diploma	M	Pensionato/a	45-55	17	0,0170
Diploma	M	Libero professionista	35-45	16	0,0160
Laurea	M	Impiegato	35-45	16	0,0160
Diploma	M	Pensionato/a	55-65	15	0,0150
Diploma	F	Impiegato	<25	14	0,0140
Diploma	F	Impiegato	55-65	14	0,0140
Laurea	M	Libero professionista	>65	14	0,0140
Diploma	F	Impiegato	45-55	13	0,0130
Diploma	F	Casalinga	55-65	13	0,0130
Laurea	M	Libero professionista	55-65	13	0,0130
Diploma	F	Pensionato/a	>65	12	0,0120
Diploma	M	Commerciante	45-55	12	0,0120
Laurea	F	Impiegato	35-45	12	0,0120
Laurea	M	Imprenditore/Dirigente	35-45	11	0,0110
Diploma	M	Commerciante	35-45	10	0,0100
Diploma	M	Pensionato/a	>65	10	0,0100
Laurea	F	Impiegato	25-35	10	0,0100

Questi risultati confermano ovviamente i risultati dell’analisi univariata, anche se evidenziano alcuni incroci di modalità interessanti. Ad esempio, risulta piuttosto nutrito il gruppo delle casalinghe con diploma e entrate in contatto con la società tra i 55 e i 65 anni.

Notiamo comunque che tali considerazioni riguardano solo il *profilo tipico dei clienti della società* e non sono necessariamente i segmenti di clientela più promettenti. Possiamo sostanzialmente dire che sono questi i profili più comuni dei clienti della società.

Nella valutazione di questo risultato, è necessario utilizzare una certa cautela. Infatti, stiamo considerando la popolazione dei clienti della società, che è una sottopopolazione della popolazione italiana. Ovviamente, l’importanza relativa dei diversi profili potrebbe riflettere l’importanza relativa che questi profili hanno nella popolazione italiana (più precisamente, potrebbe anche darsi che gli impiegati diplomati di sesso maschile con età compresa tra i 35 e i 45 anni siano una delle categorie più frequenti anche nell’intera popolazione e non solo nella popolazione dei clienti dell’azienda). Quindi i profili più rilevanti non sono necessariamente quelli su cui la società “fa più presa”. **[Attrattività dell’azienda]**

2. Caratteristiche dei nuclei familiari dei clienti

Consideriamo ora quali sono le **caratteristiche dei nuclei familiari** della clientela della società. La Tabella 2.1 contiene informazioni sui nuclei familiari dei clienti. La maggior parte (74.8%) sono sposati. Per quanto riguarda il numero di figli e di figli a carico, il **valore modale** di entrambe le distribuzioni è 0 (nessun figlio). Tuttavia, nella distribuzione univariata non si tiene conto del fatto che non hanno sicuramente figli i single. Per valutare più nel dettaglio la composizione del nucleo familiare, nella Tabella 2.2 sono riportate le tabelle a doppia entrata relative alle distribuzioni congiunte dei caratteri stato civile, numero di figli e numero di figli a carico.

Tab. 2.1.¹ Distribuzioni di frequenza univariate dei caratteri (nucleo familiare)

Stato civile	Fr. assoluta	Percentuale
Convive	13	1.3
Separato	35	3.5
Single	162	16.2
Sposato	748	74.7
Vedovo	44	4.4
Total	1002	100.0

Figli	Fr. assoluta	Percentuale
0	767	76.5
1	83	8.3
2	126	12.6
3	22	2.2
4	4	.4
Total	1002	100.0

Figli a carico	Fr. assoluta	Percentuale
0	891	88.9
1	62	6.2
2	45	4.5
3	4	.4
Total	1002	100.0

Tab. 2.2. Distribuzioni congiunte di (Stato civile, Nro di figli) e (Stato civile, Nro di figli a carico)

Stato civile	Numero di figli					Total
	0	1	2	3	4	
Single	158	3	1			162
Coniugato/a	552	64	109	20	3	748
Vedovo/a	28	7	7	2		44
Separato/a	20	7	7		1	35
Convivente	9	2	2			13
	767	83	126	22	4	1002

Stato civile	Numero di figli a carico					Total
	0	1	2	3		
Single	161	1				162
Coniugato/a	654	50	40	4		748
Vedovo/a	41	2	1			44
Separato/a	26	7	2			35
Convivente	9	2	2			13
	891	62	45	4		1002

I clienti coniugati non hanno per la maggior parte figli, anche se si nota che, tra quelli che hanno figli, la maggior parte ha un nucleo familiare costituito da 4 unità (2 figli, valore modale e **mediano**). Un buon numero di clienti con figli non ha figli a carico, o ha alcuni figli autonomi.

Se analizziamo ora i dati relativi al coniuge, notiamo che in molti casi le informazioni a riguardo sono assenti (si veda in proposito la Tabella 2.3): sia per il titolo di studio che per la professione del partner le mancate risposte superano le 750, mentre il gruppo costituito da vedovi, separati e divorziati e single non raggiunge nemmeno le 250 unità. **[Mancate risposte]**

¹ Output ottenuto con SPSS (Analyze; Descriptive Statistics, Frequencies)

Tab. 2.3. Distribuzioni di frequenza univariate dei caratteri relativi al coniuge

	Titolo di studio del coniuge	Fr. assoluta	Percentuale	Percentuale sui casi validi
Valid	Licenza Elementare	13	1.3	5.6
	Licenza Media	61	6.1	26.1
	Diploma	106	10.6	45.3
	Laurea	54	5.4	23.1
	Total	234	23.4	100.0
Missing	0	768	76.6	
Total		1002	100.0	

	Professione del coniuge	Fr. assoluta	Percentuale	Percentuale sui casi validi
Valid	libero prof	15	1.5%	6.5%
	imprend/dirig	5	0.5%	2.2%
	Impiegato	71	7.1%	30.6%
	Commerciante	5	0.5%	2.2%
	Artigiano	17	1.7%	7.3%
	Casalinga	91	9.1%	39.2%
	Pensionato	28	2.8%	12.1%
	Total	232	23.2	100.0
Missing		770	76.8	
Total		1002	100.0	100.0

Non ha senso esaminare le risposte valide, poiché rappresentano una minoranza rispetto alle risposte attese. **[vedi la Tabella a doppia entrata di professioni clienti-professioni coniugi]**

A questo punto possiamo ancora una volta tracciare i profili più rappresentativi della clientela, con riferimento ora ai caratteri personali (professione e titolo di studio) e alla condizione familiare (stato civile, numero di figli). Nella Tabella 2.4 sono riportati perciò gli incroci di modalità più rilevanti (con frequenza congiunta maggiore di 0.1).

Tab. 2.4. Distribuzioni di frequenza congiunte dei caratteri socio-demografici

Titolo di Studio	Stato civile	Professione	Numero di figli	Frequenze	
				Absolute	Relative
Diploma	Coniugato/a	Impiegato	0	152	0,1517
Laurea	Coniugato/a	Libero professionista	0	107	0,1068
Diploma	Coniugato/a	Commerciante	0	51	0,0509
Diploma	Single	Impiegato	0	45	0,0449
Laurea	Coniugato/a	Impiegato	0	43	0,0429
Diploma	Coniugato/a	Pensionato/a	0	34	0,0339
Laurea	Single	Impiegato	0	29	0,0289
Diploma	Coniugato/a	Casalinga	0	28	0,0279
Diploma	Coniugato/a	Libero professionista	0	26	0,0259
Laurea	Coniugato/a	Imprenditore/Dirigente	0	26	0,0259
Diploma	Coniugato/a	Impiegato	2	22	0,0220
Laurea	Single	Libero professionista	0	22	0,0220
Diploma	Coniugato/a	Impiegato	1	20	0,0200
Laurea	Coniugato/a	Pensionato/a	0	15	0,0150
Diploma	Coniugato/a	Imprenditore/Dirigente	0	14	0,0140
Diploma	Single	Studente/Disoccupato	0	13	0,0130
Diploma	Coniugato/a	Pensionato/a	2	12	0,0120
Laurea	Coniugato/a	Libero professionista	2	11	0,0110
Licenza media	Coniugato/a	Artigiano	2	10	0,0100
Diploma	Coniugato/a	Artigiano	0	10	0,0100
Diploma	Coniugato/a	Pensionato/a	1	10	0,0100
Diploma	Vedovo/a	Pensionato/a	0	10	0,0100

In questo caso sono gli impiegati sposati ma ancora senza figli, con un titolo di studio pari al diploma a rappresentare la fetta più grossa di clienti della società. Questo gruppo, insieme a quello dei laureati liberi professionisti, sposati senza figli compone già un quarto dell'intera clientela.

[Collassare una tabella]

Ma nel tempo si sono modificate le caratteristiche socio demografiche dei nuovi ingressi?

Osservando i grafici qui di seguito, che presentano per ogni aspetto di interesse la distribuzione di frequenza condizionata all'anno di ingresso (in classi), numerosi sono i cambiamenti registrati, [Analisi condizionata: le distribuzioni di frequenza] e la connessione di ogni carattere con l'anno d'ingresso risulta sempre significativa. Le statistiche chi-quadrato [Misure di associazione] portano, infatti, sempre al rifiuto dell'ipotesi nulla di indipendenza, essendo caratterizzate da p-value molto bassi.

In particolare, si nota che nel tempo le clienti di sesso femminile (e le casalinghe) hanno aumentato la loro importanza relativa, così come clienti che abitano presso parenti e i single. Questo può essere dovuto al fatto che aumentano nel tempo i giovani clienti, che probabilmente tendono a vivere presso i propri genitori, come mostra anche la distribuzione congiunta dei due caratteri (Abitazione, Fascia di età).

Dal punto di vista dell'età al momento dell'ingresso, notiamo un aumento delle persone con età superiore ai 65 anni. I box plot dell'età (al momento dell'ingresso) condizionati all'anno di ingresso [Analisi stratificata: box plot affiancati] evidenziano un leggero aumento nel tempo dell'età mediana. Aumenta nel tempo il terzo quartile: aumenta quindi nel tempo l'età massima del 75% dei clienti.

Diminuiscono anche i laureati a favore di persone con livelli di istruzione inferiore (licenza media e licenza elementare – titolo di studio, questo, connesso alle persone di età più elevata).

Fig. 2.1. Caratteristiche della clientela e loro variazioni per anno di ingresso (continua nelle pagine seguenti).

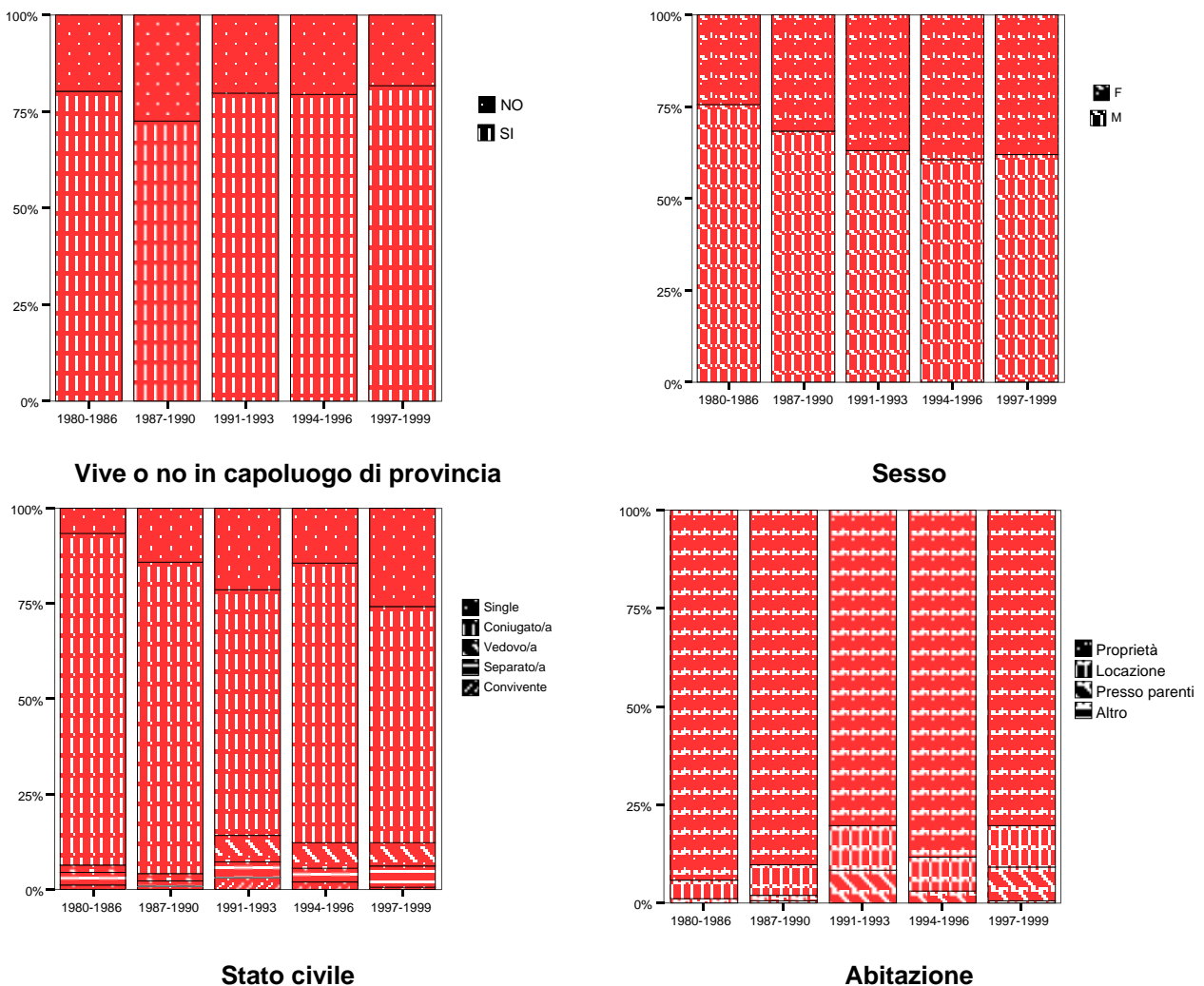
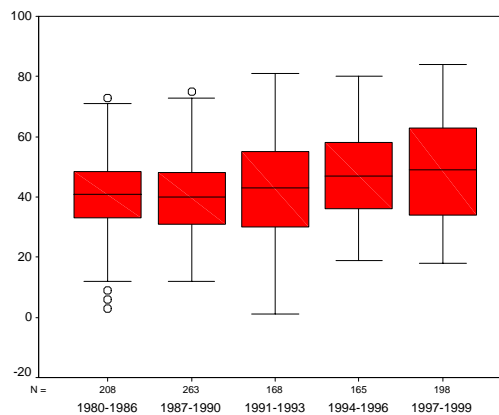
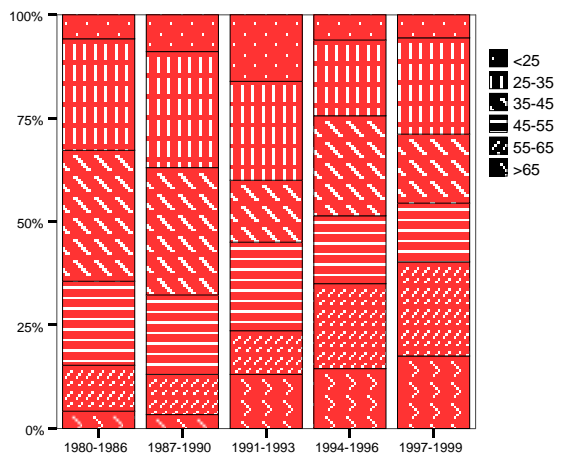


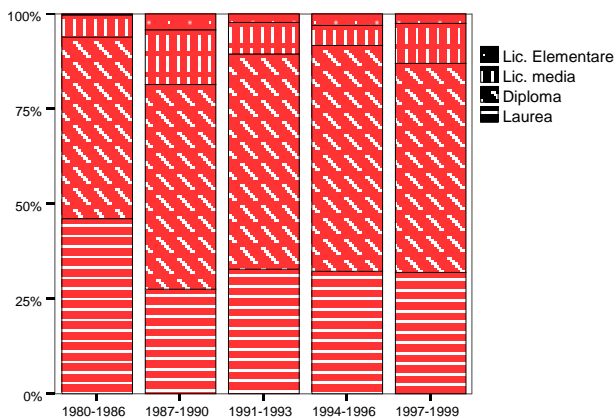
Fig. 2.1. Caratteristiche della clientela e loro variazioni per anno di ingresso (continua).

Abitazione * Età in classi Crosstabulation

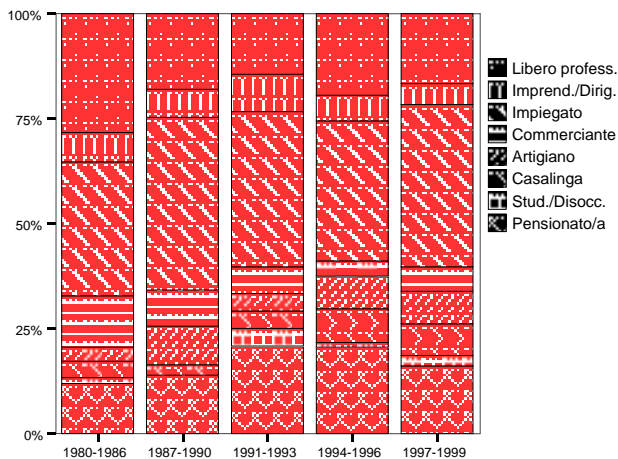
Abitazione	Età in classi						Total
	<25	25-35	35-45	45-55	55-65	>65	
Casa di proprietà	10	87	160	219	217	181	874
Casa in locazione	6	16	28	15	13	6	84
Abita presso parenti	9	20	11	1		1	42
Altro		1			1		2
Total	25	124	199	235	231	188	1002



Età al momento dell'ingresso (classi)



**Età al momento dell'ingresso (classi):
box plot affiancati**



Titolo di studio

Professione (in classi)

3. Caratteristiche socio-demografiche e scelte di investimento

Dopo aver analizzato quali sono le caratteristiche dei clienti della società è interessante per la rete di promotori comprendere se ci sono delle relazioni tra la “bontà” di un cliente e le sue caratteristiche socio-demografiche. Per misurare la bontà di un cliente, sarebbe interessante avere informazioni sull’ammontare totale di capitale versato nelle forme di investimento considerate. Tale analisi risentirebbe però dei diversi momenti di ingresso dei clienti: per i clienti di recente acquisizione non abbiamo informazioni in merito al capitale totale che potrebbero investire in un arco temporale lungo quanto quello che caratterizza i clienti di più lunga data. **[La censura delle osservazioni]**

Per questo, tenendo conto anche del fatto che i clienti sembrano essere maggiormente attivi nel 1° biennio di attività, consideriamo come variabile di bontà il **capitale totale investito nel 1° biennio di attività**, escludendo dall’analisi i clienti entrati nel 1999 che non hanno ancora maturato un biennio di attività.

3.1. Relazione tra capitale investito nel primo biennio e caratteristiche socio-demografiche.

Scopo di questa indagine è quindi quello di valutare se è possibile individuare alcune caratteristiche associabili con un’alta propensione a investire nei prodotti della società e in secondo luogo anche se esistono in generale dei “profili socio-demografici” che caratterizzano i buoni clienti.

Nei grafici che seguono (in Figura 3.1) sono riportate le **medie** del capitale investito nel primo biennio condizionate alle variabili socio-demografiche qualitative.

[Analisi stratificata: le misure di sintesi]

Siamo interessati a valutare sia se e come cambiano le medie a seconda delle caratteristiche socio-demografiche sia, contestualmente, ad individuare quali sono le caratteristiche più rilevanti in questo senso (cioè quelle condizionatamente alle quali le medie risultano più variabili). Per apprezzare più facilmente le differenze tra le medie utilizziamo una rappresentazione a barre (un diagramma in cui l’altezza della barra è proporzionale alla media condizionata). Per visualizzare in modo corretto l’importanza relativa delle diverse variabili, utilizziamo unità di misura approssimativamente uguali per tutte le variabili. **[Differenza tra medie. Rappresentazione grafica.]**

Fig. 3.1. Caratteristiche della clientela e media del capitale investito nel primo biennio (continua nelle pagine seguenti).

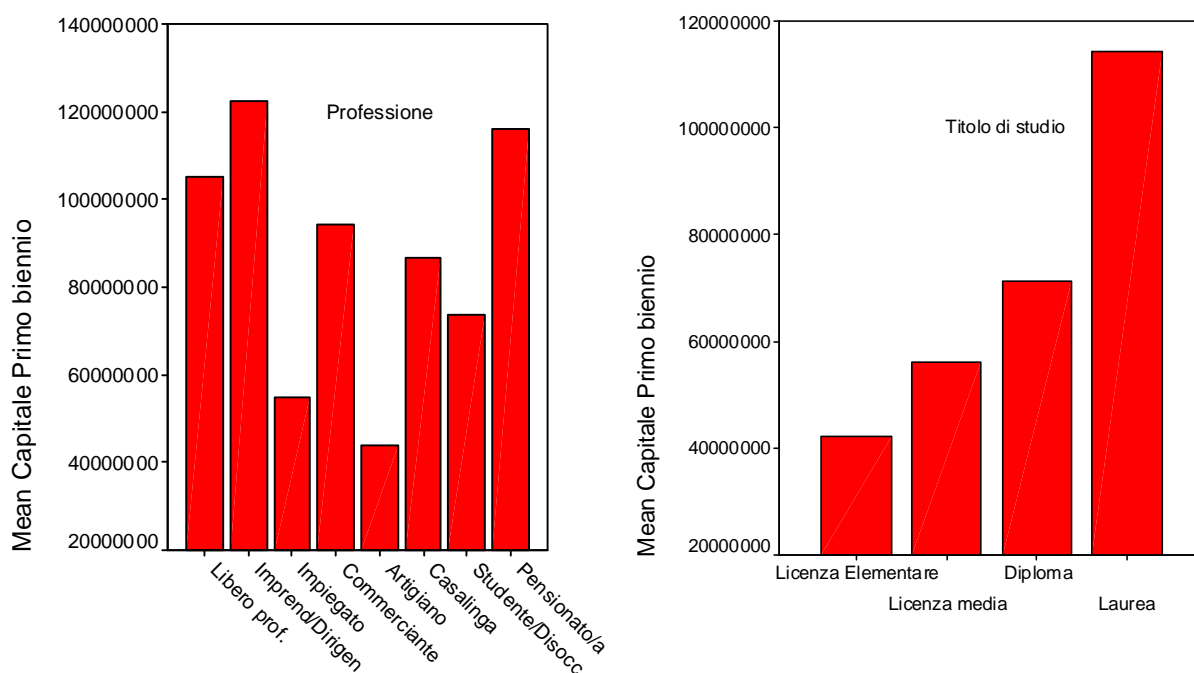
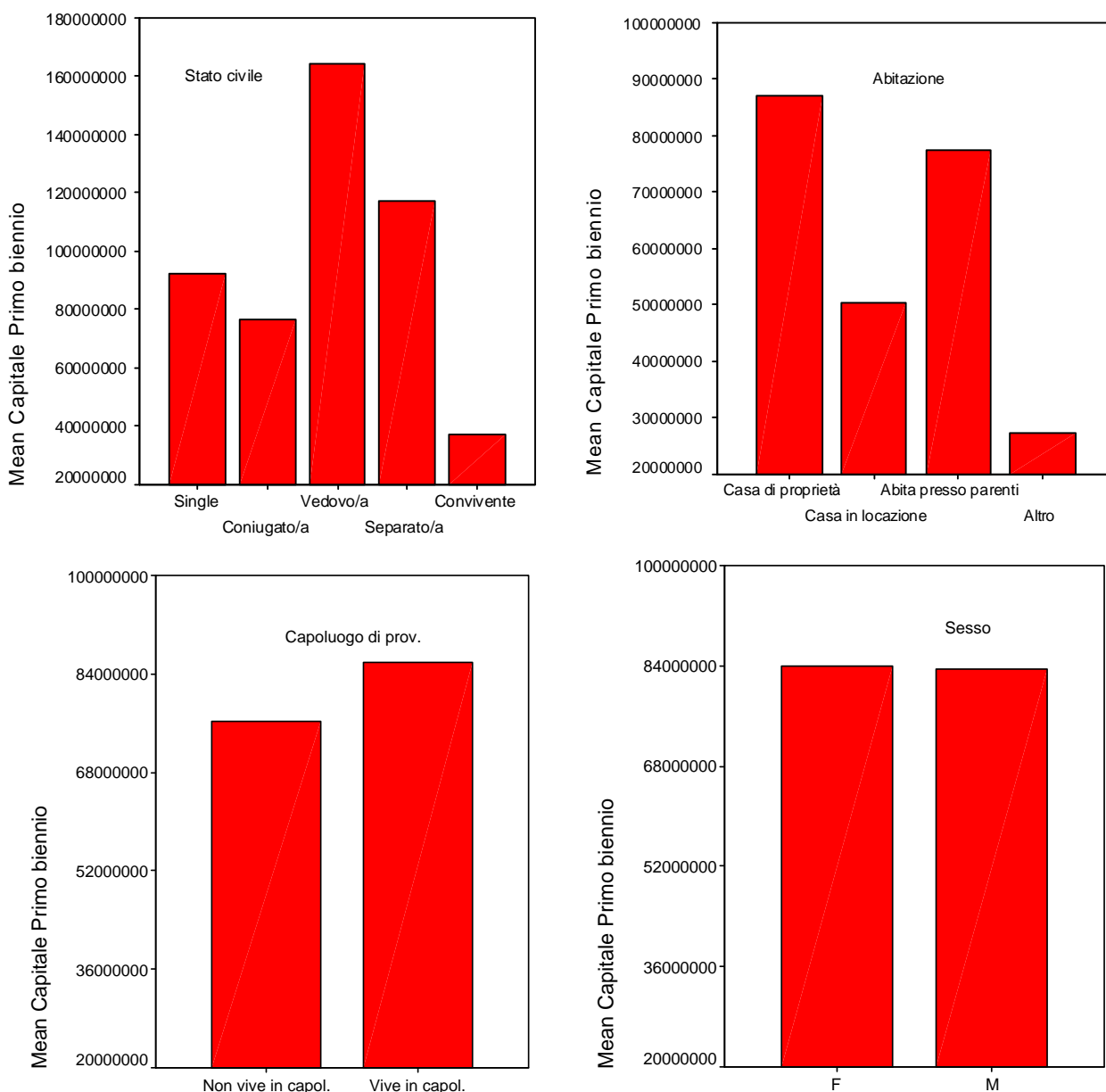


Fig. 3.1. Caratteristiche della clientela e media del capitale investito nel primo biennio (continua).



Dai grafici emergono differenze del capitale investito nel 1° biennio a seconda delle caratteristiche socio demografiche in esame. In particolare, sono gli imprenditori, i dirigenti e i liberi professionisti, insieme ai pensionati, ad investire mediamente in modo più massiccio nella società, mentre gli artigiani e gli impiegati sembrano essere più prudenti nelle loro scelte. Inoltre a titoli di studio più elevati corrispondono in media capitali investiti nel 1° biennio più alti.

Ancora, chi abita in casa di proprietà o con i genitori risulta più propenso a spendere in media in entrate cifre più alte. Per quanto riguarda lo stato civile, le situazioni di coppia precarie (vedovi e divorziati) si associano ad un impegno di maggiori capitali (i conviventi, invece, hanno quote mediamente più basse rispetto a tutte le altre categorie).

Non risultano particolarmente influenti invece il sesso e il fatto di vivere o meno in un capoluogo di provincia.

Per valutare se le considerazioni fatte in merito alle differenze tra le medie possono essere estese all'intera popolazione (e quindi in un certo senso generalizzate ad un cliente generico) oppure possono essere l'effetto di mere fluttuazioni campionarie, è opportuno valutare la significatività delle differenze osservate per mezzo di un'analisi della varianza [Analisi della varianza].

Cominciamo considerando le variabili esplicative a più livelli (Titolo di studio, professione, stato civile, abitazione¹). Per tali variabili, l'approccio standard (Tabella 3.1(a)) porta a rifiutare l'ipotesi nulla che le medie delle popolazioni siano tutte uguali: i p-value risultano infatti bassi). Per valutare quali sono le coppie di medie che risultano significativamente diverse tra loro e portano quindi al rifiuto dell'ipotesi nulla, riportiamo anche i risultati dei test post hoc [I test post-hoc standard].

Il test di Levene (Tab. 3.1(a)) porta a rifiutare l'ipotesi di omogeneità delle varianze per tutte le variabili esplicative. Riportiamo quindi anche i test adeguati nel caso di non omogeneità delle varianze [Caduta delle ipotesi dell'anova: varianze non omogenee]. Le variabili considerate risultano avere un impatto significativo sulle medie anche secondo tali test.

Tab. 3.1.² (a) Analisi della varianza del Capitale investito nel 1° biennio (in milioni)

Variabile esplicativa	Statistica F:	Significatività (p-value)
Titolo di studio	7.699	0.000
Professione	5.003	0.000
Stato civile	4.186	0.002

Test of Homogeneity of Variances

Variabile esplicativa	Levene Statistic	df1	df2	Sig.
Titolo di studio	12.005	3	947	.000
Professione	9.169	7	943	.000
Stato civile	9.782	4	946	.000

Robust Tests of Equality of Means

Variabile esplicativa	Test	Statistic	df1	df2	Sig.
Titolo di studio	Welch	10.691	3	146.954	.000
	Brown-Forsythe	11.627	3	574.021	.000
Professione	Welch	6.172	7	154.572	.000
	Brown-Forsythe	4.292	7	355.187	.000
Stato civile	Welch	7.153	4	71.898	.000
	Brown-Forsythe	3.005	4	115.511	.021

Multiple Comparisons

Gruppi confrontati		Mean Difference (I-J)	P-value			
			LSD	Bonferroni	Tamhane	Games-Howell
Licenza Elementare	Diploma	-29.1768	.322	1.000	.019	.016
Licenza Elementare	Laurea	-72.1935	.016	.095	.000	.000
Licenza media	Laurea	-58.3160	.001	.005	.001	.001
Diploma	Laurea	-43.0167	.000	.000	.002	.002
Libero professionista	Impiegato	50.5059	.000	.004	.009	.007
Libero professionista	Artigiano	61.2776	.004	.113	.001	.001
Imprenditore/Dirigente	Impiegato	67.6798	.001	.018	.395	.244
Imprenditore/Dirigente	Artigiano	78.4515	.002	.068	.181	.119
Impiegato	Commerciante	-39.4269	.033	.932	.903	.640
Impiegato	Pensionato/a	-61.2704	.000	.000	.002	.002
Commerciante	Artigiano	50.1987	.044	1.000	.578	.361
Artigiano	Pensionato/a	-72.0422	.001	.027	.000	.000
Single	Vedovo/a	26.46352	.007	.067	.654	.460
Single	Convivente	42.53048	.193	1.000	.015	.013
Coniugato/a	Vedovo/a	24.17198	.000	.003	.315	.217
Coniugato/a	Convivente	41.14376	.334	1.000	.003	.002
Vedovo/a	Convivente	47.08555	.007	.070	.034	.027

¹ I risultati per la variabile esplicativa "Abitazione" non sono riportati in quanto l'effetto sulla variabile dipendente non è significativo.

² Risultati ottenuti utilizzando SPSS (Analyze, Compare means, One-way Anova, specificando nelle options Homogeneity of variances test, test di Welch e test di Brown-Forsythe).

Il test LSD individua come significative molte più coppie di medie rispetto agli altri test. Per quanto riguarda il titolo di studio, tutti i test confermano che la media del capitale investito dai laureati è significativamente diversa dalle altre (anche se per la differenza con la media di coloro che hanno la licenza elementare il test di Bonferroni non conferma la significatività della differenza).

Con riferimento alla professione, possiamo affermare che tutti i test portano a concludere che liberi professionisti e pensionati investono mediamente di più rispetto ad artigiani e impiegati. Anche la categoria dei commercianti sembra distinguersi dalle altre, nonostante tale indicazione non sia confermata da tutti i test. Per quanto riguarda lo stato civile, la categoria che si differenzia significativamente dalle altre è quella dei vedovi.

L'ANOVA suggerisce quindi che alcune differenze sono importanti e altre no. Ad esempio, nonostante si osservino medie più elevate dei single rispetto ai coniugati (a livello *descrittivo*), sarebbe un errore per la società "puntare" sui single e trascurare i coniugati, in quanto la differenza osservata nel campione non necessariamente si riproduce nella popolazione. E' invece corretto considerare i vedovi come categoria di punta in quanto la differenza osservata con le altre categorie è significativa e quindi si può estendere all'intera popolazione (cioè il risultato descrittivo non può essere attribuito alla fluttuazione campionaria ma piuttosto ad una effettiva differenza tra le medie nella popolazione).

Di seguito sono riportati i risultati dei test di **Kolmogorov-Smirnov** e di **Shapiro Wilk** per la verifica dell'ipotesi di normalità, alle basi dell'ANOVA, che viene evidentemente rifiutata³.

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Capitale investito nel 1° biennio (in milioni)	.286	951	.000	.506	951	.000

Sebbene l'ANOVA sia abbastanza robusta a deviazioni dall'ipotesi di normalità, se la distribuzione è asimmetrica e l'asimmetria accentuata i risultati ottenuti possono non essere attendibili e si deve ricorrere a procedure non parametriche. Essendo la distribuzione del carattere fortemente asimmetrica (si veda il par. 1 del documento *Analisi delle scelte di investimento*) verificiamo i risultati ottenuti con i test non parametrici.⁴ **[Analisi della varianza non parametrica]**. Di seguito, i risultati ottenuti con il test di Kruskal-Wallis e con il test delle mediane⁵.

Tab. 3.1(b) ANOVA non parametrica del Capitale investito nel 1° biennio (in milioni)

Variabile esplicativa	Test di Kruskal Wallis		Median Test	
	Statistica test	Asymp. Sig.	Statistica test	Asymp. Sig.
Titolo di Studio	27.46	.000	15.46	.001
Professione (in classi)	35.24	.000	29.10	.000
Stato civile	6.290	.178	3.138	.535

Analizzando tali risultati troviamo conferma della significatività del titolo di studio e della professione. Risulta invece non significativo lo Stato civile (anche Abitazione si conferma non significativo). Dovremmo a questo punto procedere a verificare quali sono le popolazioni che risultano significativamente diverse tra loro utilizzando test non parametrici per l'uguaglianza tra due popolazioni **[Test non parametrici per verificare l'uguaglianza tra due popolazioni]**. Non è possibile, come nei casi visti precedentemente, procedere a verificare tali ipotesi tenendo sotto controllo la probabilità di commettere un errore di prima specie a livello di esperimento (cioè con riferimento all'intero insieme di differenze sottoposte a verifica) ma solo tenendo sotto controllo

³ Risultato ottenuto con SPSS (Analyze, Non parametric tests, 1 sample).

⁴ Ovviamente la verifica delle assunzioni alla base dell'ANOVA dovrebbe precedere l'applicazione dell'approccio standard. In questo caso, ad esempio, avremmo dovuto ricorrere immediatamente alle tecniche non parametriche. Seguiamo un percorso diverso ai fini didattici.

⁵ Risultato ottenuto con SPSS (Analyze, Non parametric tests, K independent samples).

l'errore a livello di singolo confronto (sostanzialmente la versione non parametrica del test LSD per il caso standard). Utilizzando il test di Mann-Whitney troviamo conferma che il capitale investito nel 1° biennio dai laureati ha una distribuzione significativamente diversa da quello delle altre categorie (definite in base al titolo di studio). Anche per quanto riguarda la professione, si conferma la differenza tra liberi professionisti, imprenditori e pensionati dalle categorie lavorative con capitale medio più basso.

Per quanto riguarda le due variabili dicotomiche, Vive o no in capoluogo di provincia e Sesso del cliente, possiamo valutare se il capitale medio investito cambia nei due sottocampioni indotti dalle loro (due) modalità con un test T per l'uguaglianza tra medie (campioni indipendenti). I risultati sono riportati nella Tabella 3.1.(c). **[Test T per l'uguaglianza tra medie]**

Tab. 3.1.(c)⁶ Test T sull'uguaglianza delle medie del Capitale investito nel 1° biennio (in milioni)

Var esplicativa: Vive o no in capoluogo di provincia	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	.093	.760	-.835	949	.404	-9.6182	11.51397
Equal variances not assumed			-.726	288.953	.468	-9.6182	13.24639
Var esplicativa: Sesso	Levene's Test for Equality of Variances		t-test for Equality of Means				
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Equal variances assumed	.084	.773	-.040	949	.968	-.4029	10.16278
Equal variances not assumed			-.042	744.851	.967	-.4029	9.61413

In questo caso l'ipotesi di uguaglianza tra le varianze nei due gruppi viene accettata per entrambe le variabili e il test T standard porta a rifiutare l'ipotesi che le due variabili siano significative. Riportiamo anche i risultati ottenuti con i test non parametrici (test di Mann-Whitney e test di Kolmogorov-Smirnov). **[Test non parametrici per verificare l'uguaglianza tra due popolazioni]**.

Riportiamo nella Tabella 3.1(d) solo i risultati relativi alla seconda variabile, l'unica per la quale si osserva un risultato diverso rispetto a quello ottenuto con il test T. Considerando i ranghi al posto delle medie si conclude a favore della significatività della variabile. Ciò potrebbe essere dovuto al fatto che l'esigua differenza tra le medie sia dovuta alla presenza di valori anomali evidenti nei box plot affiancati (i valori anomali possono ovviamente accentuare come offuscare le differenze tra le diverse popolazioni). **[vedi i box plot]**

Tab. 3.1(d)⁷ Test non parametrico sulla differenza tra le medie del Capitale investito nel 1° biennio (var esplicativa: Vive o no in capoluogo di provincia)

Test di Mann-Whitney		Test di Kolmogorov-Smirnov		
Mann-Whitney U	68247	Most Extreme Differences	Absolute	.112
Wilcoxon W	91038		Positive	.023
			Negative	-.112
Z	-2.931	Kolmogorov-Smirnov Z		1.443
Asymp. Sig. (2-tailed)	.003	Asymp. Sig. (2-tailed)		.031

Passiamo ora a considerare la dipendenza del capitale investito dai caratteri quantitativi (numero di figli, età al momento dell'ingresso). La natura delle coppie di caratteri considerati (quantitativo quello dipendente, quantitativo con poche modalità quello esplicativo) consente di analizzare la dipendenza in due modi: facendo riferimento al loro eventuale legame lineare o indagando sull'esistenza di una relazione non necessariamente lineare o analitica.

Studiare la relazione *lineare* che lega il capitale al numero di figli, ad esempio, significa chiedersi

⁶ Risultato ottenuto con SPSS (Analyze, Compare means, Independent-samples T test)

⁷ Risultato ottenuto con SPSS (Analyze, Non parametric tests, 2 independent samples).

se all'aumentare del numero di figli aumenta o diminuisce il capitale investito nel 1° biennio. Studiare la *dipendenza in media* significa invece chiedersi se al variare del numero di figli si osserva una differenza nel capitale medio investito (e questa relazione non è necessariamente lineare o diretta).

Cominciamo valutando se la relazione che eventualmente lega le coppie di caratteri considerate è o no lineare. Nella Tabella 3.2 riportiamo i **coefficienti di correlazione lineare** tra il capitale investito nel 1° biennio e le variabili quantitative. Data la forte asimmetria della variabile dipendente (il capitale investito), consideriamo oltre al coefficiente di correlazione lineare (influenzato, come noto, dalla presenza di valori anomali) anche due misure robuste di concordanza: il **coefficiente tau di Kendall** e il **coefficiente di Spearman**. *[Non robustezza del coefficiente di correlazione lineare]*

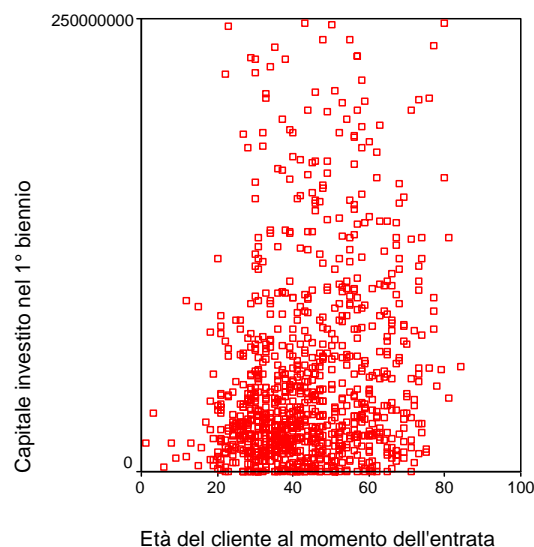
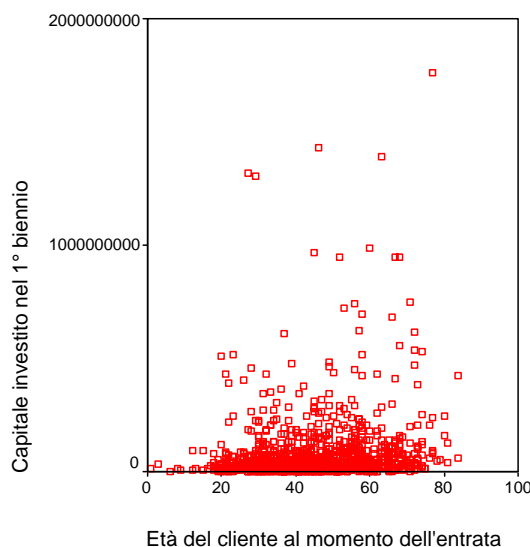
Tab. 3.2 Coefficienti di correlazione tra Capitale investito nel 1° biennio e variabili quantitative

	Età al momento dell'entrata	Numero di figli	Numero di figli a carico
Pearson Correlation	,176	-,041	-,061
Sig. (2-tailed)	,000	,194	,053
Kendall's tau_b t	,152	-,014	-,020
Sig. (2-tailed)	,000	,566	,430
Spearman's rho t	,225	-,019	-,024
Sig. (2-tailed)	,000	,555	,440

I tre coefficienti forniscono le stesse indicazioni: si registra una correlazione non prossima allo zero solo tra capitale investito nel 1° biennio e età al momento dell'ingresso: persone più mature sembrano essere anche più "coraggiose" (o comunque aver più liquidità da investire in prodotti fondiari o assicurativi). Solo per tale coppia di caratteri viene rifiutata **l'ipotesi di correlazione nulla**: il **p-value** è molto basso a differenza di quanto accade per le altre due coppie di caratteri, per le quali l'ipotesi nulla viene accettata se si fissa un **livello di significatività** pari a 0.05.).

Ovviamente, rifiutare l'ipotesi di non correlazione **non consente di concludere che i due caratteri sono fortemente correlati**. Per valutare più accuratamente la affidabilità dei coefficienti considerati, riportiamo il **diagramma di dispersione** tra Capitale investito nel 1° biennio e l'età al momento dell'entrata. Dal grafico non è possibile fare considerazioni sull'associazione tra i due caratteri: nel grafico a destra è stata considerata la porzione del piano cartesiano entro la quale si collocano la maggior parte delle osservazioni. Notiamo che il basso valore del coefficiente di correlazione risulta piuttosto affidabile, descrivendo una relazione lineare di bassa entità.

[Cautele nella valutazione del coefficiente di correlazione lineare]



La correlazione non è invece significativamente diversa da zero per il numero di figli e numero di

figli a carico. I risultati ottenuti sono confermati anche dall'analisi dei diagrammi di dispersione e dei box plot affiancati, riportati nella Figura 3.2 insieme ai diagrammi delle medie condizionate.

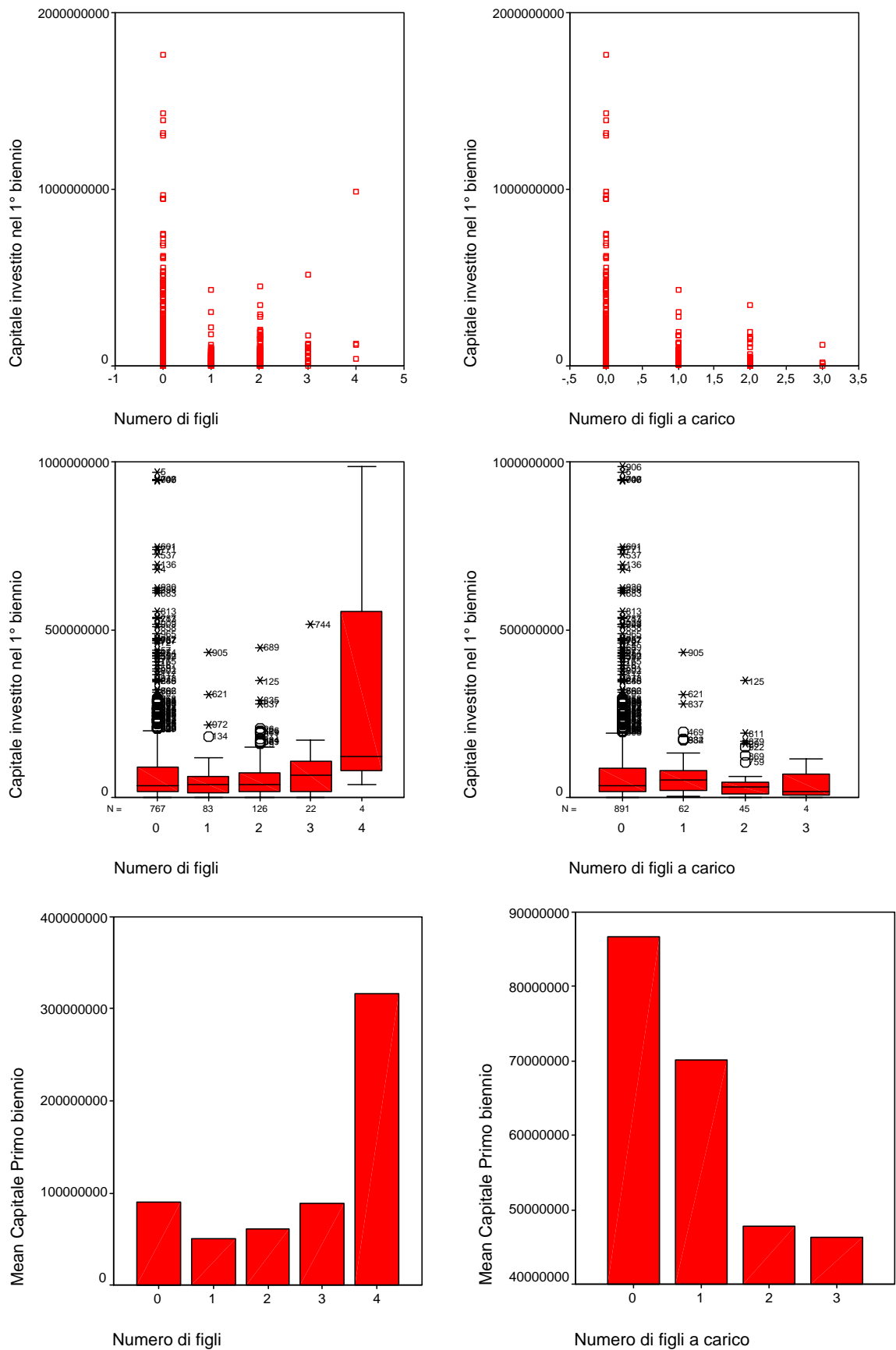


Fig. 3.2. Relazione tra Capitale investito nel 1° biennio, Numero di figli e Numero di figli a carico

Dalla Figura 3.2 si nota che famiglie con 4 figli hanno mediamente un investimento di entità di molto superiore alle altre famiglie, ma questo non è abbastanza per determinare una correlazione positiva significativa, anche perché probabilmente la numerosità di questo gruppo di persone è piuttosto esigua a confronto degli altri gruppi.

Ovviamente, il fatto che il capitale investito nel 1° biennio non sia legato linearmente al numero di figli (a carico o no) non implica che tra i caratteri considerati non sussista un altro tipo di relazione, eventualmente non lineare. Ad esempio, considerando il grafico delle medie condizionate al numero dei figli ci si può rendere conto che esiste una relazione, anche se non lineare. Per indagare su tale tipo di relazione, consideriamo l'ANOVA del capitale investito nel 1° biennio dal numero di figli, raggruppando le modalità 3 e 4 in un'unica classe.

Tab. 3.3. Analisi della varianza del Capitale investito nel 1° biennio (in milioni)

Variabile esplicativa: Numero di figli a carico

Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
7.574	3	947	.000

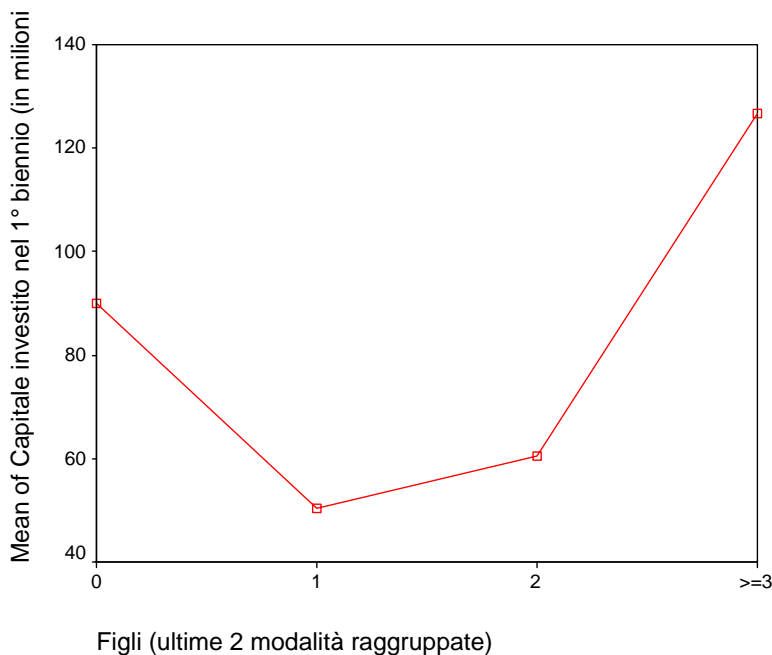
ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	228568.633	3	76189.544	3.506	.015
Within Groups	20582343.602	947	21734.259		
Total	20810912.236	950			

Multiple Comparisons: LSD

(I) Figli	(J) Figli	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1	39.5769	17.27384	.022	5.6774	73.4763
0	2	29.3418	14.37924	.042	1.1230	57.5607
>=3	1	76.5192	34.26249	.026	9.2800	143.7584
>=3	2	66.2842	32.89826	.044	1.7223	130.8461

Stima delle medie condizionate utilizzando il modello



Si può notare che il numero dei figli ha un effetto sul capitale investito, anche se la relazione che lega tra loro i due caratteri non è lineare. In particolare i migliori investitori sono coloro che non hanno figli o che hanno famiglie numerose. I valori medi che caratterizzano questi due gruppi di individui risultano significativamente diversi dai valori medi degli altri due gruppi. La maggior

“bontà” dei clienti con almeno tre figli potrebbe essere dovuta al fatto che a potersi “permettere” famiglie numerose sono i soggetti con reddito elevato. La significatività della variabile “Numero di figli” è confermata dal test non parametrico di Kruskal-Wallis (ma non dal test delle mediane).

[Anova non parametrica]. Per quanto riguarda il numero di figli a carico, anche l’analisi della varianza porta a concludere che i comportamenti di famiglie con diversi figli a carico non sono in media significativamente diversi tra loro.

Dopo aver visto gli effetti marginali, ha senso anche vedere gli effetti congiunti, per comprendere quali differenze tra categorie di persone risultano significative controllando per tutte le altre caratteristiche. [Anova a due (o più) vie: cenni] Riportiamo il modello con gli effetti principali più significativi. I risultati in questo senso sono chiari: le uniche variabili importanti per l’analisi dei capitali spesi nel primo biennio risultano l’anno di ingresso (in classi), l’età del cliente al momento dell’ingresso, lo stato civile, il titolo di studio, la professione e il numero di figli. Sostanzialmente, si registra sia l’effetto del ciclo economico del paese (a momenti economicamente più propizi corrispondono investimenti di maggiore entità) che della posizione sociale dei soggetti.

Si tenga conto del fatto che, a differenza delle analisi della varianza viste prima, la significatività di un carattere esplicativo in questo caso vale *tenendo conto di tutti gli altri*. In sostanza, se due clienti del tutto identici differiscono solo per la modalità assunta in corrispondenza di un certo carattere, il capitale medio investito subirà una variazione significativa.

Tab. 3.4.⁸ Risultati del modello.

Tests of Between-Subjects Effects

Dependent Variable: Capitale investito nel 1° biennio (in milioni)

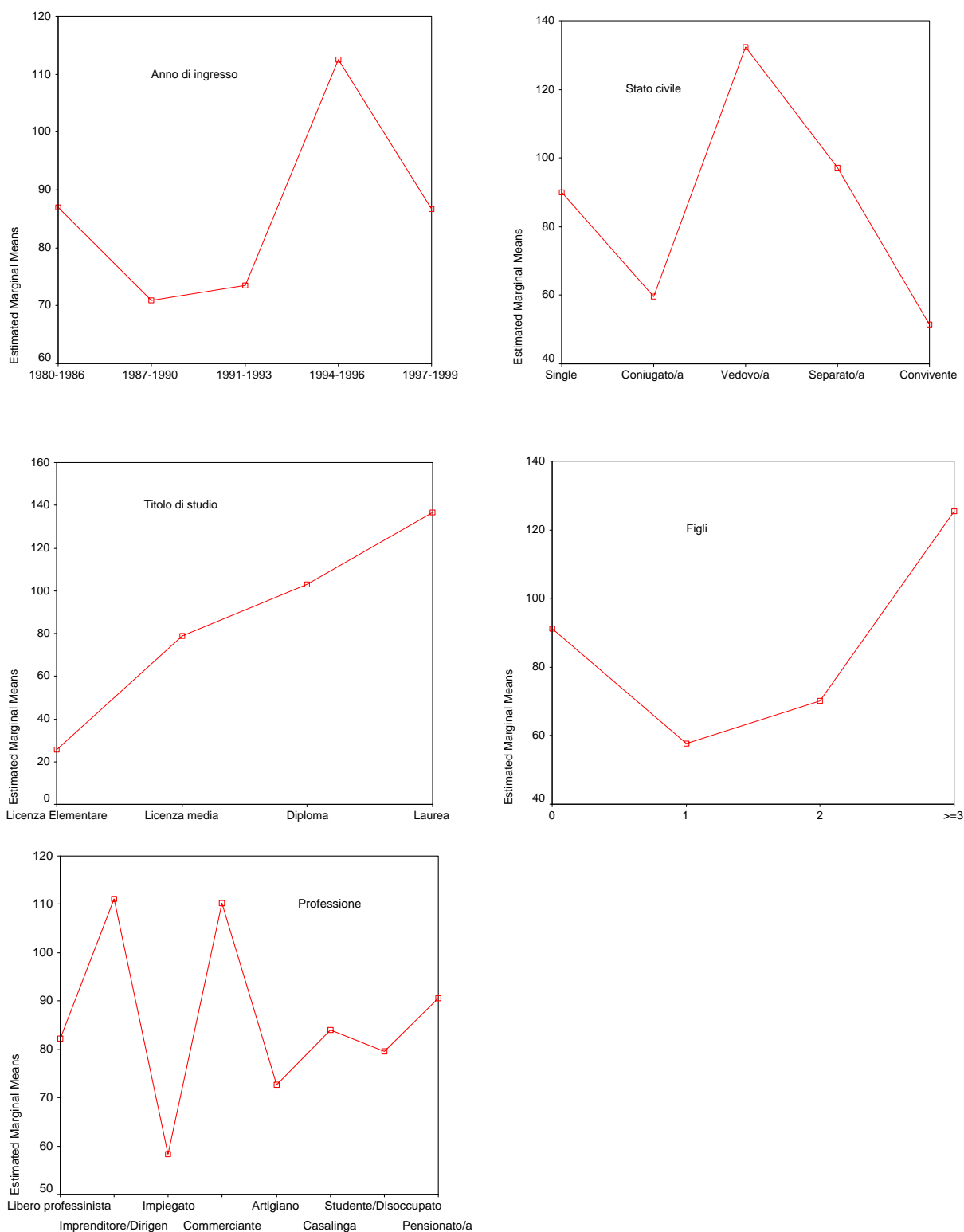
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2121009.031	22	96409.501	4.787	.000
Intercept	30718.248	1	30718.248	1.525	.217
Professione (in classi)	286791.822	7	40970.260	2.034	.048
Stato civile	275679.771	4	68919.943	3.422	.009
Numero di figli	141494.939	3	47164.980	2.342	.072
Età al momento dell'ingresso	216771.982	1	216771.982	10.763	.001
Titolo di studio	350910.092	3	116970.031	5.808	.001
Anno di ingresso (in classi)	188855.370	4	47213.843	2.344	.053
Error	18689903.204	928	20139.982		
Total	27464171.183	951			
Corrected Total	20810912.236	950			

R Squared = .102 (Adjusted R Squared = .081)

Nella Figura 3.3. riportiamo i grafici delle medie marginali dei diversi caratteri (qualitativi o trattati come tali) stimate utilizzando il modello. Anche condizionatamente a tutte le altre variabili simultaneamente, sono i commercianti e gli imprenditori o dirigenti a impegnarsi più attivamente nella società, mentre gli impiegati risultano la categoria meno proficua dal punto di vista dell’azienda. Come ci suggeriva poi l’analisi marginale, titoli di studio più elevati si accompagnano a capitali stanziati di maggiore entità.

⁸ Risultato ottenuto con SPSS (Analyze, General linear model, Univariate, Dependent=Capitale, Fixed Factors=Professione, Stato civile, Titolo di studio, Anno di ingresso, Numero di figli; Covariates: Età al momento dell’ingresso).

Fig. 3.3. Stima dal modello degli effetti marginali.



L'ultima analisi della varianza considerata tiene conto di tutte le variabili simultaneamente; tuttavia in tale analisi si sono considerati solo gli effetti *principali* trascurando gli effetti di *interazione* tra le variabili esplicative considerate. Saremmo a questo punto tentati di valutare la significatività delle interazioni. Tuttavia alcuni tentativi in questo senso scoraggiano, in quanto gli effetti interazione sono quasi tutti poco significativi. Siamo quindi interessati a valutare, almeno a livello descrittivo,

se i profili di clienti ottenuti considerando (combinando le modalità delle) variabili più significative (a livello di effetto principale) differiscono in qualche modo con riferimento alla media del capitale investito (nel 1° biennio). Per questo motivo, nella tabella seguente, accanto ad ogni gruppo “tipico” (cioè le combinazioni più frequenti di caratteristiche) viene indicato sia l’investimento medio che quello mediano.

Tab. 3.5. Medie e mediane dei capitali investiti nel 1° biennio per le categorie più frequenti di clienti.

Stato civile	Tit. di Studio	N° di figli	Professione (classi)	N	Media in milioni	Mediana in milioni
Single	Laurea	0	Libero professionista	20	231,60	78,82
Coniugato/a	Laurea	0	Imprenditore/Dirigente	25	175,09	50,56
Coniugato/a	Laurea	0	Pensionato/a	15	152,01	82,16
Coniugato/a	Diploma	0	Imprenditore/Dirigente	14	108,16	83,99
Coniugato/a	Diploma	0	Pensionato/a	32	107,14	62,00
Coniugato/a	Diploma	0	Commerciante	50	104,42	38,67
Coniugato/a	Laurea	0	Libero professionista	106	101,42	49,31
Single	Laurea	0	Impiegato	24	81,15	33,82
Coniugato/a	Diploma	0	Casalinga	26	72,29	41,85
Coniugato/a	Diploma	2	Pensionato/a	12	71,96	55,27
Coniugato/a	Diploma	0	Libero professionista	23	63,85	41,55
Single	Diploma	0	Studente/Disoccupato	13	61,63	10,16
Coniugato/a	Diploma	0	Impiegato	141	54,93	26,83
Coniugato/a	Diploma	2	Impiegato	22	50,90	43,67
Coniugato/a	Laurea	2	Libero professionista	11	49,21	28,77
Single	Diploma	0	Impiegato	40	45,81	25,38
Coniugato/a	Laurea	0	Impiegato	42	42,47	30,94
Coniugato/a	Lic. media	2	Artigiano	10	36,79	34,31
Coniugato/a	Diploma	1	Pensionato/a	10	33,70	32,32
Coniugato/a	Diploma	0	Artigiano	10	33,27	6,56
Coniugato/a	Diploma	1	Impiegato	19	29,18	22,92

La tabella ci mostra diversi elementi di interesse: innanzi tutto per quasi tutte le categorie analizzate il divario tra media e mediana è notevole, ad indicare la presenza di valori particolarmente estremi che trascinano la media verso valori alti. Un aspetto interessante da questo punto di vista riguarda il fatto che l’ordinamento dei profili secondo il valore medio non coincide con l’ordinamento dei profili secondo il valore mediano. In questo senso, possiamo dire che i “profili” caratterizzati da elevato valore medio sono relativi a clienti piuttosto variabili, alcuni dei quali molto redditizi. Risultano invece più “affidabili” i gruppi di individui con elevato investimento mediano.

[Media, mediana e media troncata per distribuzioni asimmetriche]

Nella Tabella precedente abbiamo valutato la redditività media dei clienti appartenenti ai profili più frequenti. Ovviamente, non è detto che tali categorie di clienti siano necessariamente le più redditizie, cioè quelle che investono in media maggior capitale. Per studiare se esistono e quali sono i profili più redditizi, nella Tabella 3.6 consideriamo il quadro delle categorie che portano un maggior contributo alla media generale. La tabella mostra capitali medi assai più alti della precedente; tuttavia molti valori si riferiscono a gruppi esigui di soggetti, formati da 1, 2 o 3 persone soltanto. Ad esempio, la categoria che investe mediamente in modo più cospicuo è costituita da una sola persona, e quindi l’azienda dovrà tenerne conto prima di pensare a questo gruppo come alla categoria sulla quale investire di più. **[Contributi marginali e di gruppo]**

Una considerazione confortante viene dal fatto che alcuni dei profili più frequenti (ad esempio i liberi professionisti coniugati, laureati e senza figli) sono caratterizzati da un contributo alla media abbastanza elevato, anche se il valore mediano non risulta particolarmente soddisfacente. Notiamo comunque che il divario tra media e mediana risulta ridotto solo quando si considerano piccoli gruppi di individui.

Le considerazioni precedenti evidenziano che sulla base delle caratteristiche socio-demografiche prese in considerazione non è possibile individuare dei gruppi di clienti con comportamenti omogenei almeno con riferimento al capitale investito nel 1° biennio.

Tab. 3.6. Medie e mediane dei capitali investiti nel primo biennio per le categorie che danno i maggiori contributi alla media.

Stato civile	Titolo di Studio	Numero di figli	Professione (in classi)	N	Media in milioni	Mediana in milioni	Contributo alla media
Coniugato/a	Laurea	4	Pensionato/a	1	985,86	985,86	11,79
Vedovo/a	Diploma	0	Casalinga	2	572,23	572,23	6,84
Separato/a	Diploma	0	Commerciante	2	458,47	458,47	5,48
Separato/a	Laurea	2	Pensionato/a	1	447,42	447,42	5,35
Vedovo/a	Lic. media	0	Pensionato/a	3	325,53	26,49	3,89
Vedovo/a	Laurea	0	Pensionato/a	2	312,40	312,40	3,73
Coniugato/a	Diploma	1	Commerciante	1	306,70	306,70	3,67
Single	Diploma	0	Libero professionista	2	305,20	305,20	3,65
Separato/a	Laurea	0	Pensionato/a	2	304,39	304,39	3,64
Coniugato/a	Diploma	0	Studente/Disoccupato	1	287,28	287,28	3,43
Single	Laurea	0	Imprenditore/Dirigente	3	276,33	73,57	3,30
Vedovo/a	Diploma	0	Impiegato	3	255,73	122,80	3,06
Single	Laurea	0	Pensionato/a	7	249,48	205,89	2,98
Separato/a	Laurea	0	Libero professionista	1	234,52	234,52	2,80
Single	Laurea	0	Libero professionista	20	231,60	78,82	2,77
Coniugato/a	Laurea	3	Pensionato/a	3	221,02	117,79	2,64
Vedovo/a	Diploma	0	Pensionato/a	8	211,11	109,52	2,52
Vedovo/a	Diploma	1	Pensionato/a	1	181,97	181,97	2,18
Coniugato/a	Laurea	0	Imprenditore/Dirigente	25	175,09	50,56	2,09
Separato/a	Laurea	1	Impiegato	3	157,32	20,07	1,88
Coniugato/a	Laurea	0	Pensionato/a	15	152,01	82,16	1,82
Coniugato/a	Lic. Elem.	2	Artigiano	1	149,47	149,47	1,79
Vedovo/a	Laurea	0	Libero professionista	2	133,23	133,23	1,59
Coniugato/a	Lic. media	3	Artigiano	1	123,80	123,80	1,48
Separato/a	Laurea	4	Libero professionista	1	122,72	122,72	1,47
Coniugato/a	Diploma	4	Casalinga	1	121,53	121,53	1,45
Separato/a	Diploma	0	Impiegato	2	117,69	117,69	1,41
Coniugato/a	Diploma	0	Imprenditore/Dirigente	14	108,16	83,99	1,29
Coniugato/a	Diploma	0	Pensionato/a	32	107,14	62,00	1,28
Coniugato/a	Laurea	2	Pensionato/a	5	106,08	101,99	1,27
Convivente	Diploma	2	Imprenditore/Dirigente	1	105,34	105,34	1,26
Coniugato/a	Lic. media	0	Pensionato/a	8	104,68	58,10	1,25
Coniugato/a	Diploma	0	Commerciante	50	104,42	38,67	1,25
Coniugato/a	Laurea	2	Impiegato	7	103,26	58,71	1,23
Coniugato/a	Laurea	0	Libero professionista	106	101,42	49,31	1,21
Separato/a	Laurea	2	Impiegato	1	100,46	100,46	1,20
Vedovo/a	Diploma	1	Casalinga	2	99,16	99,16	1,19
Coniugato/a	Lic. media	3	Pensionato/a	3	98,83	106,47	1,18
Coniugato/a	Laurea	3	Libero professionista	1	96,87	96,87	1,16
Coniugato/a	Diploma	3	Impiegato	2	96,84	96,84	1,16
Separato/a	Diploma	1	Artigiano	1	94,40	94,40	1,13
Coniugato/a	Diploma	2	Libero professionista	4	90,82	82,27	1,09
Coniugato/a	Laurea	2	Imprenditore/Dirigente	5	89,38	45,88	1,07
Coniugato/a	Diploma	2	Artigiano	6	86,93	67,74	1,04
Coniugato/a	Lic. media	3	Commerciante	1	86,13	86,13	1,03
Single	Diploma	0	Imprenditore/Dirigente	2	83,00	83,00	0,99

3.2. Relazione tra forma di investimento di ingresso e caratteristiche socio-demografiche.

L'analisi delle scelte di investimento effettuate dai clienti della società ha evidenziato l'importanza della forma di investimento di ingresso (che spesso è anche l'unica). La società, come ultimo sforzo di indagine, è interessata a comprendere se ci sono clienti particolarmente attratti da alcune forme di investimento. Per condurre questo tipo di analisi, è necessario ricorrere allo studio delle tabelle di contingenza ottenute incrociando la forma di investimento di ingresso con le diverse categorie socio-demografiche. Saremo inoltre interessati a valutare l'associazione tra le coppie di caratteri via via considerate e, ovviamente, a verificare statisticamente la significatività dell'associazione eventualmente osservata a livello campionario. **[Misure di associazione]**

Nella Tabella 3.7(a) riportiamo la **tabella a doppia entrata** con la distribuzione congiunta della prima forma di investimento e dell'età di ingresso (in classi). Oltre alle frequenze congiunte assolute riportiamo anche le distribuzioni della prima forma di investimento condizionate alla fascia di età.

[Analisi condizionata: le distribuzioni di frequenza] Le distribuzioni condizionate sono diverse l'una dall'altra, ad indicare la presenza di un'associazione tra i due caratteri. Per valutare se l'**ipotesi nulla di indipendenza** può essere rifiutata, riportiamo le misure di associazione e i rispettivi **p-value**.

Tutte le misure di associazione risultano significativamente diverse da zero, anche se gli indici non assumono valori particolarmente elevati. Si osserva che tutti i clienti che entrano in contatto con l'azienda quando hanno meno di 55 anni prediligono in entrata l'assicurazione o il fondo 4. I clienti entrati in età più avanzata appaiono invece maggiormente attratti dal fondo 2 e dal fondo 6 (se hanno meno di 65 anni) o dal fondo 5 (se hanno più di 65 anni).

La prima forma di investimento risulta significativamente associata anche con lo stato civile. Notiamo comunque che il Lambda risulta poco significativo per questa coppia di variabili. In realtà bisogna prestare attenzione al fatto che alcune categorie di "Stato civile" sono caratterizzate da bassa frequenza, di modo che la tabella si presenta piuttosto sparsa (cioè caratterizzata da molte celle vuote). In questo caso, la differenza tra le distribuzioni condizionate potrebbe essere dovuta anche alla particolare struttura della tabella e non necessariamente alla presenza di un'associazione tra i due caratteri. **[Cautela nella valutazione delle misure di associazione]** Tuttavia, anche limitando l'analisi agli stati civili più frequenti, le misure di associazione risultano significative.

Per quanto riguarda la professione, è evidente un'associazione con la prima forma di investimento. Le distribuzioni condizionate risultano significativamente diverse tra loro (Il Chi-quadro è significativamente diverso da zero). Notiamo che la forma di investimento prediletta da quasi tutte le categorie lavorative è l'assicurazione sulla vita che attira in particolare i commercianti. I pensionati risultano invece maggiormente attratti dal Fondo2, così come gli studenti e i disoccupati (anche se il numero di clienti in questa categoria è così esiguo che è meglio non trarre conclusioni a partire dall'analisi delle distribuzioni condizionate).

L'analisi delle altre distribuzioni congiunte consente di trarre le seguenti conclusioni.

La prima forma di investimento risulta associata, anche se in modo meno forte rispetto ai tre caratteri appena considerati, con l'abitazione. Tuttavia, se vengono eliminate le categorie meno frequenti (abita presso parenti e altro) non si può concludere che le distribuzioni della prima forma di investimento siano significativamente diverse nella sottopopolazione di coloro che abitano in caso di proprietà e in quella di coloro che abitano in affitto **[vedi la tabella]**

Per quanto riguarda il sesso e il fatto che il cliente viva o meno in capoluogo di provincia, l'unica misura di associazione significativa è il Chi-quadro. L'analisi delle distribuzioni congiunte non consente però di trarre alcun tipo di conclusioni su particolari attrazioni tra coppie di modalità.

In sostanza, l'unica informazione rilevante per la società riguarda il fatto che, mentre in generale un generico soggetto è attratto dall'assicurazione, come forma di investimento che lo spinge ad entrare in contatto con la società, i potenziali clienti con età più avanzata e/o pensionati potrebbero essere maggiormente invogliati ad acquistare forme di investimento diverse (soprattutto il Fondo2).

Tab. 3.7(a) Distribuzione congiunta della prima forma di investimento e dell'età al momento dell'ingresso e misure di associazione

Età al momento dell'ingresso in classi		Prima forma di investimento acquistato										Total	
		asvita	fondo0	fondo1	fondo2	fondo3	fondo4	fondo5	fondo6	fondo7	fondo8		fondo9
<25	Count	21	4	8	23	3	13	1	2	7		1	83
	% within Età al momento dell'ingr.	25.3%	4.8%	9.6%	27.7%	3.6%	15.7%	1.2%	2.4%	8.4%		1.2%	100.0%
25-35	Count	75	1	23	53	3	42	9	5	29	6		246
	% within Età al momento dell'ingr.	30.5%	.4%	9.3%	21.5%	1.2%	17.1%	3.7%	2.0%	11.8%	2.4%		100.0%
35-45	Count	91	3	13	44	4	45	7	1	23	12	2	245
	% within Età al momento dell'ingr.	37.1%	1.2%	5.3%	18.0%	1.6%	18.4%	2.9%	.4%	9.4%	4.9%	.8%	100.0%
45-55	Count	57	1	13	45	4	26	8	7	17	4	1	183
	% within Età al momento dell'ingr.	31.1%	.5%	7.1%	24.6%	2.2%	14.2%	4.4%	3.8%	9.3%	2.2%	.5%	100.0%
55-65	Count	24	2	12	47	5	14	9	15	14	4		146
	% within Età al momento dell'ingr.	16.4%	1.4%	8.2%	32.2%	3.4%	9.6%	6.2%	10.3%	9.6%	2.7%		100.0%
>65	Count	12		8	42	1	5	12	3	10	6		99
	% within Età al momento dell'ingr.	12.1%		8.1%	42.4%	1.0%	5.1%	12.1%	3.0%	10.1%	6.1%		100.0%
Total	Count	280	11	77	254	20	145	46	33	100	32	4	1002
	% within Età al momento dell'ingr.	27.9%	1.1%	7.7%	25.3%	2.0%	14.5%	4.6%	3.3%	10.0%	3.2%	.4%	100.0%

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.	
Chi-square tests	Pearson Chi-Square	139.218			.000	
Directional measures Nominal by Nominal	Lambda	Symmetric	.068	.014	4.536	.000
		Età al momento dell'ingr. Dependent	.060	.023	2.496	.013
	Goodman and Kruskal tau	Prima forma di investimento Dependent	.076	.017	4.269	.000
		Età al momento dell'ingr. Dependent	.026	.005		.000
	Uncertainty Coefficient	Prima forma di investimento Dependent	.020	.004		.000
		Symmetric	.037	.006	6.206	.000
Età al momento dell'ingr. Dependent		.040	.006	6.206	.000	
Symmetric measures Nominal by Nominal	Phi	.373			.000	
	Cramer's V	.167			.000	

Tab. 3.7(b) Distribuzione congiunta della prima forma di investimento e dello stato civile

Stato civile		Prima forma di investimento acquistato											Total
		asvita	fondo0	fondo1	fondo2	fondo3	fondo4	fondo5	fondo6	fondo7	fondo8	fondo9	
Single	Count	40	7	13	37	4	20	5	2	25	8	1	162
	% within Stato civile	24.7%	4.3%	8.0%	22.8%	2.5%	12.3%	3.1%	1.2%	15.4%	4.9%	.6%	100.0%
Coniugato/a	Count	213	4	51	197	14	121	35	27	64	20	2	748
	% within Stato civile	28.5%	.5%	6.8%	26.3%	1.9%	16.2%	4.7%	3.6%	8.6%	2.7%	.3%	100.0%
Vedovo/a	Count	10		5	14	2	1	4	1	5	2		44
	% within Stato civile	22.7%		11.4%	31.8%	4.5%	2.3%	9.1%	2.3%	11.4%	4.5%		100.0%
Separato/a	Count	12		5	6		2	2	3	3	2		35
	% within Stato civile	34.3%		14.3%	17.1%		5.7%	5.7%	8.6%	8.6%	5.7%		100.0%
Convivente	Count	5		3			1			3		1	13
	% within Stato civile	38.5%		23.1%			7.7%			23.1%		7.7%	100.0%
Total	Count	280	11	77	254	20	145	46	33	100	32	4	1002
	% within Stato civile	27.9%	1.1%	7.7%	25.3%	2.0%	14.5%	4.6%	3.3%	10.0%	3.2%	.4%	100.0%

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.	
Chi-square tests	Pearson Chi-Square	84.946			.000	
Directional measures Nominal by Nominal	Lambda	Symmetric	.007	.006	1.184	.236
		Stato civile Dependent	.012	.013	.905	.366
		Prima forma di investimento acquistato Dependent	.006	.007	.817	.414
	Goodman and Kruskal tau	Stato civile Dependent	.027	.009		.000
		Prima forma di investimento acquistato Dependent	.007	.002		.004
		Uncertainty Coefficient	Symmetric	.026	.006	4.482
	Stato civile Dependent		.043	.009	4.482	.002
Prima forma di investimento acquistato Dependent	.018		.004	4.482	.002	
Symmetric measures Nominal by Nominal	Phi	.291			.000	
	Cramer's V	.146			.000	

Tab. 3.7(c) Distribuzione congiunta della prima forma di investimento e della professione (in classi)

Professione (in classi)		Prima forma di investimento acquistato											Total
		asvita	fondo0	fondo1	fondo2	fondo3	fondo4	fondo5	fondo6	fondo7	fondo8	fondo9	
Libero professionista	Count	58	1	17	50	3	21	7	8	24	6		195
	% within Professione	29.7%	.5%	8.7%	25.6%	1.5%	10.8%	3.6%	4.1%	12.3%	3.1%		100.0%
Imprenditore/Dirigente	Count	18		6	14	1	10	2	2	10	2	2	67
	% within Professione	26.9%		9.0%	20.9%	1.5%	14.9%	3.0%	3.0%	14.9%	3.0%	3.0%	100.0%
Impiegato	Count	99	5	30	82	8	62	15	8	38	20	1	368
	% within Professione	26.9%	1.4%	8.2%	22.3%	2.2%	16.8%	4.1%	2.2%	10.3%	5.4%	.3%	100.0%
Commerciante	Count	35		3	11		16	3	4	3	2		77
	% within Professione	45.5%		3.9%	14.3%		20.8%	3.9%	5.2%	3.9%	2.6%		100.0%
Artigiano	Count	21	1	6	13	1	13	3	1	5	2		66
	% within Professione	31.8%	1.5%	9.1%	19.7%	1.5%	19.7%	4.5%	1.5%	7.6%	3.0%		100.0%
Casalinga	Count	11	2	4	13	2	8	4	2	3			49
	% within Professione	22.4%	4.1%	8.2%	26.5%	4.1%	16.3%	8.2%	4.1%	6.1%			100.0%
Studente/Disoccupato	Count	2			7	2	1	2	1	1		1	17
	% within Professione	11.8%			41.2%	11.8%	5.9%	11.8%	5.9%	5.9%		5.9%	100.0%
Pensionato/a	Count	36	2	11	64	3	14	10	7	16			163
	% within Professione	22.1%	1.2%	6.7%	39.3%	1.8%	8.6%	6.1%	4.3%	9.8%			100.0%
Total	Count	280	11	77	254	20	145	46	33	100	32	4	1002
	% within Professione	27.9%	1.1%	7.7%	25.3%	2.0%	14.5%	4.6%	3.3%	10.0%	3.2%	.4%	100.0%

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.	
Chi-square tests	Pearson Chi-Square	122.993			.000	
	Lambda	Symmetric	.027	.009	2.932	.003
Directional measures Nominal by Nominal	Professione (in classi) Dependent	Prima forma di investimento acquistato	.002	.007	.229	.819
		Dependent	.048	.016	3.049	.002
	Goodman and Kruskal tau	Professione (in classi) Dependent	.017	.004		.000
		Prima forma di investimento acquistato	.016	.004		.000
	Uncertainty Coefficient	Symmetric	.031	.005	5.865	.001
		Professione (in classi) Dependent	.032	.005	5.865	.001
Prima forma di investimento acquistato		.029	.005	5.865	.001	
Symmetric measures Nominal by Nominal	Phi	.350			.000	
	Cramer's V	.132			.000	