

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

Facoltà di Ingegneria  
Dipartimento di Tecnica e Gestione dei sistemi industriali

Tesi di Laurea di Primo Livello



**CORSO DI LAUREA IN INGEGNERIA GESTIONALE**

**ANALISI STATISTICA DELLA RELAZIONE TRA  
CONDIZIONI METEOROLOGICHE E VENDITE DEI  
GIORNALI QUOTIDIANI**

**Relatore:** Ch.<sup>mo</sup> Dott. Livio Corain

**Correlatore:** Ch.<sup>mo</sup> Prof. Luigi Salmaso

**Laureando:** Marco Vanacore

**ANNO ACCADEMICO 2010-2011**



# INDICE

<b>SOMMARIO</b>	<b>1</b>
<b>INTRODUZIONE</b> .....	<b>3</b>
<b>CAPITOLO 1. Analisi delle vendite di un giornale quotidiano</b> .....	<b>5</b>
1.1 Obiettivi e modalità di svolgimento dell'attività di ricerca.....	5
1.2 Sviluppo della relazione di vicinanza tra edicole/aree e centraline .....	7
1.3 Costruzione dei database .....	9
<b>CAPITOLO 2. Aspetti statistici per l'analisi delle vendite del giornale</b> .....	<b>13</b>
2.1 Definizione del modello statistico.....	13
2.2 Approfondimento metodologico: la regressione polinomiale e i modelli con variabili Dummy .....	14
2.3 Procedura di stima del modello statistico.....	16
2.4 Il software statistico MINITAB .....	18
<b>CAPITOLO 3. Risultati e discussione</b> .....	<b>21</b>
3.1 Serie temporali delle vendite ed analisi del trend .....	21
3.2 Analisi statistica a 18 mesi .....	37
3.3 Analisi statistica previsionale .....	38
<b>CONCLUSIONI</b>	<b>55</b>
<b>BIBLIOGRAFIA</b>	<b>57</b>



# SOMMARIO

---

L'obiettivo di questa tesi di Laurea è l'analisi statistica della relazione tra condizioni meteorologiche e vendite dei giornali quotidiani.

Nel capitolo 1 "Analisi delle vendite di un giornale quotidiano" vengono illustrate le attività preliminari del progetto e gli obiettivi finali, nello specifico capire quali variabili meteo incidano sulle vendite e sviluppare un modello statistico che ci permetta di elaborare previsioni attendibili sul livello di vendite dei quotidiani in funzioni delle variabili significative.

L'obiettivo del secondo capitolo "Aspetti statistici per l'analisi delle vendite del giornale" è quello di presentare le modalità di costruzione del modello. A partire da un primo modello iniziale, che include l'effetto ciclico, del prezzo del quotidiano oltre alle variabili meteo contenute, si opera con analisi statistiche tali da scartare quelle variabili risultanti non significative.

Il terzo capitolo "Risultati e discussione" è dedicato ad analizzare i risultati ottenuti e ad utilizzare il modello per effettuare previsioni, al fine di verificarne la validità a posteriori. In particolare vengono evidenziate per ogni edicola e area le variabili risultate significative, con i rispettivi coefficienti di regressione che andranno a costruire il modello finale. Una volta ottenuto il modello si operano le previsioni e le si confrontano con i reali dati di vendita, allo scopo di testare validità e accuratezza del progetto.

Il lavoro di tesi si conclude con alcuni commenti in merito ai risultati osservati soprattutto mediante i grafici: oltre a porre particolare evidenza ai casi in cui il modello funziona bene e risulti dunque uno strumento utile all'azienda, si fa anche riferimento ai casi laddove risultino esserci delle anomalie, delle quali si tenterà di darne una spiegazione tale per cui anche per le nostre analisi valga il detto che l'eccezione conferma la regola.



# INTRODUZIONE

---

Ottenere informazioni appropriate finalizzate a raggiungere un obiettivo aziendale è sicuramente una fase essenziale dell'attività di una qualsivoglia impresa. Nella maggior parte dei casi i dati sono relativi ai casi di un campione. I campioni sono estratti dalla popolazione in modo tale da essere il più possibile rappresentativi della popolazione stessa. I dati sono i valori osservati con riferimento alle variabili rilevate: nella analisi statistiche si considerano diversi tipi di variabili.

Nella prima fase del lavoro di tesi, è avvenuta la raccolta dei dati di vendita del "Mattino di Padova" a partire da Gennaio 2010 a Giugno 2011 compresi, fornitici direttamente dall'azienda stessa. Il primo passo è stato quello di individuare, se esistente, una possibile relazione tra il numero di quotidiani venduti e le condizioni meteo presenti in prossimità delle edicole. Tali aspetti climatici sono stati messi a disposizione dall' Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto (A.R.P.A.V). A tal proposito, per ogni giorno dell'arco temporale considerato sono stati catalogati in un file Excel tutti i parametri ambientali di rilievo per la ricerca: la radiazione solare, livello di precipitazioni, la temperatura dell'aria a 2 metri (media, minima, massima) e l'umidità relativa (minima, massima) rilevate dalle 18 centraline disposte nel territorio della provincia di Padova. A partire da questi dati è stato possibile costruire un dataset contenente per ogni giorno le rilevazioni meteo corrispondenti alle diverse aree oggetto di studio.

Una volta costruito il dataset, si è proseguito analizzando i dati per mezzo del software Minitab. Il primo obiettivo è stato quello di evidenziare per ogni edicola e area le variabili significative, al fine di depurare il modello grezzo di partenza e ottenere quello da utilizzare successivamente per le previsioni. In particolare, il software permette di eseguire test F parziali aventi come output l'elenco di tutte le variabili con i rispettivi p-value. Si è operato a step, rimuovendo via via le variabili aventi p-value sopra i valori limite, fino ad ottenere un elenco di variabili significative cioè aventi tutte p-value inferiore allo 0,1. Si è utilizzata questa procedura poiché al diminuire del numero di variabili i test statistici diventano sempre più precisi. Una volta ottenuto ciò, si è testata preliminarmente la significatività del modello sui 16 mesi, facendo un controllo sul coefficiente di correlazione. Il coefficiente di correlazione non ha una chiave di lettura assoluta, però maggiore è il suo valore e maggiore è l'accuratezza del modello.

Nella seconda fase del lavoro di tesi, si sono analizzati i risultati ottenuti con il software tramite la costruzione di grafici. Dapprima si è costruito il trend delle vendite di tutte le edicole e aree, per evidenziare l'andamento delle vendite dovuto solo ed esclusivamente al fattore ciclico e al prezzo. Successivamente, si è utilizzato l'output del software Minitab al fine di costruire il modello definitivo, facendo cioè uso solo delle variabili significative con il rispettivo coefficiente di regressione e quindi includendo anche aspetti meteorologici. Il modello è stato dunque utilizzato allo scopo di ottenere delle previsioni di vendita, utilizzando i valori delle variabili individuate, ovvero quanto possono incidere le condizioni meteorologiche sulle vendite dei quotidiani. Una volta ottenute le previsioni di vendita, si è operato al fine di testare la validità del modello a posteriori. Si sono dunque costruiti dei grafici mostrandoti il trend, l'andamento delle vendite previste e l'andamento delle vendite reali. Da questi grafici è stato possibile fare un rapido confronto fra quanto si è previsto e quanto è realmente avvenuto. Si sono riscontrati casi in cui il modello è risultato adeguato e funzionante, ma anche casi dove la sua accuratezza non è risultata sufficiente rendendolo pertanto uno strumento poco adeguato.

# **CAPITOLO 1.   Analisi delle vendite di un giornale quotidiano**

---

In questo capitolo vengono descritti i dettagli specifici della ricerca alla quale la presente tesi di laurea fa riferimento ovvero le analisi delle vendite di un giornale quotidiano, specificatamente del "Mattino di Padova".

## **1.1 Obiettivi e modalità di svolgimento dell'attività di ricerca**

L'obiettivo che si prefigge la ricerca è l'analisi delle vendite del giornale quotidiano denominato "Mattino di Padova" al fine di sviluppare e sperimentare un modello statistico predittivo a supporto della distribuzione del Mattino di Padova nelle edicole del territorio della provincia di Padova.

A tal scopo, il gruppo coinvolto nell'attività di ricerca ha realizzato le seguenti attività preliminari:

- Richiesta all'A.R.P.A.V (Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto), in riferimento a tutte le stazioni sia agrometeorologiche, sia meteorologiche della Provincia di Padova (ovvero Agna ,Balduina Sant'Urbano, Ca' di Mezzo (Codevigo), Campodarsego, Codevigo, Este (loc. Calaone), Faedo Cinto Euganeo, Galzignano (Via Ca Demia), Grantorto, Legnaro, Masi, Montagnana, Ponte Zata (Tribano), Teolo, Trebaseleghe, Stazioni Meteorologiche, Cittadella, Monte Grande Rovolon, Orto Botanico di Padova), dei dati giornalieri del periodo 01/01/2010-30/06/2011 per i seguenti parametri climatici (Figura 1):
  - Precipitazioni (mm)
  - Radiazione solare (MJ/m<sup>2</sup>)
  - Temperatura dell'aria a 2 metri: minima, media e massima (°C)
  - Umidità relativa a 2m: minima e massima (%)

Provincia di PADOVA				
Stazioni Agrometeorologiche	Coordinate Gauss-Boaga fuso Ovest		Quota m s.l.m.	Anno attivazione
	X	Y		
<a href="#">Aqna (169)</a>	1732500	5004921	2	1992
<a href="#">Balduina Sant'Urbano (152)</a>	1703222	5001188	8	1994
<a href="#">Ca' di Mezzo (Codevigo) (211)</a>	1746929	5012991	6	1996
<a href="#">Campodarsego (179)</a>	1727668	5042147	15	1992
<a href="#">Codevigo (175)</a>	1743297	5014716	0	1992
<a href="#">Este (loc. Calaone) (180)</a>	1708384	5013285	69	1991
<a href="#">Faedo Cinto Euganeo (142)</a>	1711449	5020414	247	1994
<a href="#">Galzignano (Via Ca Demia) (265)</a>	1715064	5019974	13	2004
<a href="#">Grantorto (177)</a>	1714504	5052654	31	1991
<a href="#">Legnaro (111)</a>	1731296	5025799	8	1991
<a href="#">Masi (151)</a>	1695166	4999008	8	1994
<a href="#">Montagnana (106)</a>	1693723	5012987	13	1990
<a href="#">Ponte Zata (Tribano) (182)</a>	1723829	5007659	4	1996
<a href="#">Teolo (170)</a>	1709767	5024532	158	1992
<a href="#">Trebalegha (122)</a>	1736015	5054913	23	1995
Stazioni Meteorologiche	Coordinate Gauss-Boaga fuso Ovest		Quota m s.l.m.	Anno attivazione
	X	Y		
<a href="#">Cittadella (110)</a>	1717437	5060795	56	1991
<a href="#">Monte Grande Rovolon F2 (201)</a>	1709358	5026653	454	1992
<a href="#">Orto Botanico di Padova (234)</a>	1725465	5031295	12	2000

Figura 1. Stazioni Agrometeorologiche e Meteorologiche della Provincia di Padova.

- costruzione di un database ad hoc su base Excel per l'inserimento delle variabili meteorologiche;
- inserimento nel suddetto database dei valori corrispondenti a tali fattori ambientali, con una copertura finale di 18 mesi (gen. 2010 – giu. 2011);
- collegamento del database meteo con il database vendite.

La ricerca si propone principalmente i due seguenti obiettivi:

1. stabilire, sia a livello aggregato di zona geografica della provincia di Padova sia a livello di singola edicola, se esistono delle associazioni significative tra le vendite e le caratteristiche meteo;
2. costruire e validare un modello statistico capace di fare delle previsioni possibilmente affidabili delle vendite in funzione delle caratteristiche climatiche e di tutti i possibili fattori potenzialmente impattanti sulle vendite.

## 1.2 Sviluppo della relazione di vicinanza tra edicole/aree e centraline

Per costruire un modello coerente abbiamo relazionato le edicole con le centraline utilizzando il criterio della minima distanza tra le stesse; a tale scopo, per ciascun osservatorio abbiamo convertito (Figura 2) le relative coordinate Gauss-Boaga fuso Ovest (messe a disposizione dall'Arpav) nell'equivalente proiezione Universale Trasversa di Mercatore (proiezione UTM) mediante il software "Traspunto".

The screenshot shows the 'Interfaccia per l'applicativo TRASPUNTO' window. It is divided into several sections:

- DATO INGRESSO:** Sistema-Datum:  Gauss Boaga - Roma 40,  UTM - ED 50,  UTM - WGS 84. Fuso:  32 - OVEST,  33 - EST. Coordinate:  Piane,  Geografiche.
- DATO USCITA:** Sistema-Datum:  Gauss Boaga - Roma 40,  UTM - ED 50,  UTM - WGS 84. Fuso:  32 - OVEST,  33 - EST. Coordinate:  Piane,  Geografiche.
- FORMATO DATI:** Formato Coord. Geografiche:  Gradi.PrimiSecondi,  GradiPrimiSecondi,  Gradi Primi Secondi,  Gradi\* Primi\* Secondi",  Radianti. Ordine Coordinate:  Nord - Est,  Est - Nord. Geografiche:  fi - lambda,  lambda - fi.
- PUNTO:** Coordinate Ingresso: Nord: 5004921, Est: 1732500. Coordinate Uscita: Nord: 5004900.157, Est: 732468.986.

Buttons at the bottom: PUNTO, FILE ELENCO, FILE DXF, FILE SHP-SHX-DBF, FILE BMP-BPw, DTM, ?, TRASFORMA.

Figura 2. Esempio conversione coordinate Gauss-Boaga-UTM.

Il risultato ottenuto è stato successivamente trasformato in latitudine/longitudine, utilizzando sempre il medesimo programma (Figura 3 e Figura 4).

Geographic (degrees decimal)		To/From	UTM	
lon	<input type="text" value="11.95775295753781"/>	<input type="button" value="&gt;&gt;"/>	x (easting)	<input type="text" value="732468.986"/>
lat	<input type="text" value="45.15927081986975"/>	<input type="button" value="&lt;&lt;"/>	y (northing)	<input type="text" value="5004900.157"/>
			zone	<input type="text" value="32"/>
			hemisphere	<input checked="" type="radio"/> N <input type="radio"/> S

*Geographic coordinates are entered and displayed in decimal degrees. Negative numbers indicate West longitudes and South latitudes. UTM coordinates are entered and displayed in meters.  
The ellipsoid model used for computations is WGS84.*

Figura 3. Esempio conversione coordinate UTM-lon/lat.

N. Centr.	Tipo centralina	Nome centralina	Coordinate Gauss-Boaga fuso Ovest X	Coordinate Gauss-Boaga fuso Ovest Y	Quota m s.l.m.	Anno attivazione	Latitudine	Longitudine
1	Agrometeorologiche	Agna	1732500	5004921	2	1992	45,1592708198697	11,9577529575378
2	Agrometeorologiche	Balduina Sant'Urbano	1703222	5001188	8	1994	45,1347423693073	11,5841384503179
3	Agrometeorologiche	Ca' di mezzo (Codevigo)	1746929	5012991	6	1996	45,2289034258602	12,1450601994531
4	Agrometeorologiche	Campodarsago	1727668	5042147	15	1992	45,4954940822988	11,9134780567580
5	Stazioni Meteorologiche	Cittadella	1717437	5060795	56	1991	45,6664060973797	11,7909998681411
6	Agrometeorologiche	Codevigo	1743297	5014716	0	1992	45,2436722472122	12,0997079352278
7	Agrometeorologiche	Este (loc. Calzaone)	1708364	5013285	69	1991	45,2420110708111	11,6547744288720
8	Agrometeorologiche	Faedo Cinto Euganeo	1711449	5020414	247	1994	45,3051965442029	11,6968209836609
9	Agrometeorologiche	Galzignano (Via Ca' Demia)	1715064	5019974	13	2004	45,3001425126664	11,7426884569772
10	Agrometeorologiche	Grantorto	1714504	5052654	31	1991	45,5941299715090	11,7498072069021
11	Agrometeorologiche	Legnaro	1731296	5025799	8	1991	45,3473474551145	11,9521728201124
12	Agrometeorologiche	Masi	1695166	4999008	8	1994	45,1174098157725	11,4809339664414
13	Agrometeorologiche	Montagnana	1693723	5012987	13	1990	45,2435204792550	11,4680376446452
14	Stazioni Meteorologiche	Monte Grande Rovolon F2	1723829	5007659	4	1996	45,3619209458329	11,6728184856944
15	Stazioni Meteorologiche	Orto Botanico di Padova	1709767	5024532	158	1992	45,3986542171498	11,8803447376703
16	Agrometeorologiche	Teolo	1736015	5054913	23	1995	45,3427279424400	11,6771349219717
17	Agrometeorologiche	Trebaseleghe	1709358	5026653	454	1992	45,6074653900379	12,0263173181927
18	Agrometeorologiche	Tribano (Ponte Zata)	1725465	5031295	12	2000	45,1866894198086	11,8488001236038

Figura 4. Elenco coordinate centraline.

A questo punto, utilizzando congiuntamente le coordinate delle centraline e delle edicole, abbiamo calcolato la distanza tra le stesse sfruttando il teorema di Pitagora ed evidenziando quindi quella minimale (si veda Figura 5).

Cod. Edicola	Teolo	Trebaseleghe	Tribano (Ponte Zata)	N. Centr. + vicina	Area
2040380	11,7	34,8	15,8	9	TERME
2040381	20,0	21,3	28,3	15	PADOVA
2040382	13,5	47,5	9,7	7	MONSELICENSE
2040384	13,4	52,0	16,7	7	ESTENSE
2040394	19,1	22,4	27,2	15	PADOVA
2040399	20,7	21,9	27,1	15	PADOVA
2040400	15,1	48,1	8,1	18	MONSELICENSE
2040401	3,2	43,1	22,6	16	TERME
2040402	5,8	34,7	21,5	14	TERME
2040403	9,4	30,6	25,7	14	CINTURA URBANA
2040404	4,3	35,8	22,8	14	TERME
2040405	5,2	38,0	26,9	14	TERME
2040406	3,9	36,1	22,8	14	TERME
2040407	8,7	31,4	25,4	14	CINTURA URBANA
2040408	4,6	35,6	24,0	14	TERME
2040410	9,0	31,1	24,2	14	CINTURA URBANA
2040411	8,1	32,4	22,4	14	CINTURA URBANA
2040412	3,7	36,3	22,8	14	TERME

Figura 5. Esempio distanza e relazione Edicola-centralina

Per i rapporti area-centralina, invece, si è tenuto conto di un metodo differente: la centralina di riferimento è quella che risulta avere il conteggio maggiore relativo al

numero di edicole nella medesima area per le quali si rivela anche come "centralina più vicina". Per una maggiore comprensione del criterio viene di seguito illustrato un esempio:

Conteggio di Cod. Edicola		
Area	N. Centr. + vicina	Totale
CAMPOSAMPIERESE	4	28
	5	1
	9	1
	10	2
	15	6
	17	18

Figura 6. Esempio relazione Area-centralina

L'area Camposampierese viene relazionata con la centralina 4, essendo massimo il relativo conteggio.

### 1.3 Costruzione dei database

Per ottenere la base dati necessaria alle elaborazioni statistiche utili a raggiungere i due obiettivi prefissati, sono stati costruiti tre database, denominati rispettivamente:

1. Database meteo;
2. Database vendite;
3. Database previsioni.

Il database meteo (Figura 7), è costituito da 536 rilevazioni meteo giornaliere per un totale di 67536 misurazioni climatiche, classificate come precedentemente descritto.

Data Ora	Precip_C1	Rad_sol_C1	Temp_aria_min_C1	Temp_aria_med_C1	Temp_aria_max_C1	Umid_rel_min_C1	Umid_rel_max_C1	Precip_C2	Rad_sol_C2	Temp_aria_min_C2	Temp_aria_med_C2	Temp_aria_max_C2
01/01/2010	6,4	1,33	0,9	4,3	5,8	89	100	2,5	1,287	0,6	3,9	3,9
02/01/2010	0,2	5,567	-2,2	0,7	4,9	48	100	0,4	5,429	-2,8	0,1	0,1
03/01/2010	2,8	1,764	-1,4	-0,4	1,3	70	100	0,4	1,62	-2,2	-0,3	-0,3
04/01/2010	5,2	1,092	-0,2	0,8	2,7	97	100	4,6	1,017	-0,1	0,3	0,3
05/01/2010	1,2	4,389	0	2,2	3,7	73	100	0,4	3,09	0,3	1,4	1,4
06/01/2010	1,2	5,56	1	4	7,0	78	100	2	5,768	0,6	3,5	3,5
07/01/2010	11,8	0,872	4,8	5,6	6,4	87	100	13	1,249	4,1	5,0	5,0
08/01/2010	9,0	2,297	5,5	6,9	8,5	87	100	10	1,45	4,1	6,3	6,3
09/01/2010	0,2	5,906	0,0	4,9	9	72	100	0,6	6,053	1,8	4,9	4,9
10/01/2010	0	3,709	0,2	2,3	4,5	84	100	0	3,27	-0,6	1,8	1,8
11/01/2010	0	1,958	1,9	3,6	5,7	92	100	0	1,902	1,4	2,8	2,8
12/01/2010	0	2,220	2,6	4,1	6,4	78	100	0,4	2,220	1,9	3,7	3,7
13/01/2010	0	2,975	2,2	4	6,4	73	100	0	1,958	1,8	3,2	3,2
14/01/2010	0	4,029	1	4,3	8,1	56	95	0	3,791	0,7	4	4
15/01/2010	0	7,242	0,2	2,4	6,6	53	88	0	7,594	-0,7	2,1	2,1
16/01/2010	0	2,184	0,8	2,1	4,8	69	92	0	1,807	0,4	2	2
17/01/2010	0	6,295	-0,7	1,5	5,2	79	100	0,2	2,768	-1,4	0,6	0,6
18/01/2010	0,2	1,92	-1,1	0,1	1,5	100	100	0	1,958	-0,8	0,1	0,1
19/01/2010	0	1,538	-0,7	0,6	2	98	100	0	1,734	-0,9	0,2	0,2
20/01/2010	0	2,573	0,7	1,8	3,1	87	100	0	1,939	0,2	1,3	1,3
21/01/2010	0	4,023	0	2,5	4,8	70	98	0	4,268	0,4	2,3	2,3
22/01/2010	0	6,778	-2,1	0,5	5	60	100	0	6,188	-1,7	0,5	0,5
23/01/2010	0	1,845	-1,9	-0,1	1,1	76	100	0	2,372	-2,6	-0,2	-0,2
24/01/2010	0	4,029	-1,3	0,8	2,7	68	89	0	3,389	-1,1	0,9	0,9
25/01/2010	0	1,356	-0,5	0,4	2,4	70	100	4,4	1,218	-0,2	0,3	0,3
26/01/2010	0	8,692	-2,8	-0,1	4,2	46	95	0	9,263	-4,8	-0,1	-0,1
27/01/2010	0	4,87	-4,5	-1	3,3	59	100	0,2	4,927	-2,5	-0,5	-0,5
28/01/2010	1	3,496	-4,2	0	2,9	68	100	0,8	2,906	-3,3	-0,6	-0,6
29/01/2010												

Figura 7. Alcuni record del database meteo.

Sulla base della suddivisione interna adottata dalla divisione diffusione e marketing del quotidiano, la provincia di Padova è stata suddivisa in 8 aree geografiche il cui dettaglio viene riportato in Tabella 1

*Tabella 1. Dettaglio della suddivisione del territorio della provincia di Padova in 8 aree.*

Area n.	Descrizione area	Abbreviazione area
1	Padova (dentro le mura)	CAMP
2	Cintura urbana	CINT
3	Camposampierense	CITT
4	Cittadelense	ESTE
5	Estense	MONS
6	Monselicense	PADO
7	Piovese	PIOV
8	Terme Euganee	TERM

Infine, il database meteo è stato collegato al "database vendite" (Figura 8) in modo da costituire il "database previsioni", che comprende due versioni: la prima utile ad una analisi aggregata per area (Figura 9), la seconda finalizzata allo studio delle vendite per singola edicola.

A	B	C	D	E	G	H	I	J	K	L	M	N	
Data	Comune	CAP	Località/Quartiere	Cod. Edicola	Inviato	Reso	Vendita	Esaurito	Gior	set	Settimana	Mese	Zona
02/01/2010	Padova	35100	PADOVA	2040001	60	15	45	0	7	1	1	1	PADOVA
03/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	2	1	1	PADOVA
04/01/2010	Padova	35100	PADOVA	2040001	40	1	39	0	2	2	1	1	PADOVA
05/01/2010	Padova	35100	PADOVA	2040001	40	1	39	0	3	2	1	1	PADOVA
06/01/2010	Padova	35100	PADOVA	2040001	40	26	14	0	4	2	1	1	PADOVA
07/01/2010	Padova	35100	PADOVA	2040001	75	38	37	0	5	2	1	1	PADOVA
08/01/2010	Padova	35100	PADOVA	2040001	60	23	37	0	6	2	1	1	PADOVA
09/01/2010	Padova	35100	PADOVA	2040001	50	6	44	0	7	2	1	1	PADOVA
10/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	3	1	1	PADOVA
11/01/2010	Padova	35100	PADOVA	2040001	45	16	29	0	2	3	1	1	PADOVA
12/01/2010	Padova	35100	PADOVA	2040001	43	2	41	0	3	3	1	1	PADOVA
13/01/2010	Padova	35100	PADOVA	2040001	50	8	42	0	4	3	1	1	PADOVA
14/01/2010	Padova	35100	PADOVA	2040001	50	4	46	0	5	3	1	1	PADOVA
15/01/2010	Padova	35100	PADOVA	2040001	49	11	38	0	6	3	1	1	PADOVA
16/01/2010	Padova	35100	PADOVA	2040001	50	0	50	1	7	3	1	1	PADOVA
17/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	4	1	1	PADOVA
18/01/2010	Padova	35100	PADOVA	2040001	35	11	24	0	2	4	1	1	PADOVA
19/01/2010	Padova	35100	PADOVA	2040001	45	13	32	0	3	4	1	1	PADOVA
20/01/2010	Padova	35100	PADOVA	2040001	45	1	44	0	4	4	1	1	PADOVA
21/01/2010	Padova	35100	PADOVA	2040001	40	0	40	1	5	4	1	1	PADOVA
22/01/2010	Padova	35100	PADOVA	2040001	40	1	39	0	6	4	1	1	PADOVA
23/01/2010	Padova	35100	PADOVA	2040001	55	7	48	0	7	4	1	1	PADOVA
24/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	5	1	1	PADOVA
25/01/2010	Padova	35100	PADOVA	2040001	45	19	26	0	2	5	1	1	PADOVA
26/01/2010	Padova	35100	PADOVA	2040001	43	4	39	0	3	5	1	1	PADOVA
27/01/2010	Padova	35100	PADOVA	2040001	50	9	41	0	4	5	1	1	PADOVA
28/01/2010	Padova	35100	PADOVA	2040001	40	9	31	0	5	5	1	1	PADOVA
29/01/2010	Padova	35100	PADOVA	2040001	45	10	35	0	6	5	1	1	PADOVA
30/01/2010	Padova	35100	PADOVA	2040001	50	11	47	0	7	5	1	1	PADOVA
31/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	6	1	1	PADOVA
02/02/2010	Padova	35100	PADOVA	2040003	65	27	38	0	7	1	1	1	PADOVA
03/02/2010	Padova	35100	PADOVA	2040003	1	1	0	0	1	2	1	1	PADOVA
04/02/2010	Padova	35100	PADOVA	2040003	35	1	34	0	2	2	1	1	PADOVA

*Figura 8. Alcuni record del database vendite.*

#	A	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
1	Data	VEN_426	VEN_429	VEN_437	VEN_443	VEN_458	VEN_491	Prezzo	Giorno	Giorno*2	Giorno*3	Mese	Gior_sett	Rad_sol_C1	Temp_aria_min_C1	Temp_aria_med_C1	Temp_aria_max_C1
2	02/01/2010	14	40	10	90	228	12	1	1	1	1	1	7	1,33	0,9		4,3
3	03/01/2010	19	52	7	128	255		1	2	4	8	1	1	5,567	-2,2		0,7
4	04/01/2010	12	21	3	87	165	15	1	3	9	27	1	2	1,764	-1,4		-0,4
5	05/01/2010	13	27	7	92	185	7	1	4	16	64	1	3	1,092	-0,2		0,8
6	06/01/2010	10	39	8	167	255		1	5	25	125	1	4	4,355	0		2,2
7	07/01/2010	11	26	4	108	188	19	1	6	36	216	1	5	5,56	1		4
8	08/01/2010	12	24	3	83	176	13	1	7	49	343	1	6	0,872	-4,8		5,6
9	09/01/2010	13	40	5	95	200	11	1	8	64	512	1	7	2,297	5,5		6,9
10	10/01/2010	17	46	7	146	274		1	9	81	729	1	1	5,906	0,8		-4,9
11	11/01/2010	12	24	3	89	128	15	1	10	100	1000	1	2	3,709	0,2		2,3
12	12/01/2010	11	38	5	99	152	13	1	11	121	1331	1	3	1,958	1,9		3,6
13	13/01/2010	12	29	12	148	165	11	1	12	144	1728	1	4	2,228	2,6		-4,1
14	14/01/2010	11	25	5	91	179	10	1	13	169	2197	1	5	2,975	2,2		4
15	15/01/2010	13	26	5	98	172	12	1	14	196	2744	1	6	4,029	1		-4,3
16	16/01/2010	15	37	5	104	204	6	1	15	225	3375	1	7	7,242	0,2		2,4
17	17/01/2010	11	58	10	160	290		1	16	256	4096	1	1	2,184	0,8		2,1
18	18/01/2010	14	23	6	89	140	32	1	17	289	4913	1	2	6,295	-0,7		1,5
19	19/01/2010	14	25	5	82	144	14	1	18	324	5832	1	3	1,92	-1,1		0,1
20	20/01/2010	9	20	1	129	164	16	1	19	361	6859	1	4	1,538	-0,7		0,6
21	21/01/2010	12	30	8	90	155	9	1	20	400	8000	1	5	2,573	0,7		1,8
22	22/01/2010	13	22	5	90	148	12	1	21	441	9261	1	6	4,023	0		2,5
23	23/01/2010	9	26	5	100	213	8	1	22	484	10648	1	7	6,778	-2,1		0,5
24	24/01/2010	15	46	8	154	322		1	23	529	12167	1	1	1,845	-1,9		-0,1

Figura 9. Alcuni record del database previsioni.



## **CAPITOLO 2. Aspetti statistici per l'analisi delle vendite del giornale**

---

Questo capitolo è specificatamente finalizzato a presentare nel dettaglio gli aspetti statistici che hanno interessato l'attività di analisi delle vendite del giornale, in particolare la procedura di ricerca delle variabili significative, l'utilizzo di test F parziali al fine di evidenziare i p-value al di sopra dei valori limite e la costruzione del modello previsionale.

### **2.1 Definizione del modello statistico**

Premessa necessaria all'analisi statistica finalizzata a stabilire se esistono delle associazioni significative tra le caratteristiche climatiche presenti in prossimità dell'edicola o area e le vendite/esauriti del quotidiano stesso è il tenere in debita considerazione l'effetto ciclico (giorno della settimana e mese dell'anno) che è presente naturalmente nel fenomeno delle vendite dei quotidiani. Dopo aver incluso nel modello statistico sia l'effetto ciclico sia il prezzo di vendita del quotidiano (aumentato, a partire dal 1° gennaio 2011, da 1 a 1,20 Euro) è possibile studiare in modo adeguato l'eventuale impatto su vendite/esauriti della rilevanza e tipologia delle condizioni meteorologiche. A tal fine, separatamente per ciascuna delle 8 aree geografiche in cui è stata suddivisa la provincia di Padova, si è adottato un modello rappresentato in Figura 10.

*Figura 10. Rappresentazione grafica del modello statistico per lo studio delle associazioni tra le caratteristiche meteorologiche e le vendite/esauriti.*

Più precisamente, in prima istanza si possono considerare i seguenti modelli lineari:

$$V_i = \text{Trend}_i + \text{GSett}_i + \text{Mese}_i + \text{GSett}_i * \text{Mese}_i + \text{Prezzo}_i + \text{FattClim}_i + \varepsilon_i,$$

dove  $V_i$  indica le vendite osservate il giorno  $i$ -esimo,  $\text{Trend}_i$  l'eventuale trend presente nelle serie temporale delle vendite (effetto di lungo periodo),  $\text{GSett}_i$  e  $\text{Mese}_i$  indicano l'effetto ciclico del giorno della settimana e del mese di riferimento dello specifico giorno  $i$ -esimo,  $\text{Prezzo}_i$  il prezzo di vendita del quotidiano (aumentato, a partire dal 1° gennaio 2011, dal valore di 1 ad 1,20 Euro),  $\text{FattClim}_i$  indica il tipo di parametro ambientale (precipitazioni, radiazione solare, temperatura (media, minima, massima), umidità relativa (minima, massima). Più precisamente,  $\text{FattClim}_i$  si riferisce all'insieme di coefficienti che misurano l'effetto del meteo sulla vendita di giornali nel giorno  $i$ -esimo. Infine,  $\varepsilon_i$  indica una componente di errore casuale (che tiene conto degli aspetti non misurati/osservabili e di quelli intrinsecamente aleatori) che si assume di media zero, indipendente ed identicamente distribuita secondo una legge gaussiana.

## 2.2 Approfondimento metodologico: la regressione polinomiale e i modelli con variabili Dummy

Si definisce regressione polinomiale quella relazioni in cui le variabili dipendenti appaiono sotto forma di polinomio. Tra le più comuni vi è quella quadratica, il cui modello generale è:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \varepsilon_i$$

dove

$\beta_0 = Y_i$  intercetta

$\beta_1 =$  coefficiente relativo all'effetto lineare su  $Y_i$

$\beta_2 =$  coefficiente relativo all'effetto quadratico su  $Y_i$

$\varepsilon_i =$  l'errore casuale in  $Y_i$  corrispondente all' $i$ -esima osservazione

Il modello di regressione quadratica differisce da quello strettamente lineare per la presenza al quadrato di una variabile esplicativa.

Analogamente al modello di regressione multipla, i coefficienti ( $\beta_0$ ,  $\beta_1$  e  $\beta_2$ ) possono essere stimati attraverso i corrispondenti parametri campionari ( $b_0$ ,  $b_1$  e  $b_2$ ); ne deriva pertanto l'espressione campionaria del modello quadratico:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{1i}^2$$

Per verificare se ci sia una relazione significativa tra la variabile dipendente e l'insieme delle variabili esplicative si utilizza la statistica test F, la quale si basa sul rapporto tra la media dei quadrati della regressione e la media dei quadrati degli errori:

$$F = \frac{MQR}{MQE}$$

con l'ipotesi nulla e l'ipotesi alternativa specificate nella seguente maniera:

$H_0: \beta_1 = \beta_2 = 0$  (non vi è relazione significativa tra  $X_1$  e  $Y$ )

$H_1: \beta_1 \text{ e/o } \beta_2 \neq 0$  (vi è relazione significativa tra  $X_1$  e  $Y$ )

Tuttavia nel procedimento per la scelta del modello più appropriato occorre stabilire se vi sia una differenza sostanziale tra la relazione quadratica e quella lineare. I due modelli possono essere confrontati valutando l'effetto dell'aggiunta del termine quadratico a un modello in cui sia già stato inserito un termine lineare. A tal fine si può utilizzare la statistica test  $t$  formalizzando la statistica nel modo seguente:

$H_0: \text{ l' inclusione del termine quadratico non migliora in maniera significativa il modello } (\beta_2 = 0)$

$H_1: \text{ l' inclusione del termine quadratico migliora in maniera significativa il modello } (\beta_2 \neq 0)$

$$t = \frac{b_2}{S_{b_2}}$$

Un ulteriore strumento fondamentale nella selezione e nell'analisi di un modello è il coefficiente di determinazione che esprime quantitativamente la proporzione di variabilità della  $Y$  spiegata dalle variabili indipendenti. Tale indice è il risultato del rapporto tra la somma dei quadrati della regressione e la somma totale dei quadrati:

$$r_{Y12}^2 = \frac{SQR}{SQT}$$

Tuttavia trattandosi di una regressione multipla occorre "aggiustare" il coefficiente di determinazione tenendo conto anche del numero di variabili esplicative incluse nel modello e dell'ampiezza del campione:

$$r_{adj}^2 = 1 - \left[ (1 - r_{Y12 \dots p}^2) \frac{n-1}{n-p-1} \right]$$

Dove,  $p$  = numero delle variabili indipendenti incluse nel modello e  $n$  = ampiezza del campione.

In molte situazioni si rende necessario anche l'utilizzo di variabili qualitative, le quali possono essere introdotte in un modello di regressione mediante l'utilizzo della variabili Dummy. Se la variabile categorica assume solo due valori, sarà necessario inserire una sola variabile dummy,  $X_d$ , così definita:

$X_d = 0$  se si osserva il primo valore

$X_d = 1$  se si osserva il secondo valore

D'altra parte, per poter utilizzare un modello costituito anche da variabile Dummy si deve verificare che i parametri corrispondenti alle altre variabili (esplicative) non dipendano dalla presenza o meno della variabile qualitativa. A tal riguardo si introducono tante nuove variabili quante sono quelle quantitative; ciascuna di esse viene denominata termine di interazione ed è ottenuta dal prodotto della  $j$ -esima variabile esplicativa  $X_j$  con la dummy  $X_d$ . In definitiva mediante un statistica test si accerta se queste nuove variabili danno un contributo significativo o meno attraverso due step:

- a) mediante il test  $F$  si accerta se i termini d'interazione, nel loro complesso, interagiscono con le altre variabili quantitative. In caso affermativo si deve scartare il modello considerato per effettuare delle previsioni.
- b) per mezzo del test  $t$  si valuta il contributo di ciascun termine di interazione separatamente per stabilire se dovrebbe essere incluso nel modello

Formalizzando quanto finora descritto, il modello di regressione con variabile Dummy ha la seguente struttura:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \beta_d X_{di} + \epsilon_i$$

dove

$\beta_0 = Y_i$  intercetta

$\beta_{1,2,\dots,p}$  = coefficiente relativo all'effetto lineare di  $X_{1,2,\dots,p}$  su  $Y_i$

$\beta_d$  = coefficiente addizionale dovuto alla variabile dummy su  $Y_i$

$\epsilon_i$  = l'errore casuale in  $Y_i$  corrispondente all' $i$ -esima osservazione

$X_{int}$  = termine d'interazione =  $X_{1,2,\dots,p} \times X_d$

### 2.3 Procedura di stima del modello statistico

Nella prima analisi l'obiettivo è stato individuare le variabili significative e dunque escludere quelle variabili che non lo sono. Per fare ciò si è utilizzato il software "MINITAB" considerando la funzione ANOVA → GENERAL LINEAR MODEL. Attraverso

questa funzione, che opera eseguendo test F parziali, è stato possibile individuare e dunque scartare quelle variabili che presentavano un p-value elevato, cioè non significative per il nostro modello. Il nostro modello di partenza prevedeva variabili quantitative quali prezzo, giorno, giorno<sup>2</sup>, giorno<sup>3</sup> e i fattoci climatici, ma anche variabili qualitative quali mese, giorno della settimana e la interazione tra mese e giorno. Si è proceduto per passi:

- eliminazione delle variabili con p-value superiore a 0,7.
- eliminazione delle variabili con p-value superiore a 0,3.
- eliminazione della variabile con p-value più alto.
- - iterazione del punto 3 fino ad avere tutte i p-value inferiori a 0,1.

Il motivo per il quale si sono effettuati più passaggi è dato dal fatto che, a mano a mano che vengono scartate delle variabili, i test sulle statistiche diventano più accurati. Per quanto riguarda le variabili giorno, giorno<sup>2</sup>, giorno<sup>3</sup>, non sono state modificate fino a quando tutte le altre variabili rimanenti non fossero risultate significative. A quel punto si verificava se il p-value di una delle tre variabili superava lo 0,1 e in caso affermativo si eliminava quella avente grado maggiore.

Per quanto riguarda il p-value, esso rappresenta la probabilità di osservare un valore della statistica test uguale o più estremo rispetto al valore campionario della statistica test quando l'ipotesi nulla è vera. Per stabilire se tra le variabili X e Y (nel nostro caso corrispondono a fattore climatico e vendita di giornali) esiste una relazione lineare significativa si può verificare se  $\beta$  (la pendenza della retta nella popolazione) è uguale a 0. L'ipotesi nulla e alternativa sono rispettivamente:

- $H_0: \beta = 0$  (non c'è relazione);
- $H_1: \beta \neq 0$  (c'è relazione).

Se il p-value è elevato significa che è elevata la probabilità che l'ipotesi nulla sia vera. Se si rifiuta l'ipotesi nulla, si può concludere che vi è la prova dell'esistenza della relazione tra le variabili considerate.

Al termine di questa fase abbiamo ottenuto per ogni edicola e area l'elenco delle variabili significative con i rispettivi coefficienti di regressione. Immediato è stato a quel punto costruire il modello di previsione. Nella seconda analisi l'obiettivo è stato quello di testare la significatività del modello sui 16 mesi (da gennaio 2010 a aprile 2011) mediante un controllo su  $R^2$ . L' $R^2$ , o coefficiente di determinazione, è una misura della bontà dell'adattamento della regressione lineare stimata ai dati osservati. L'elevato valore di  $R^2$  indica che vi è una forte relazione lineare tra le variabili considerate.

## 2.4 Il software statistico MINITAB

Minitab è un pacchetto statistico particolarmente adeguato per la didattica. Nonostante sia molto facile da utilizzare è piuttosto potente e flessibile per data set di dimensioni non eccessive. Le sessioni di lavoro di Minitab sono interattive, dapprima i dati vengono immessi o importati in un foglio di lavoro e poi vengono manipolati o analizzati mediante una serie di comandi interattivi. All'ingresso l'ambiente Minitab si presenta come in Figura 11.

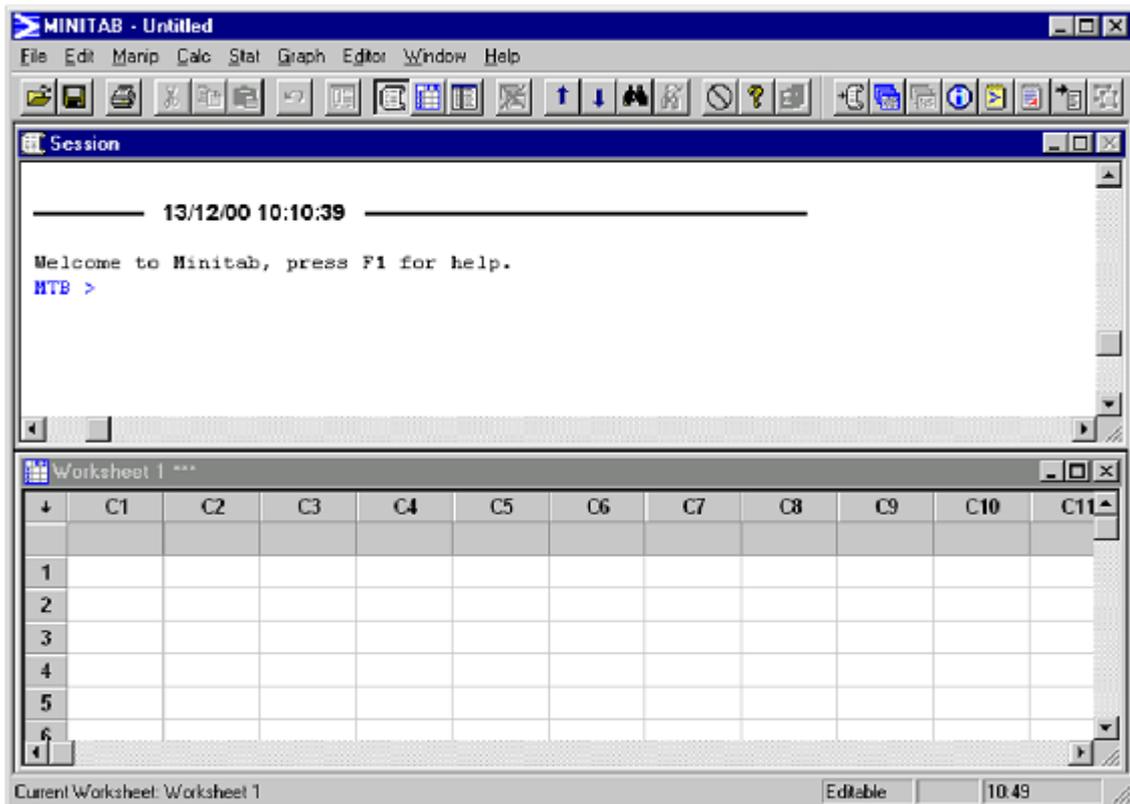


Figura 11. Schermata iniziale Minitab

Con Minitab si può memorizzare tutto ciò che riguarda un lavoro in un progetto che contiene: uno o più fogli di lavoro (Worksheets) contenenti dati; finestre di dati nel Worksheet; la finestra della sessione di lavoro con i risultati delle elaborazioni; le finestre grafiche illustranti i grafici ad alta risoluzione prodotti durante le elaborazioni e un gestore di progetti che tiene memoria di tutto il lavoro svolto. I principali usi di Minitab sono:

- Data e file management
- Analisi di regressione
- Analisi multivariate
- Test non parametrici e analisi varianza

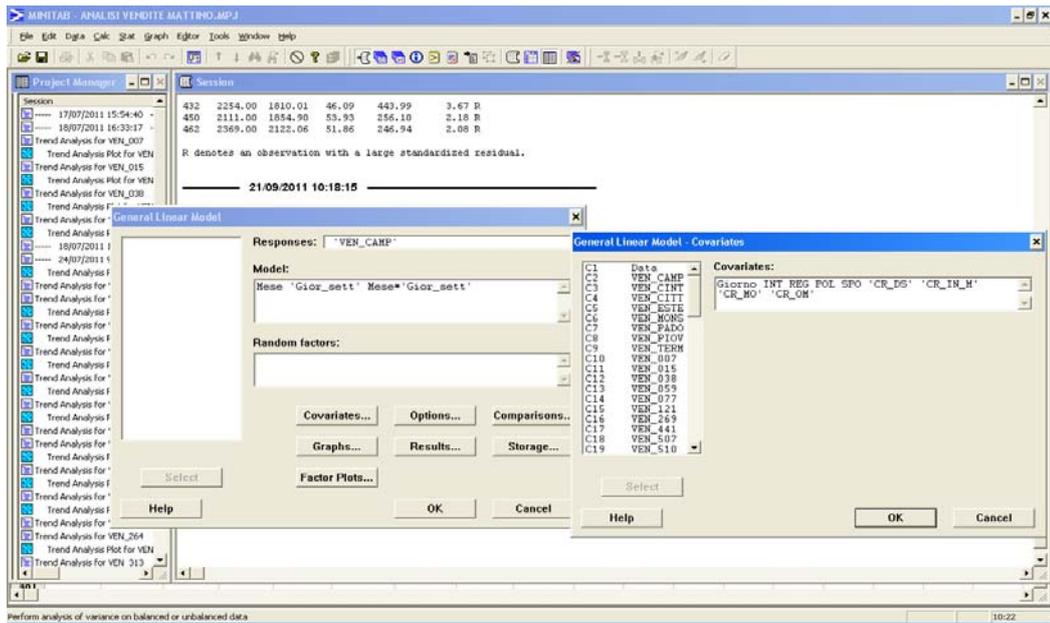


Figura 12. Processo di ricerca delle variabili significative



## CAPITOLO 3. Risultati e discussione

In questo capitolo vengono presentati i risultati ottenuti nell'analisi. In particolare, si evidenziano i grafici relativi all'andamento reale delle vendite durante l'intero arco di osservazione con il relativo il trend (dovuto alla variabile prezzo e al fattore ciclico) e quelli confrontanti l'andamento previsto con quello effettivo.

### 3.1 Serie temporali delle vendite ed analisi del trend

Sulla base dei dati di vendita nei 18 mesi, per ogni area e edicola si sono costruiti i trend dovuti solo alla variabile prezzo e al fattore ciclico.

Questi sono i trend delle 8 aree:

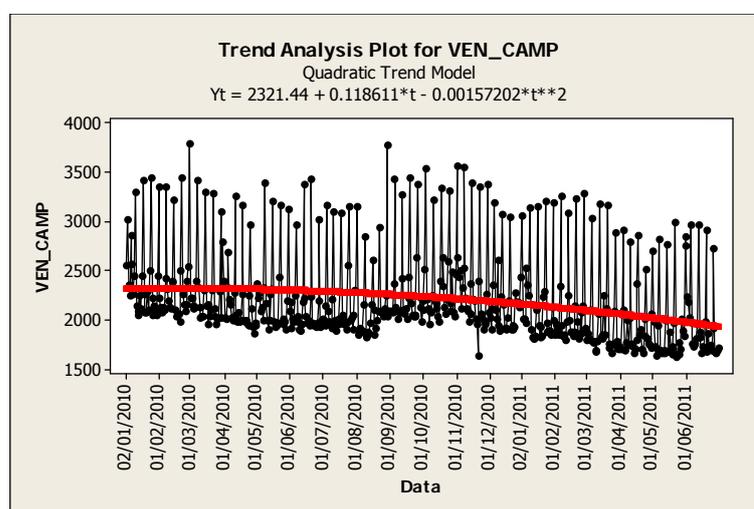


Figura 13. Serie temporale e trend analysis per le vendite dell'area Camposampierense

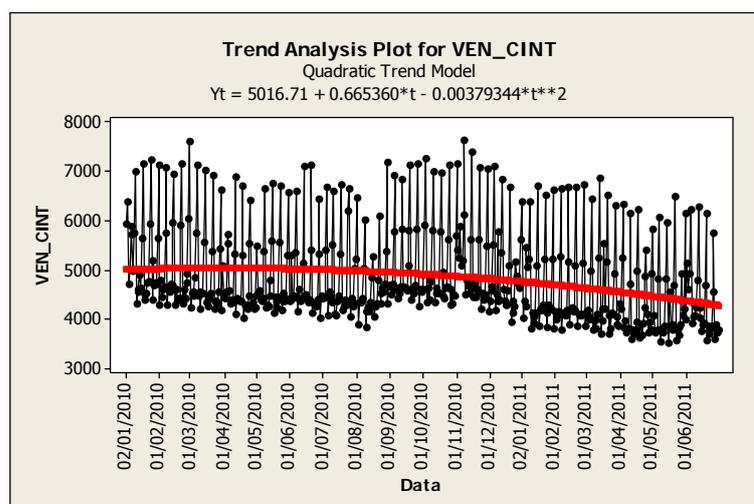


Figura 14. Serie temporale e trend analysis per le vendite della Cintura urbana

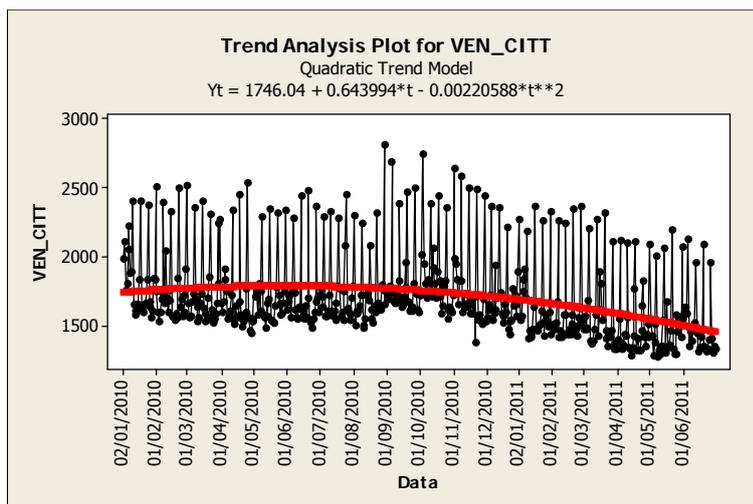


Figura 15. Serie temporale e trend analysis per le vendite dell'area Cittadelense

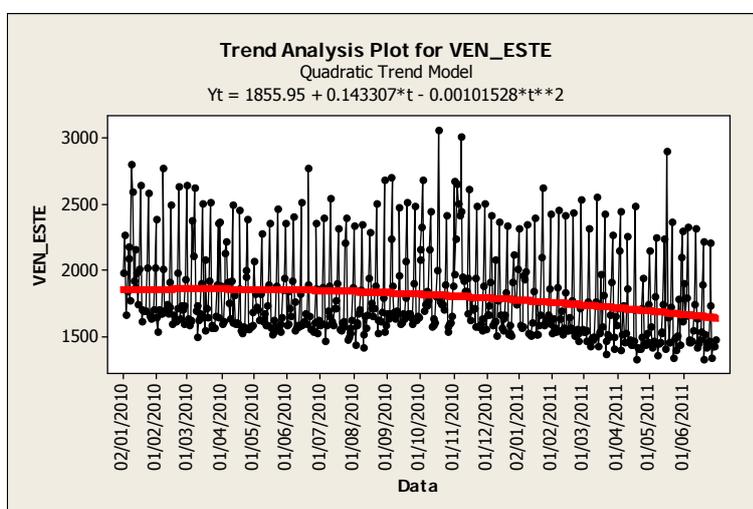


Figura 16. Serie temporale e trend analysis per le vendite dell'area Estense

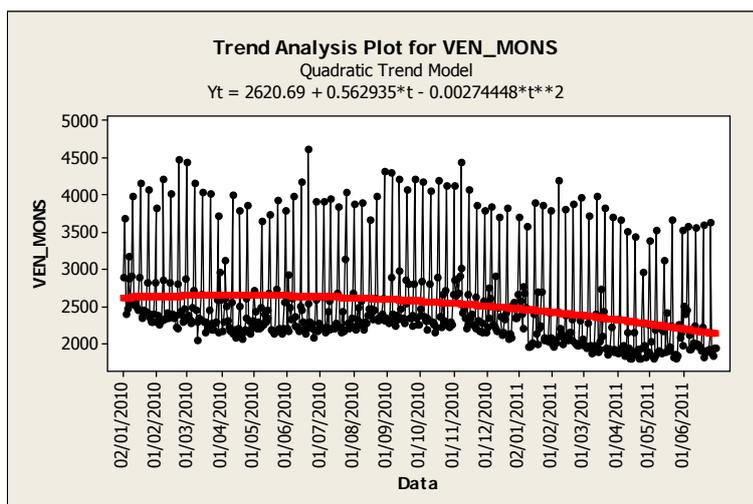


Figura 17. Serie temporale e trend analysis per le vendite dell'area Monselicense

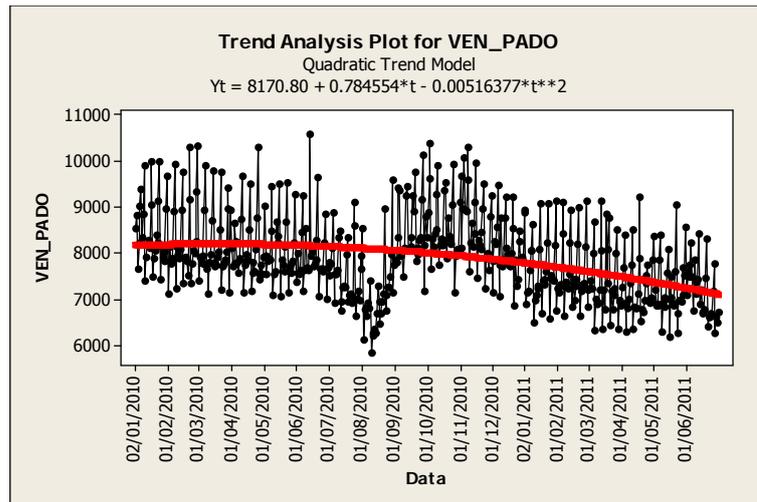


Figura 18. Serie temporale e trend analysis per le vendite di Padova

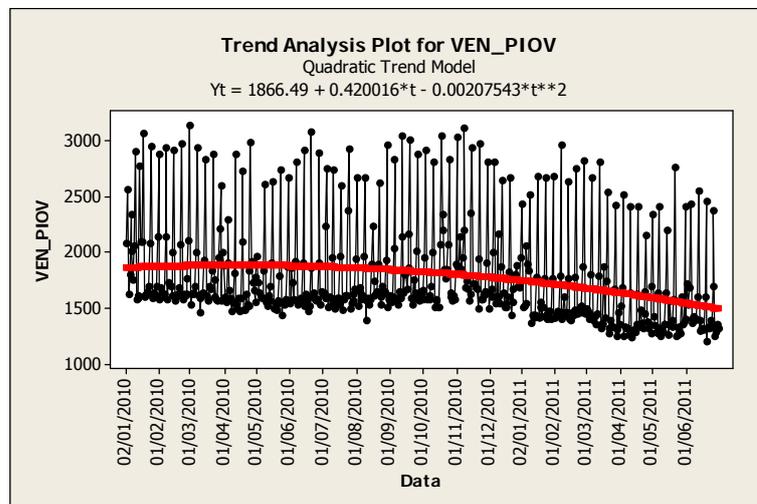


Figura 19. Serie temporale e trend analysis per le vendite di Piovese

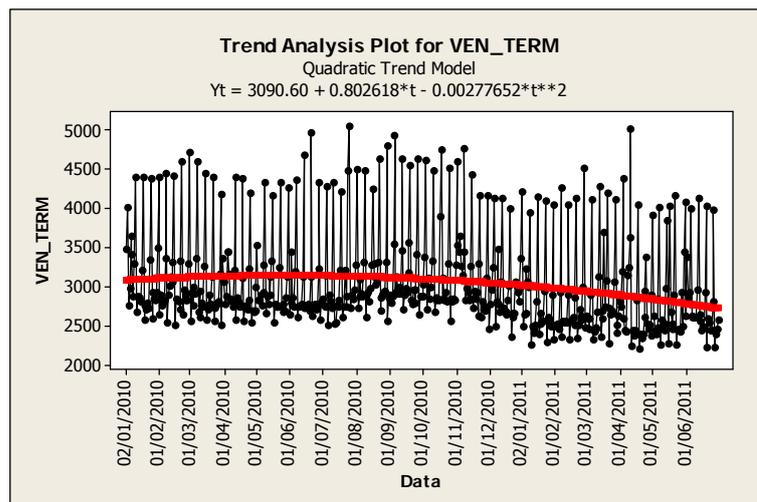


Figura 20. Serie temporale e trend analysis per le vendite di Terme Euganee

Questi sono i trend delle edicole:

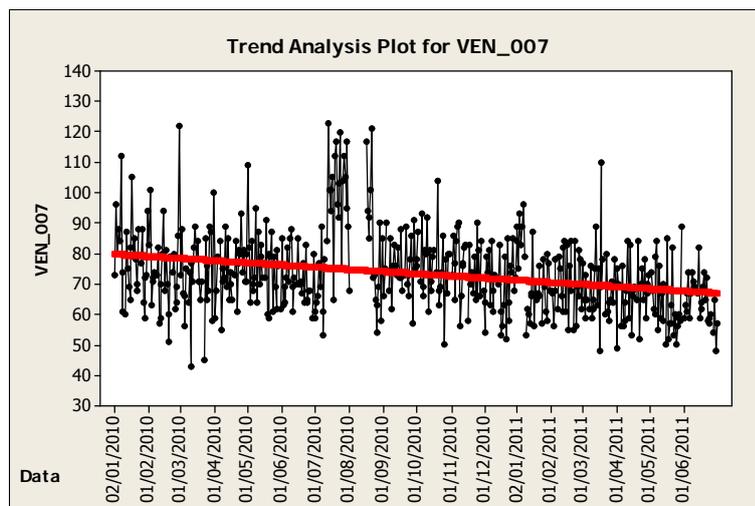


Figura 21. Serie temporale e trend analysis per le vendite dell'edicola 007

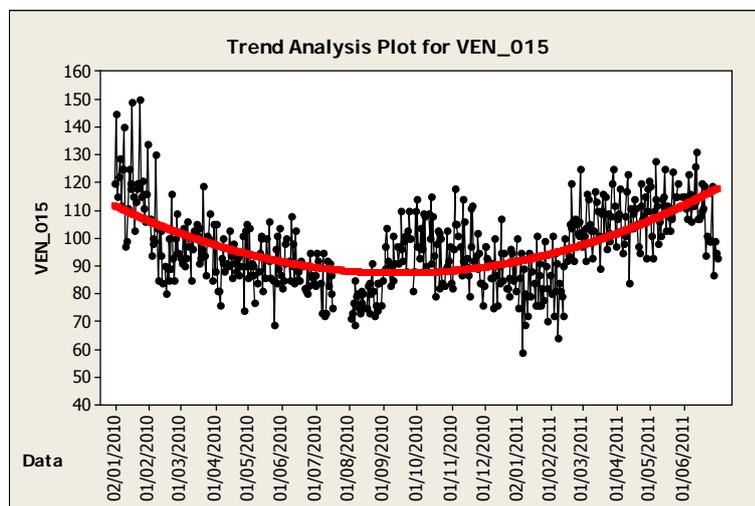


Figura 22. Serie temporale e trend analysis per le vendite dell'edicola 015

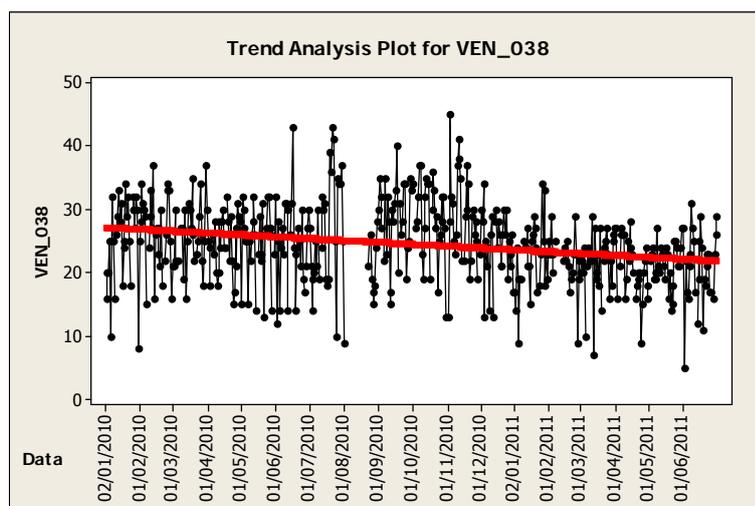


Figura 23. Serie temporale e trend analysis per le vendite dell'edicola 038

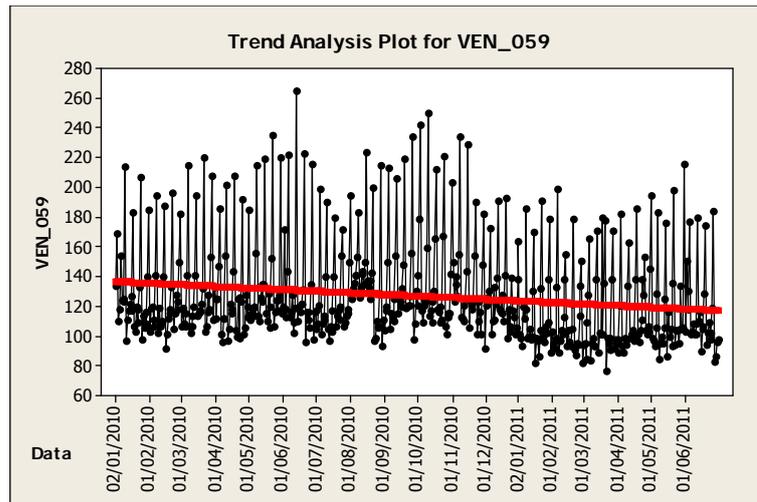


Figura 24. Serie temporale e trend analysis per le vendite dell'edicola 059

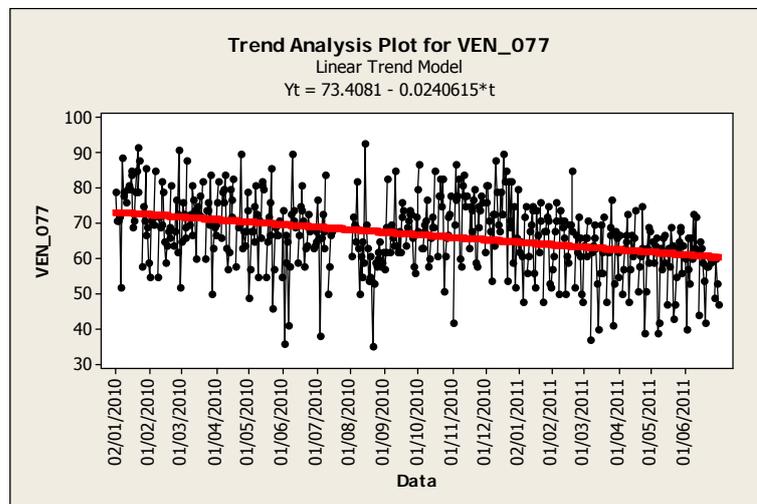


Figura 25. Serie temporale e trend analysis per le vendite dell'edicola 077

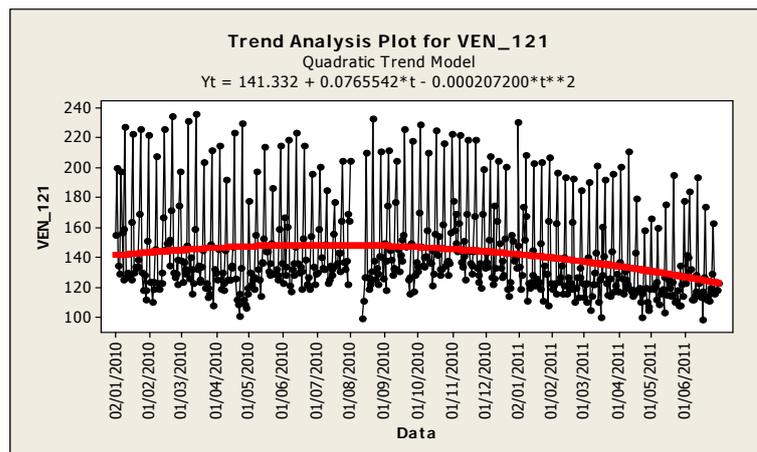


Figura 26. Serie temporale e trend analysis per le vendite dell'edicola 121

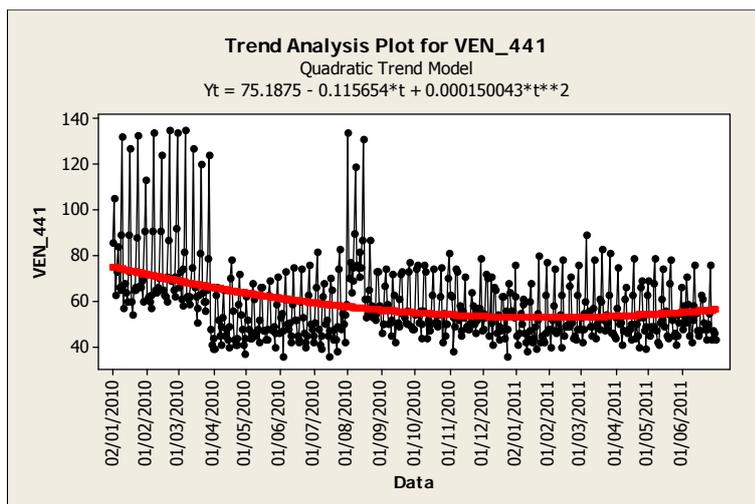


Figura 27. Serie temporale e trend analysis per le vendite dell'edicola 441

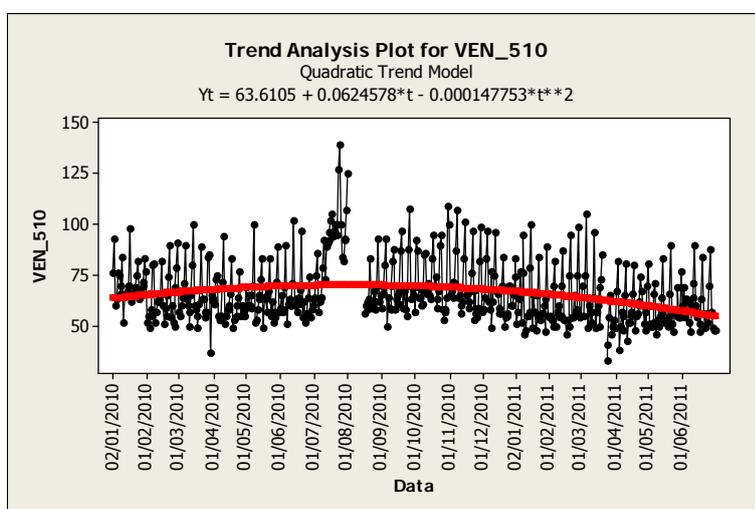


Figura 28. Serie temporale e trend analysis per le vendite dell'edicola 510

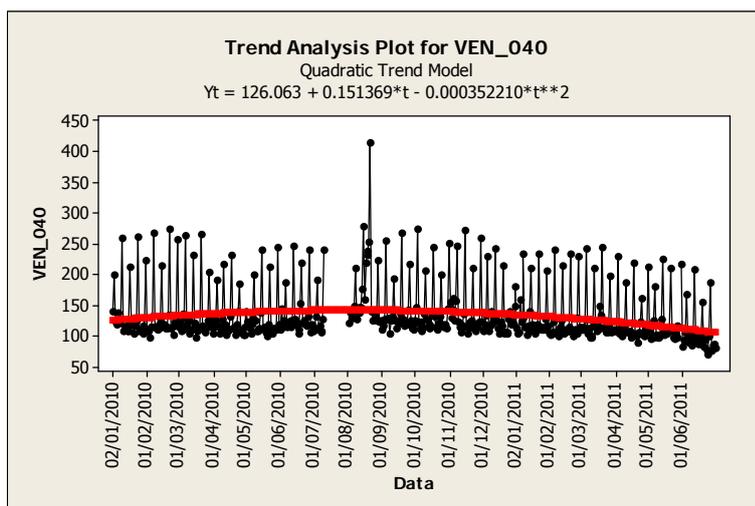


Figura 29. Serie temporale e trend analysis per le vendite dell'edicola 040.

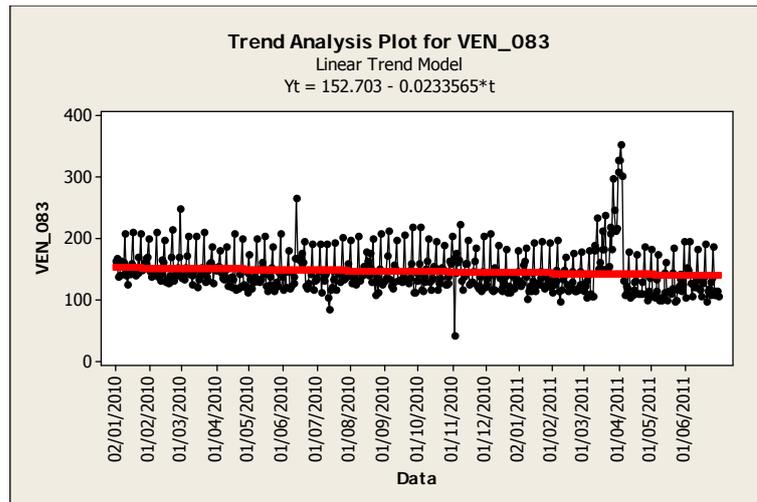


Figura 30. Serie temporale e trend analysis per le vendite dell'edicola 083

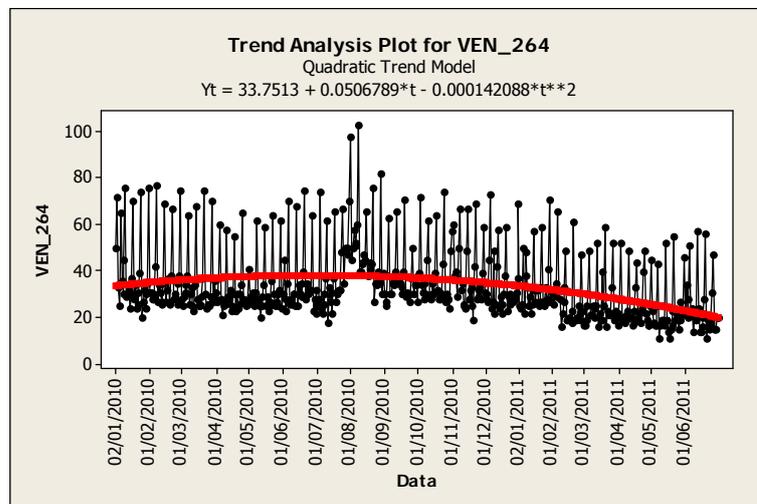


Figura 31. Serie temporale e trend analysis per le vendite dell'edicola 264

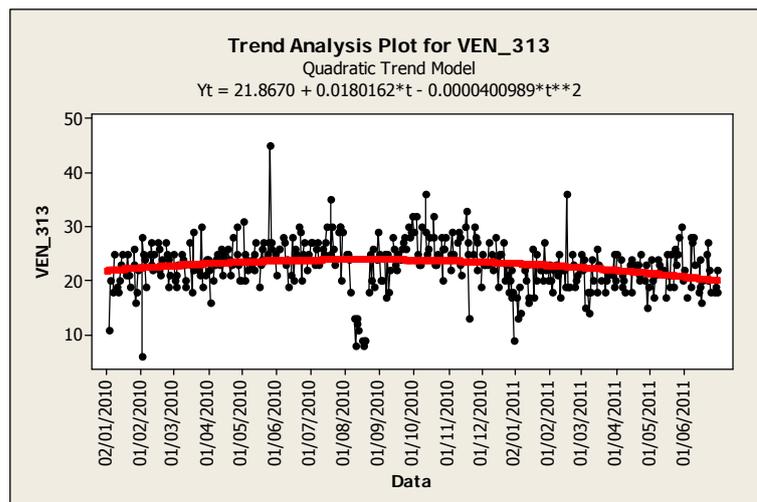


Figura 32. Serie temporale e trend analysis per le vendite dell'edicola 313

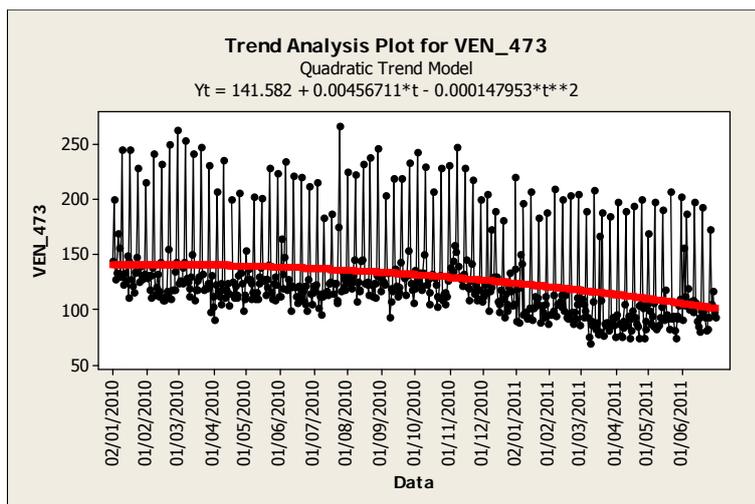


Figura 33. Serie temporale e trend analysis per le vendite dell'edicola 473

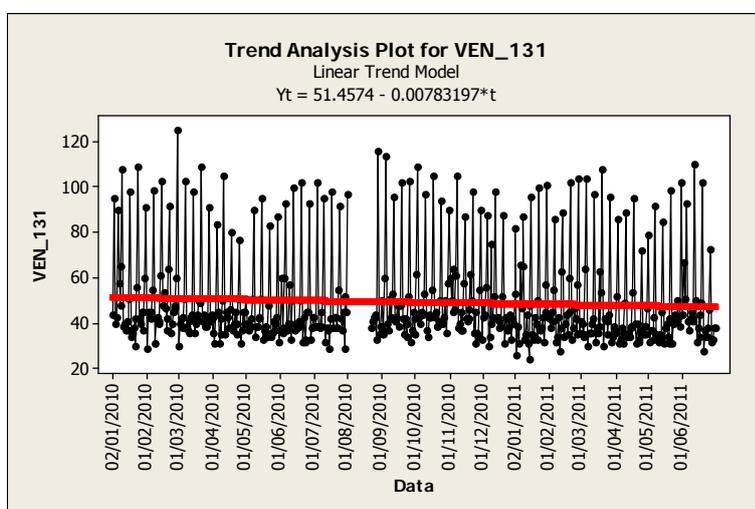


Figura 34. Serie temporale e trend analysis per le vendite dell'edicola 131

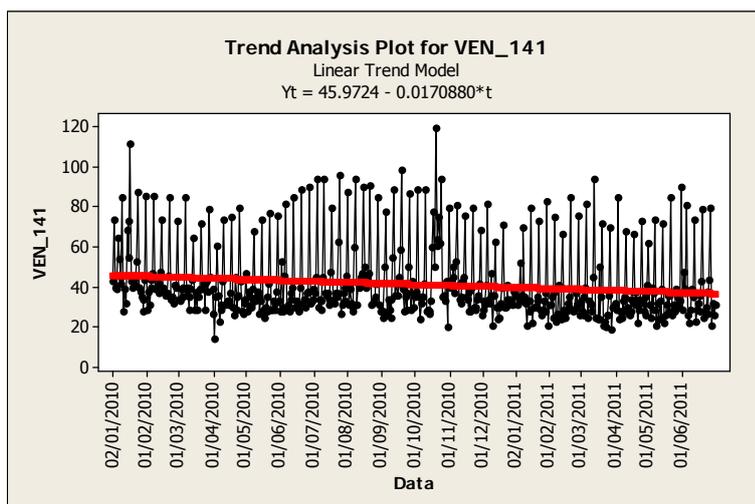


Figura 35. Serie temporale e trend analysis per le vendite dell'edicola 141

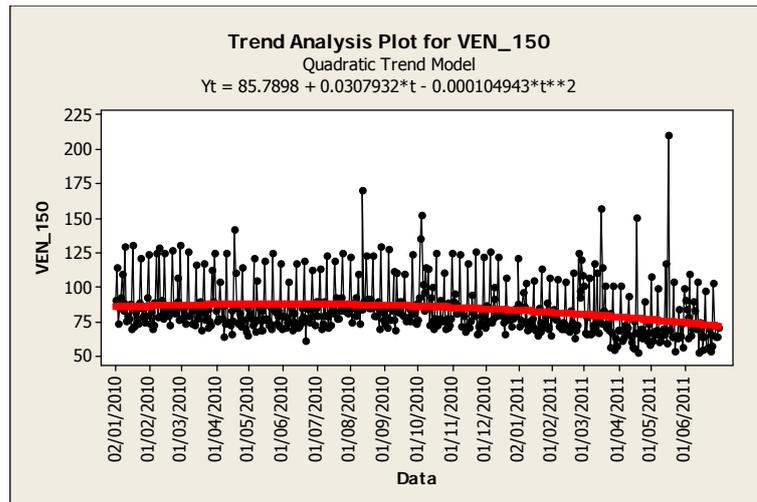


Figura 36. Serie temporale e trend analysis per le vendite dell'edicola 150

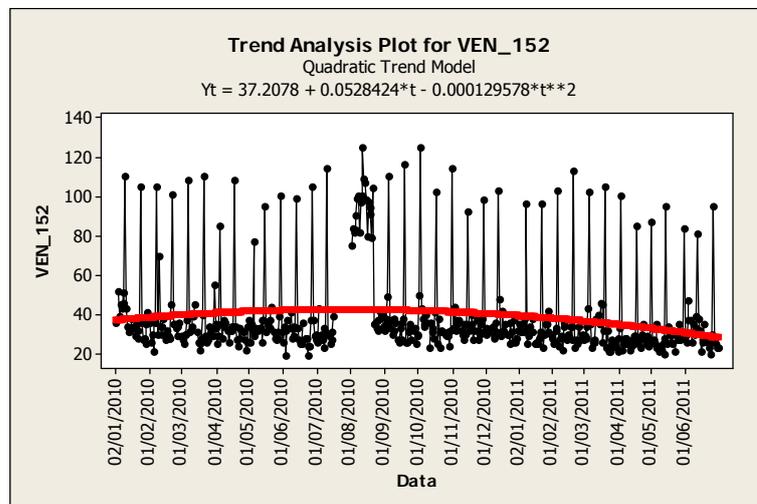


Figura 37. Serie temporale e trend analysis per le vendite dell'edicola 152

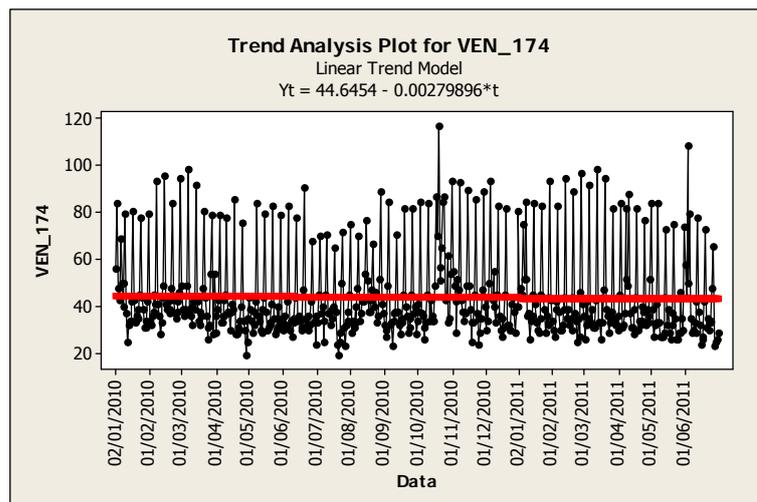


Figura 38. Serie temporale e trend analysis per le vendite dell'edicola 174

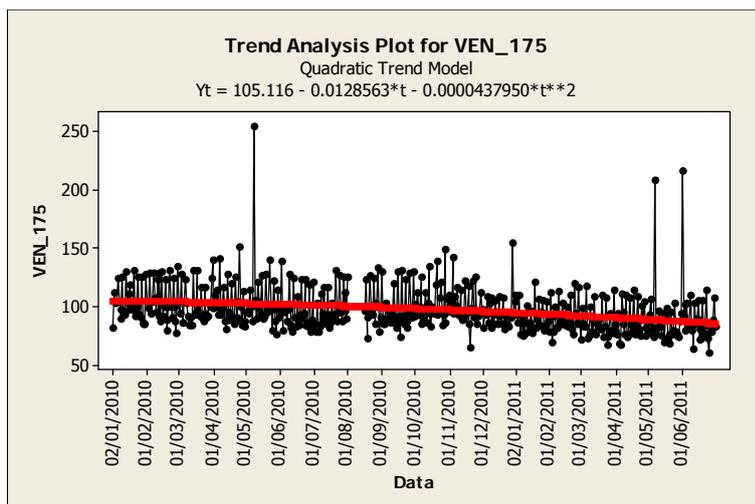


Figura 39. Serie temporale e trend analysis per le vendite dell'edicola 175

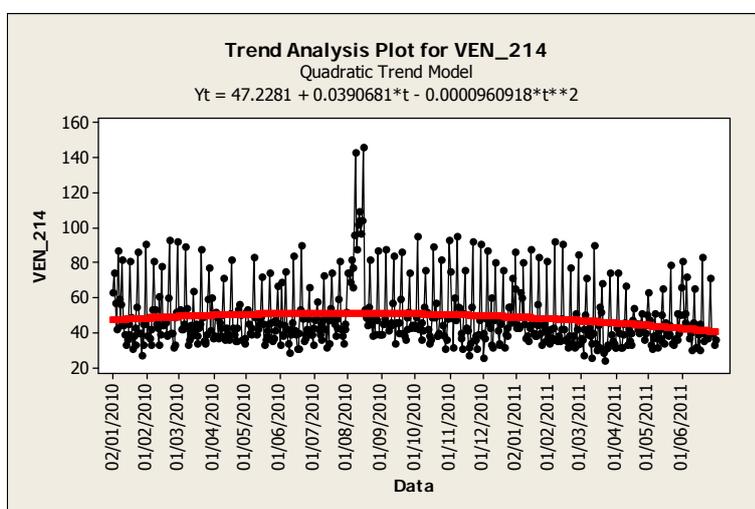


Figura 40. Serie temporale e trend analysis per le vendite dell'edicola 214

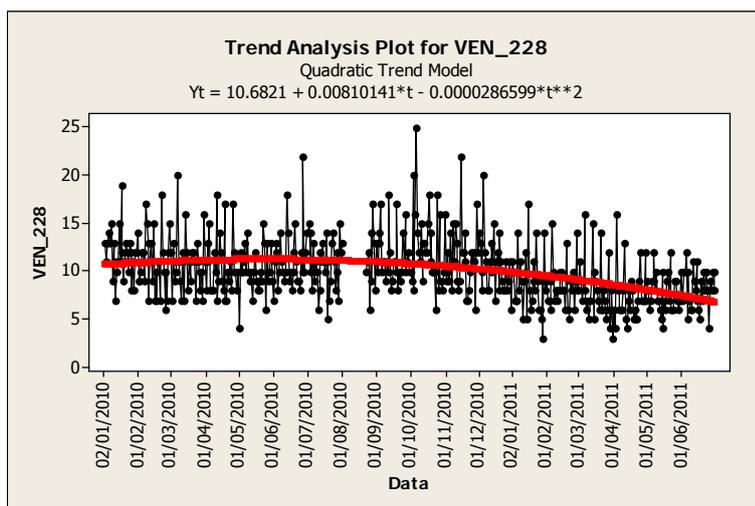


Figura 41. Serie temporale e trend analysis per le vendite dell'edicola 228

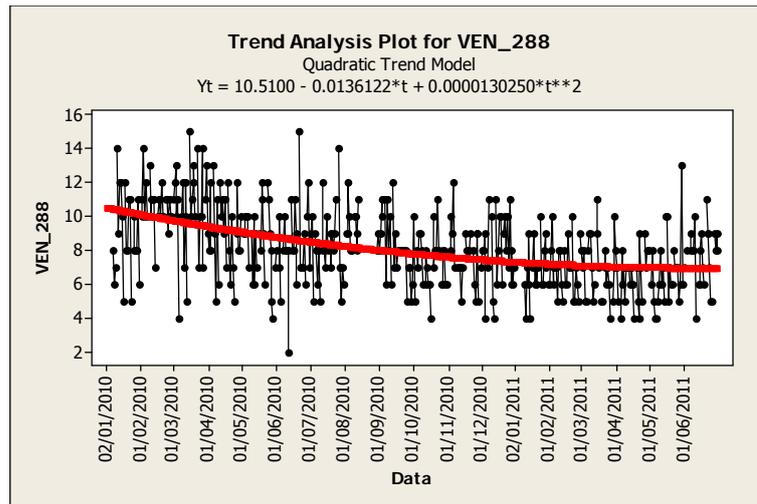


Figura 42. Serie temporale e trend analysis per le vendite dell'edicola 288

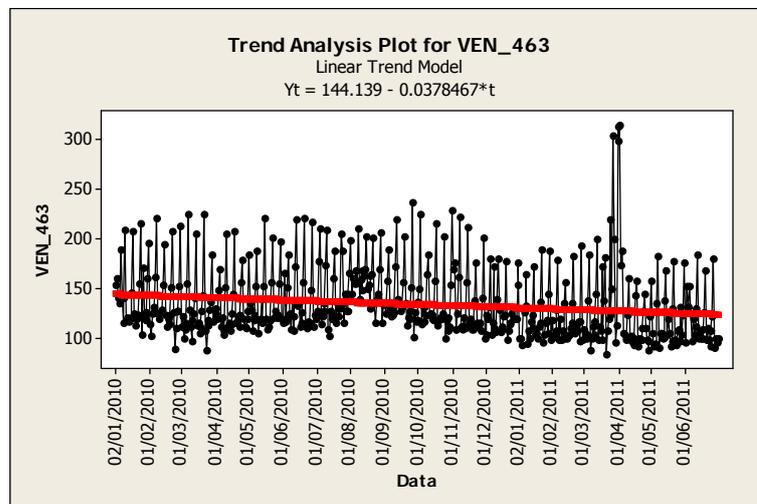


Figura 43. Serie temporale e trend analysis per le vendite dell'edicola 463

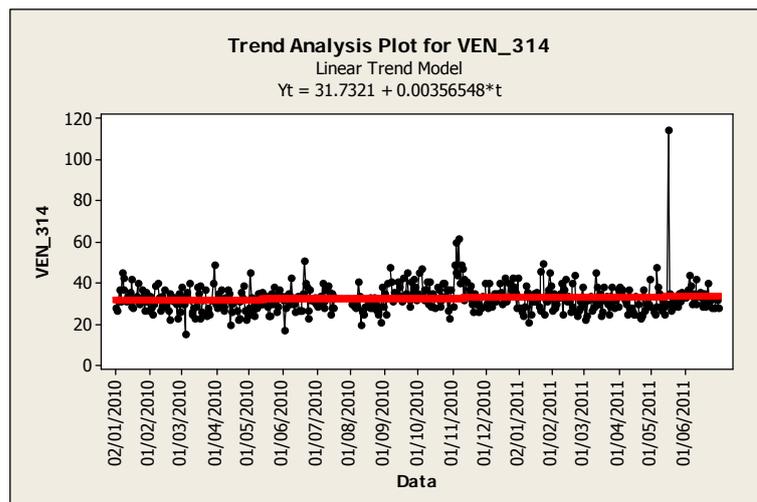


Figura 44. Serie temporale e trend analysis per le vendite dell'edicola 314

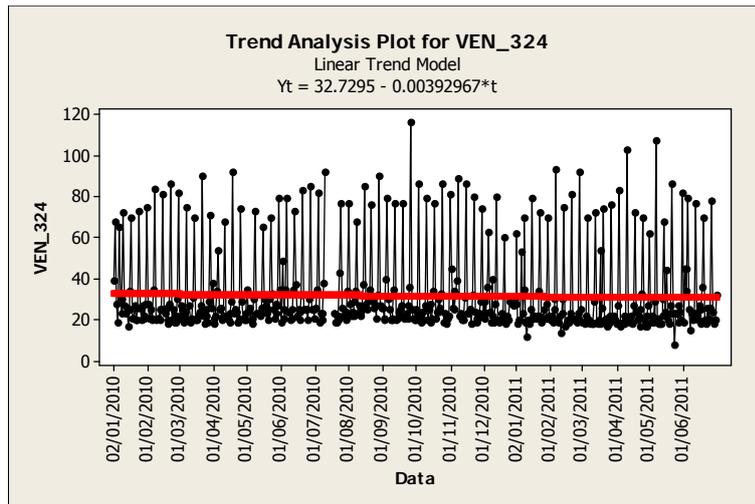


Figura 45. Serie temporale e trend analysis per le vendite dell'edicola 324

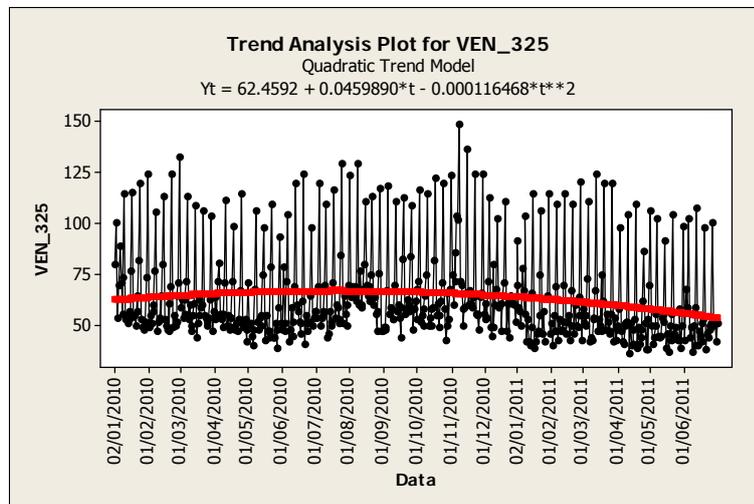


Figura 46. Serie temporale e trend analysis per le vendite dell'edicola 325

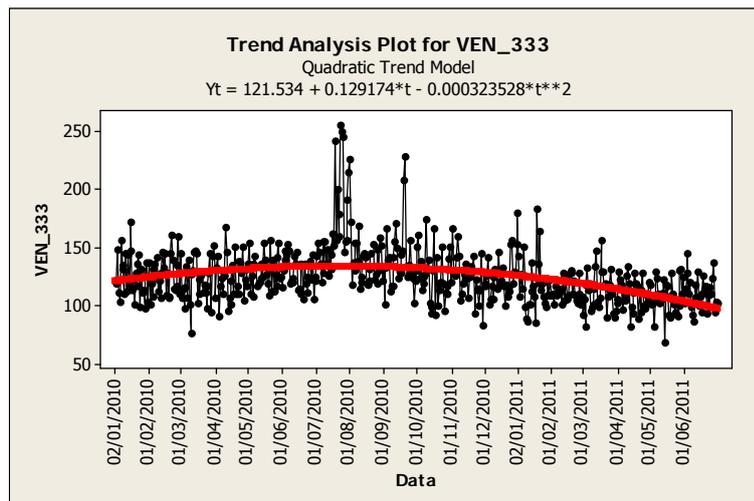


Figura 47. Serie temporale e trend analysis per le vendite dell'edicola 333

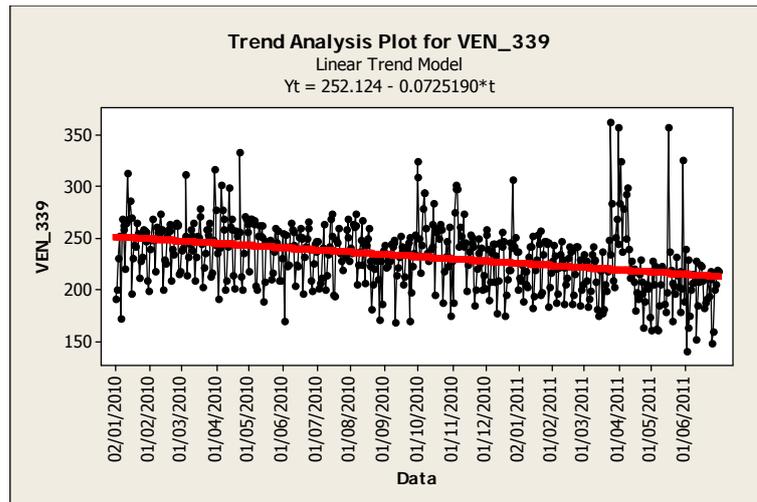


Figura 48. Serie temporale e trend analysis per le vendite dell'edicola 339

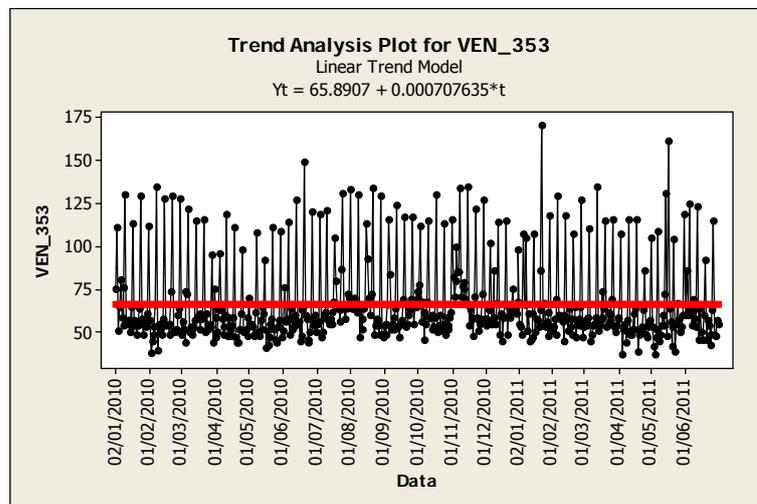


Figura 49. Serie temporale e trend analysis per le vendite dell'edicola 353

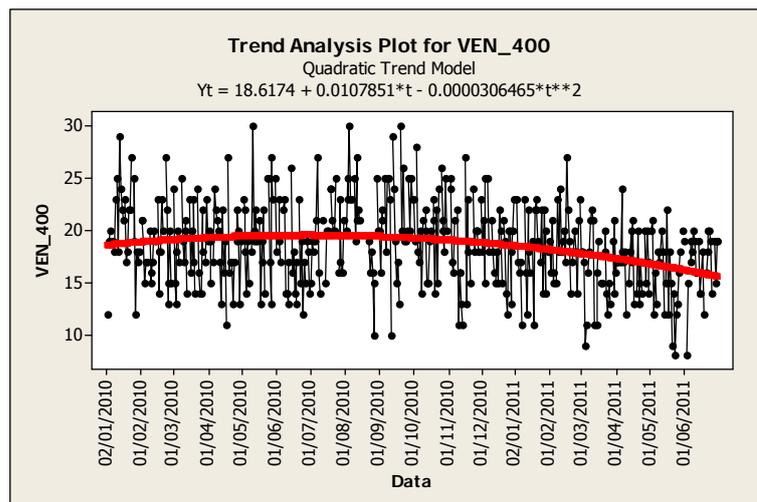


Figura 50. Serie temporale e trend analysis per le vendite dell'edicola 400

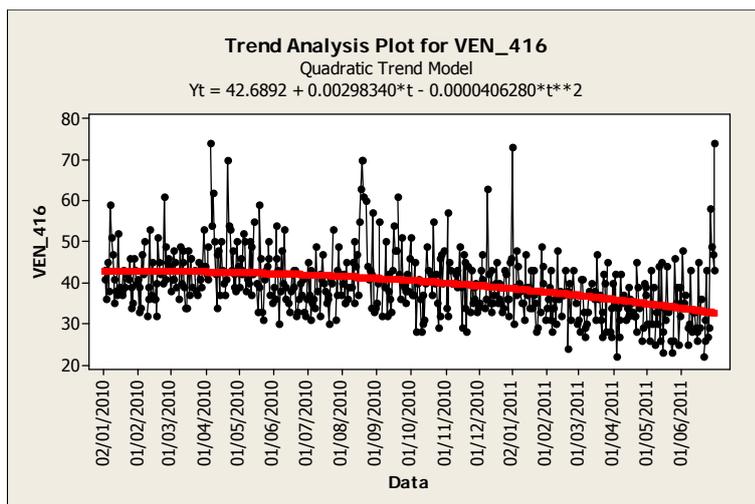


Figura 51. Serie temporale e trend analysis per le vendite dell'edicola 416

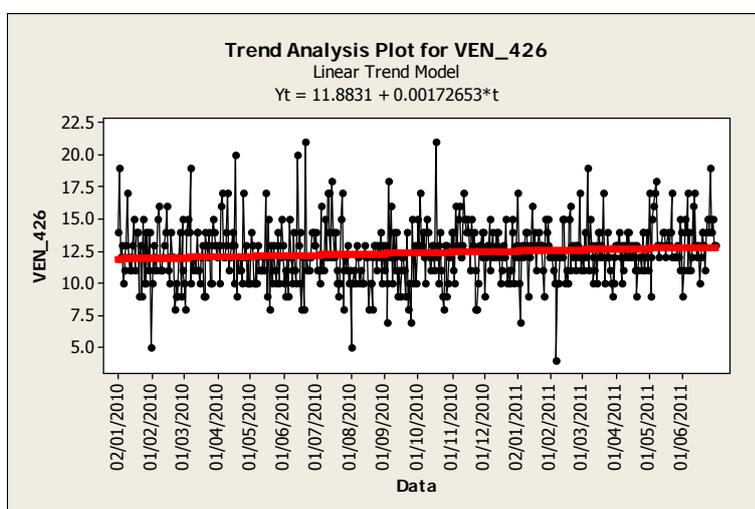


Figura 52. Serie temporale e trend analysis per le vendite dell'edicola 426

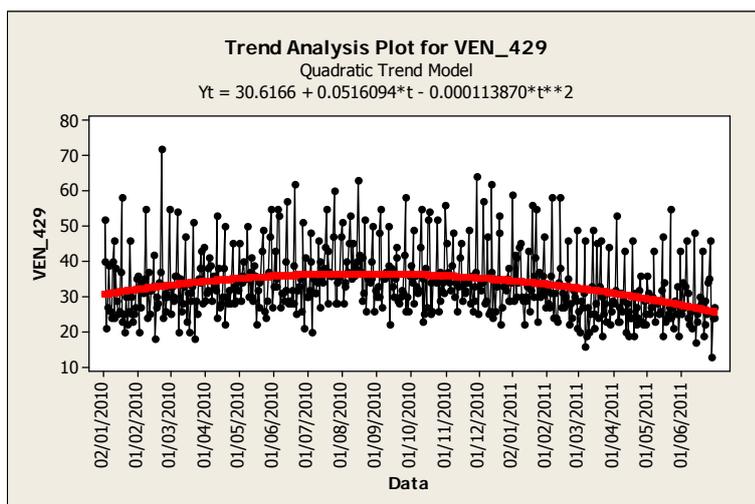


Figura 53. Serie temporale e trend analysis per le vendite dell'edicola 429

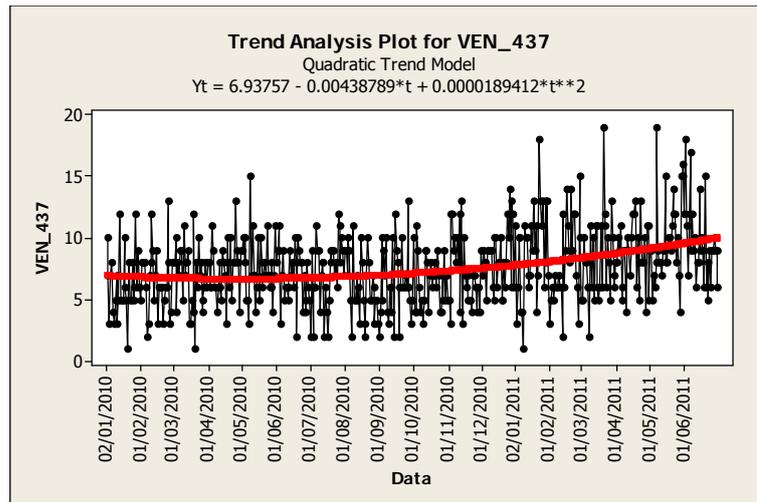


Figura 54. Serie temporale e trend analysis per le vendite dell'edicola 437

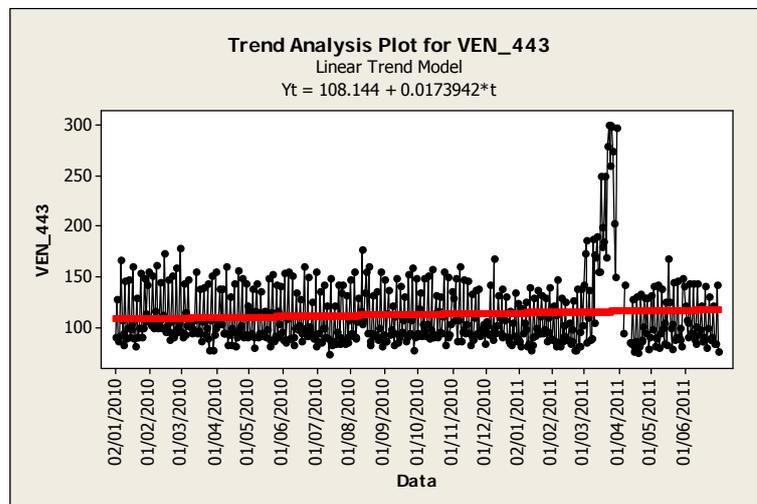


Figura 55. Serie temporale e trend analysis per le vendite dell'edicola 443

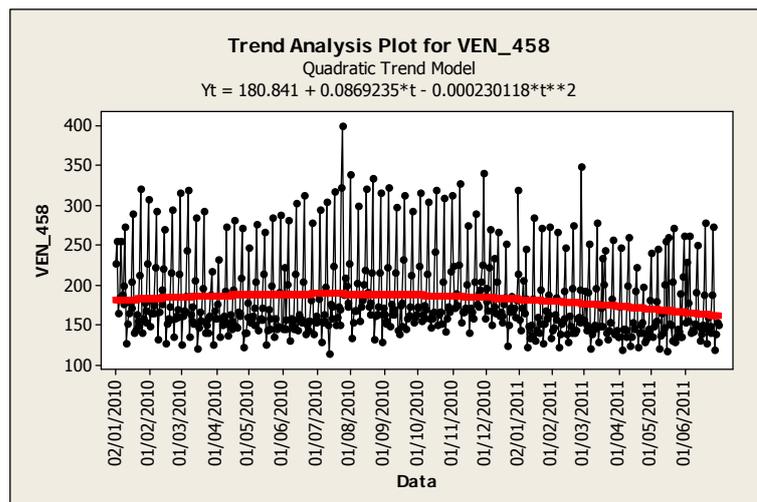
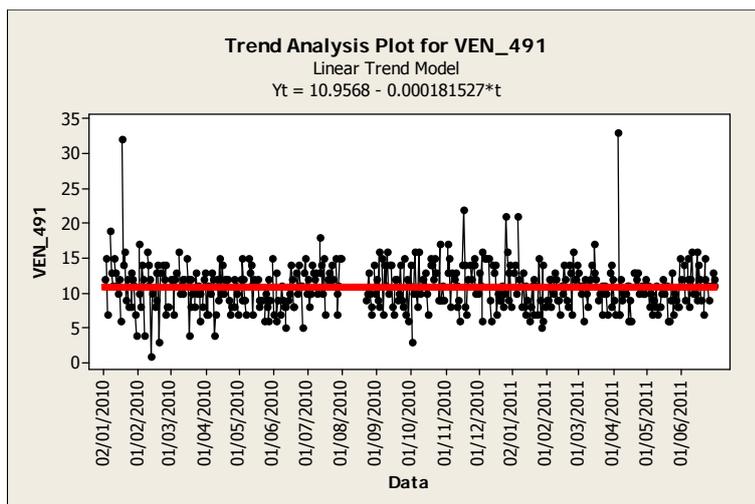


Figura 56. Serie temporale e trend analysis per le vendite dell'edicola 458



*Figura 57. Serie temporale e trend analysis per le vendite dell'edicola 491*

Come si può osservare, la componente sistematica tendenziale delle aree e delle edicole considerate è lineare (sul grafico appare come una retta) oppure quadratico (sul grafico ha l'andamento di una parabola). Nella maggior parte dei casi il trend è decrescente, andamento oggettivamente attribuibile all'aumento del prezzo del giornale a partire dagli inizi del 2011. In alcuni casi isolati, invece, il trend è crescente, anche se la pendenza positiva è minima.

### 3.2 Analisi statistica a 18 mesi

Nell'analisi statistica a 18 mesi l'obiettivo è stato quello di individuare le variabili significative. Attraverso l'utilizzo del software MINITAB e dunque della funzione "ANOVA: GENERAL LINEAR MODEL" è stato possibile scartare quelle variabili che presentavano un p-value elevato ( $>0.7$ , poi  $>0.3$  e infine  $>0.1$ ), ovvero variabili non significative per il nostro modello. Al termine di questa fase abbiamo ottenuto per ogni edicola e area l'elenco della variabili significative con i rispettivi coefficienti di regressione.

Luogo_Var_Meteo	Precip	Rad_sol	Temp_aria_min	Temp_aria_med	Temp_aria_max	Umid_rel_min	Umid_rel_max
VEN_CAMP	-	-	9,23	-12,03	-	-	-
VEN_CINT	-	-	29,58	-21,48	-	-	-
VEN_CITT	2,05	-	-	-	-	-	-
VEN_ESTE	-	-	9,42	-	-	-	-
VEN_MONS	-	-	6,27	-	-	-	-
VEN_PADO	-4,75	-	42,01	-63,57	36,32	-	-
VEN_PIOV	-	-	4,92	-	-	-	-
VEN_TERM	-	-	-	7,84	-	-	-
VEN_007	-0,15	-	-	-	-	-	-
VEN_038	-	-	-	-	-	-	-
VEN_077	-	-	-0,77	-	0,46	0,08	-
VEN_441	-	-	-	-	-	-	-
VEN_510	-	-	-	-	-0,25	-	-
VEN_040	-	-	0,57	-	-	-	-
VEN_264	-	-0,09	-	-	-	-	-

Figura 58. Tabella dei coefficienti di regressione

Luogo_Var_Meteo	Precip	Rad_sol	Temp_aria_min	Temp_aria_med	Temp_aria_max	Umid_rel_min	Umid_rel_max
VEN_313	-	-	0,11	-	-	-	-
VEN_131	-	-	0,26	-	-	-	-
VEN_150	-	-	-	0,90	-0,68	-	-
VEN_174	-	-	-	-	-0,24	-	0,12
VEN_214	-	-	0,31	-	-0,35	-	-
VEN_288	-	-	-	-	-	-	-
VEN_314	-	-	-	-	-	-	-
VEN_325	-	-	-	0,61	-0,55	-	-
VEN_339	-	-	-	1,47	-	-	-
VEN_400	0,08	-	-	-	-	-	-
VEN_426	-	-	-	-	-	-	-
VEN_437	-	-	-	-	-	-	-
VEN_458	-	-	-	-	-1,05	-	-

Legenda: - = coefficiente non significativo

Figura 59. Tabella dei coefficienti di regressione

In questa tabella sono presenti sulle colonne tutte le variabili meteorologiche considerate e sulle righe ogni area e edicola esaminata. Se si riscontra un coefficiente negativo significa che tale fattore climatico influenza in modo negativo la vendita dei quotidiani.

### 3.3 Analisi statistica previsionale

Una volta costruito il modello, lo si utilizza per effettuare delle previsioni di vendita. Per valutarne la validità, si costruiscono dei grafici che mostrano per ogni edicola e area il trend delle vendite, l'andamento previsto e l'andamento reale.

Questi sono i grafici ottenuti per le aree:

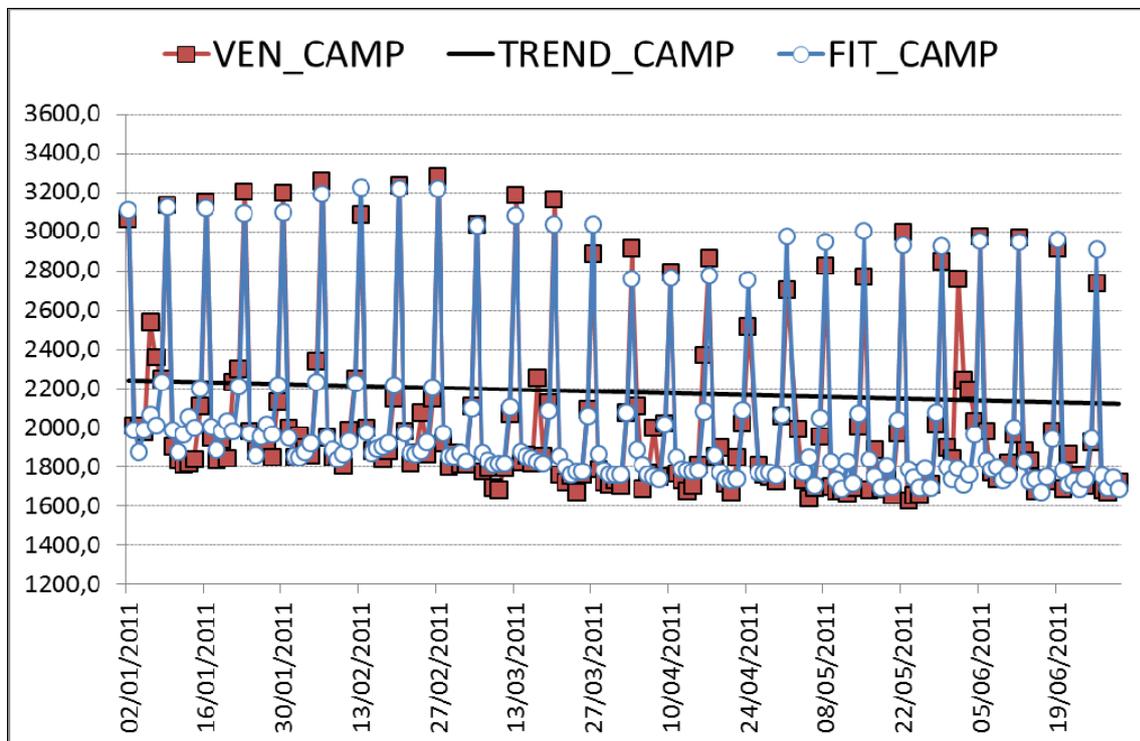


Figura 60. Vendite reali, trend e vendite previste per l'area Camposampierese

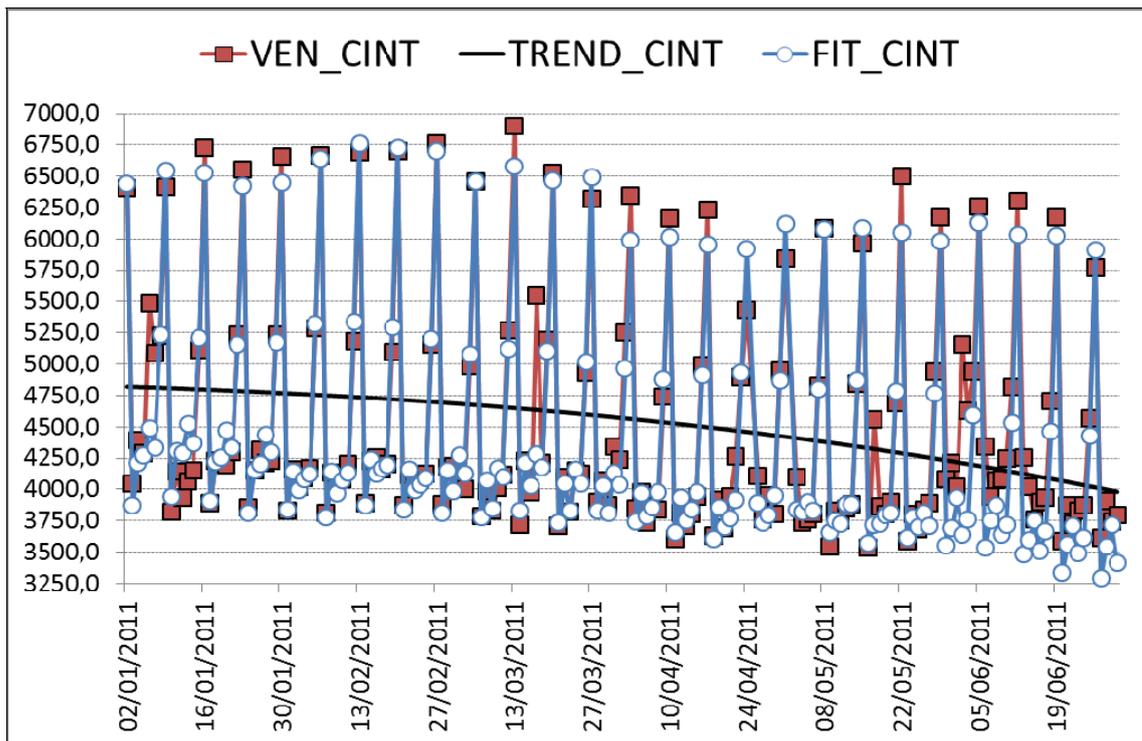


Figura 61. Vendite reali, trend e vendite previste per la Cintura Urbana

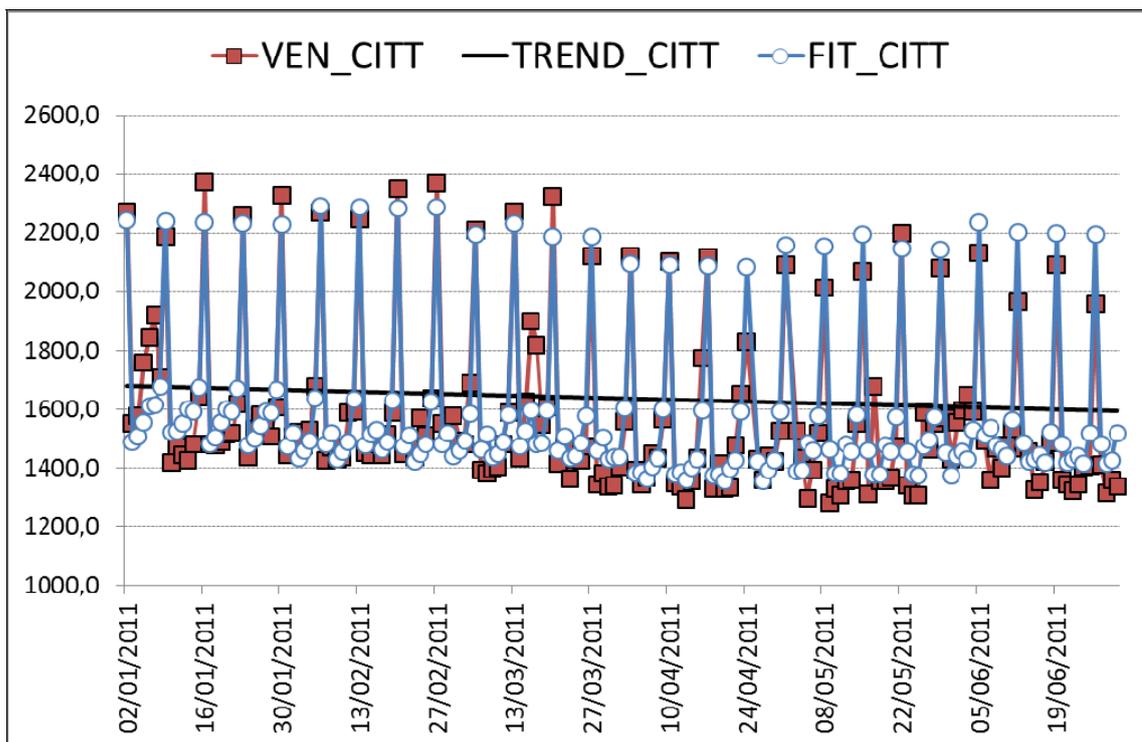


Figura 62. Vendite reali, trend e vendite previste per l'area Cittadelense

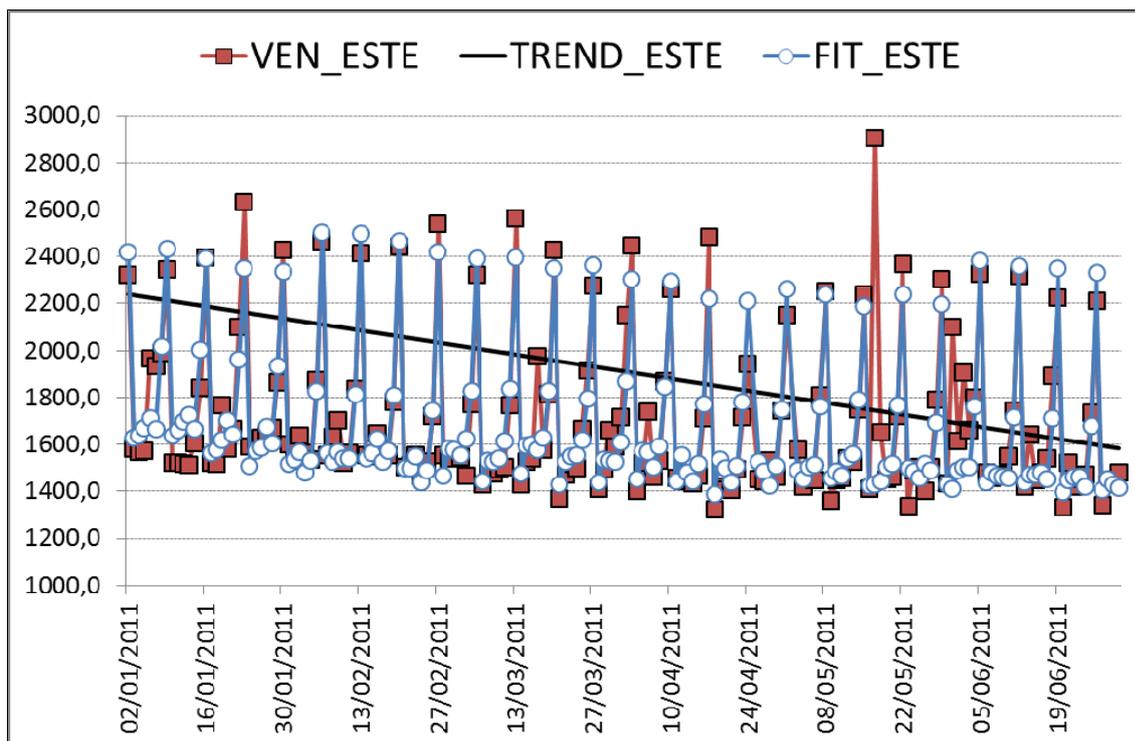


Figura 63. Vendite reali, trend e vendite previste per l'area Estense

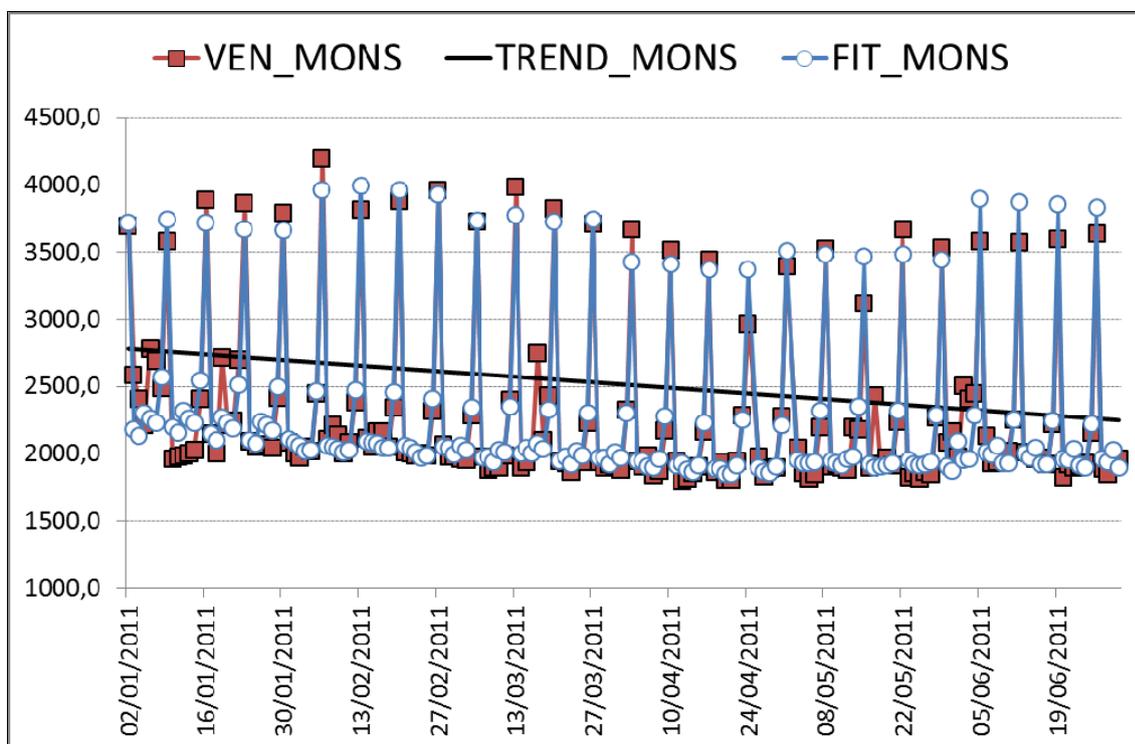


Figura 64. Vendite reali, trend e vendite previste per l'area Monselicense

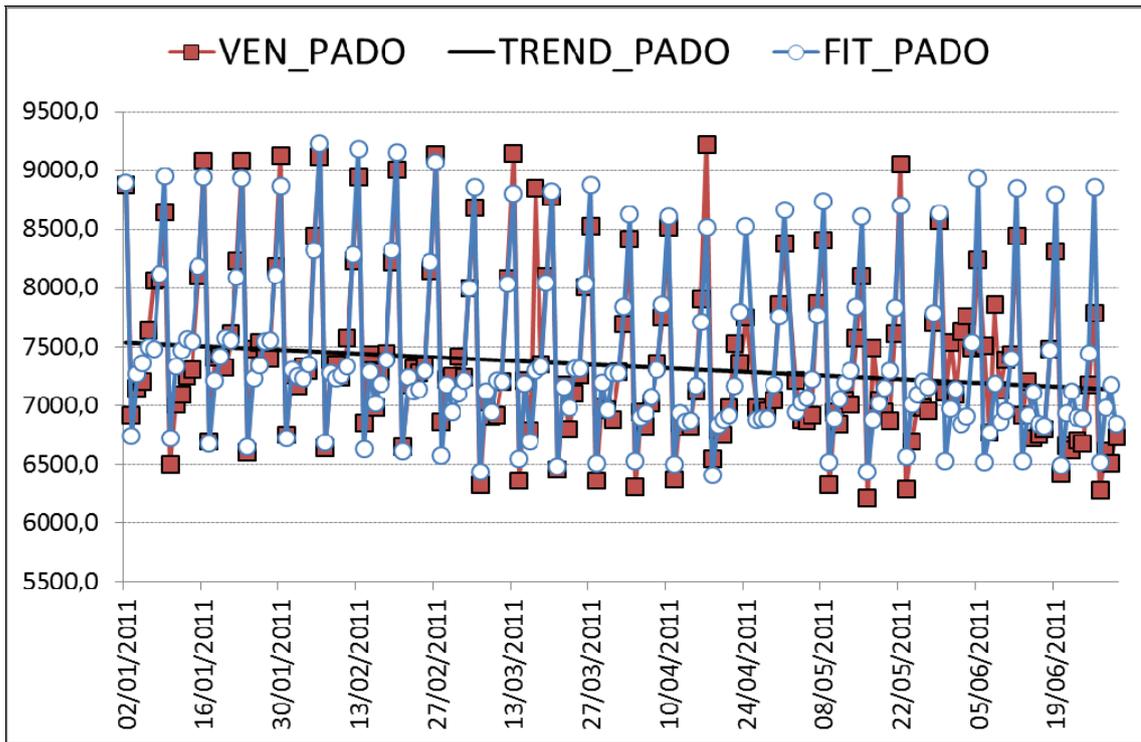


Figura 65. Vendite reali, trend e vendite previste per Padova

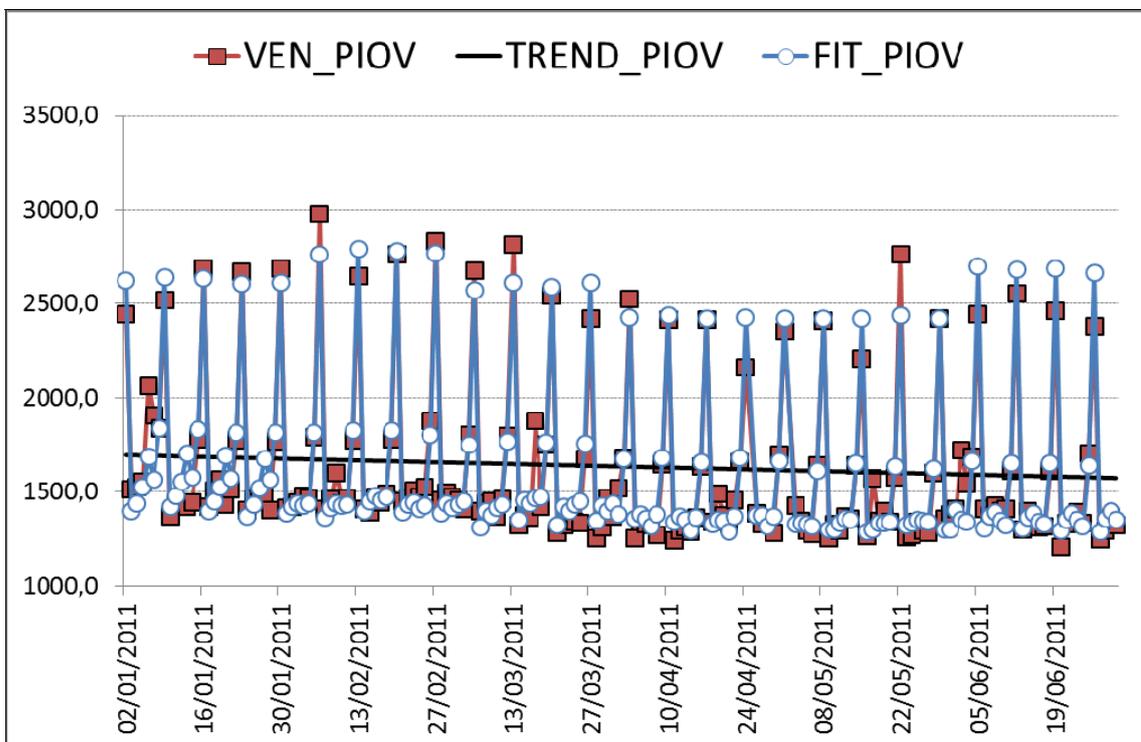


Figura 66. Vendite reali, trend e vendite previste per Piovese

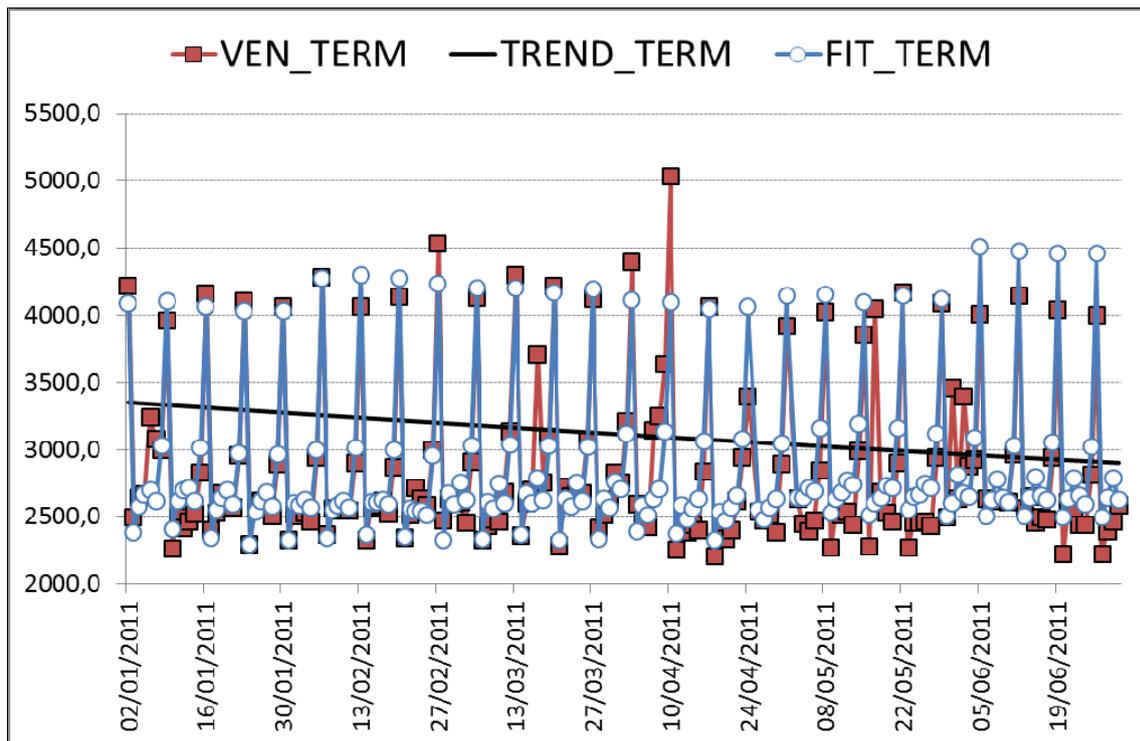


Figura 67. Vendite reali, trend e vendite previste per Terme Euganee

Questi sono i grafici ottenuti per le edicole:

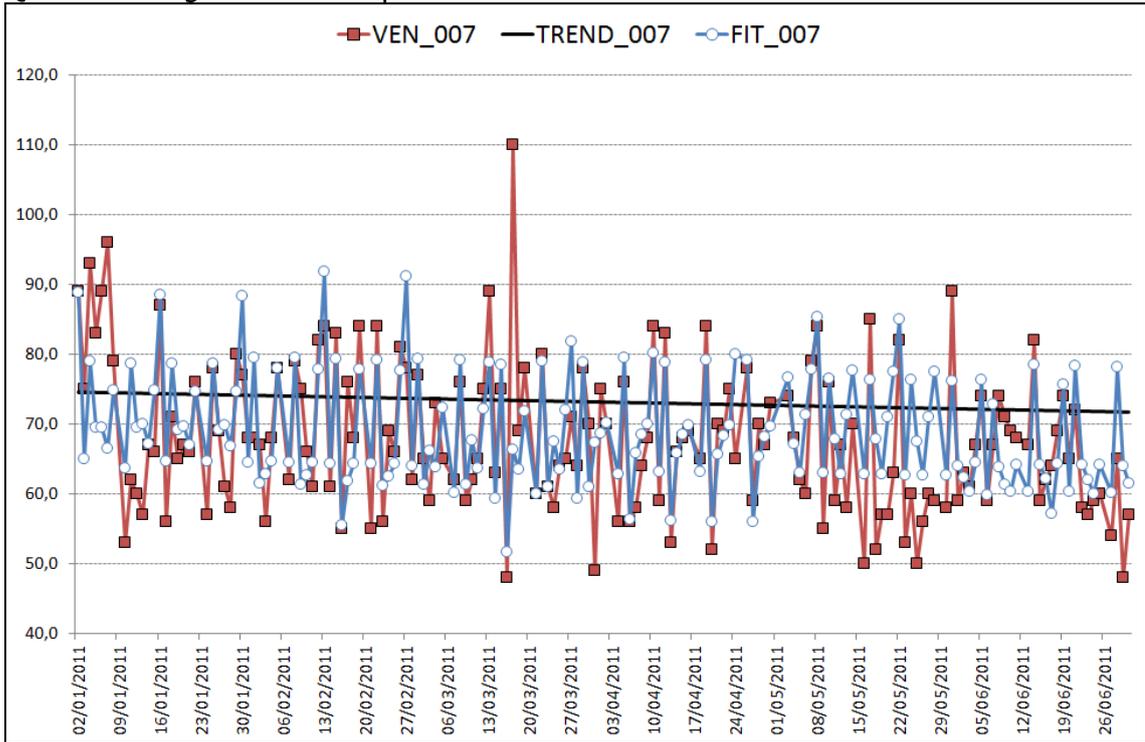


Figura 68. Vendite reali, trend e vendite previste per l'edicola 007

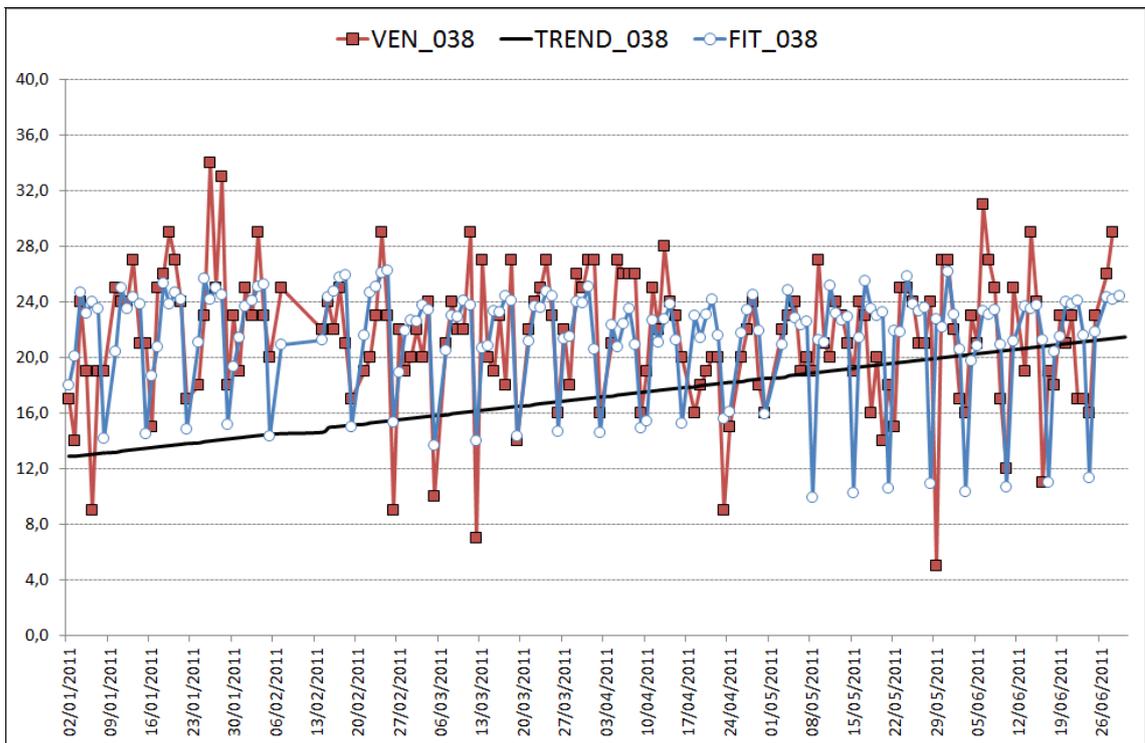


Figura 69. Vendite reali, trend e vendite previste per l'edicola 038

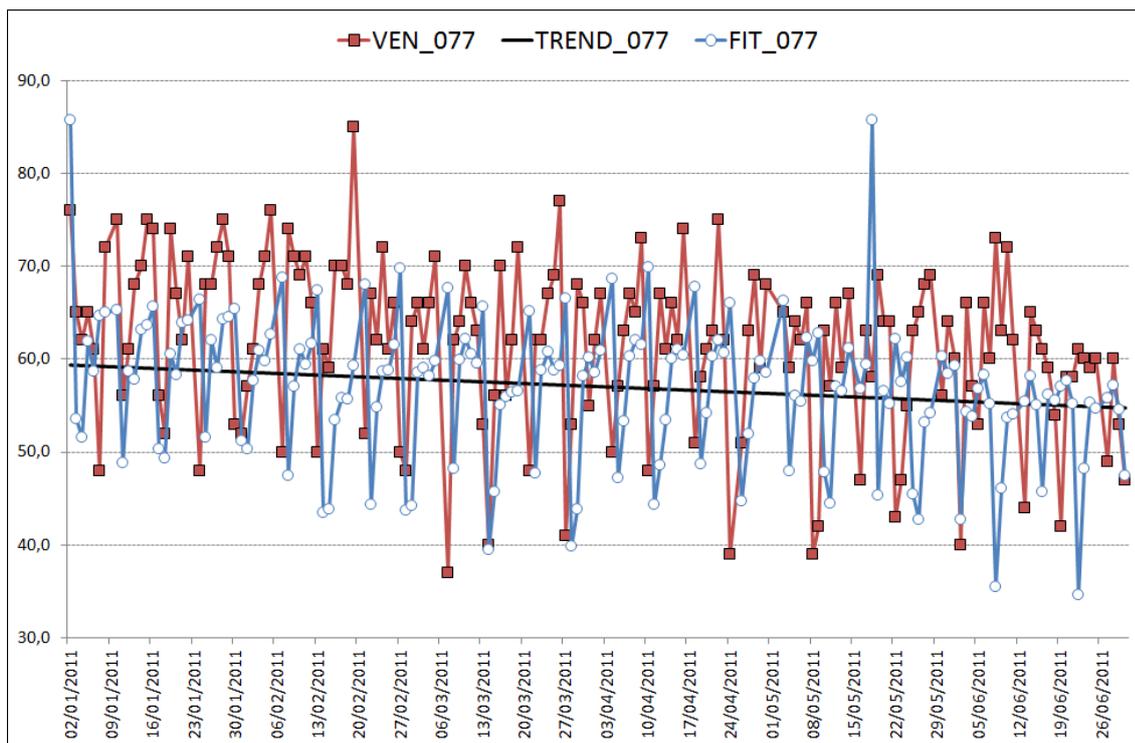


Figura 70. Vendite reali, trend e vendite previste per l'edicola 077

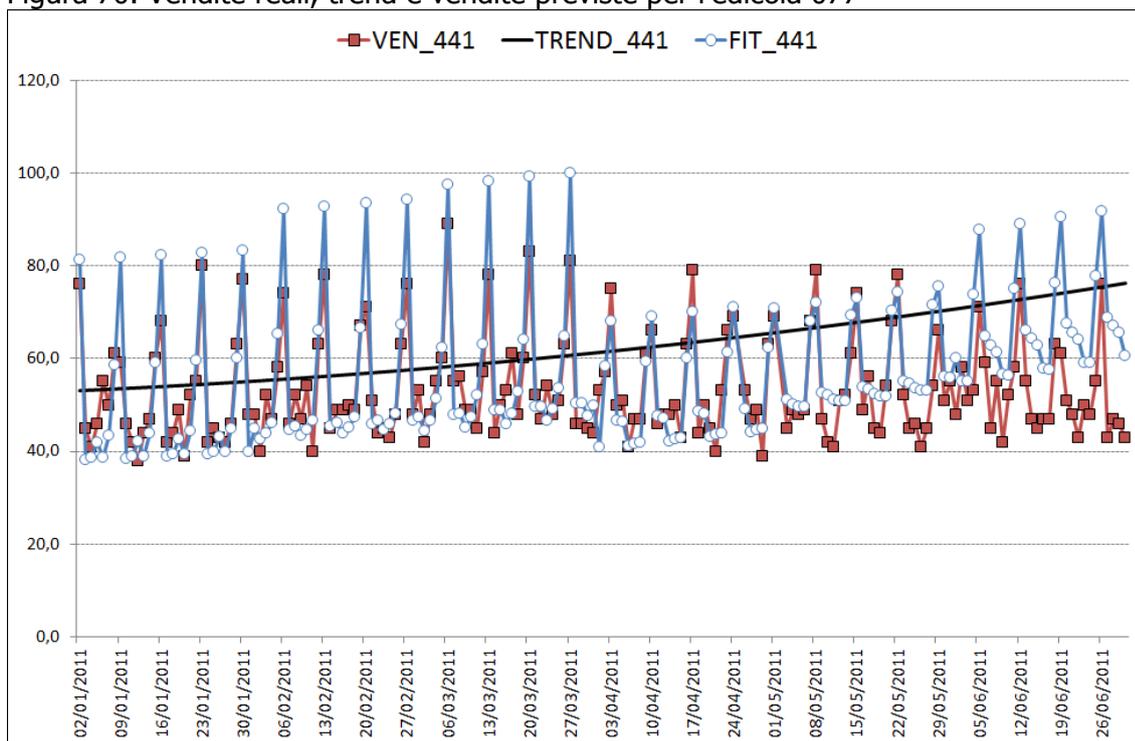


Figura 71. Vendite reali, trend e vendite previste per l'edicola 441

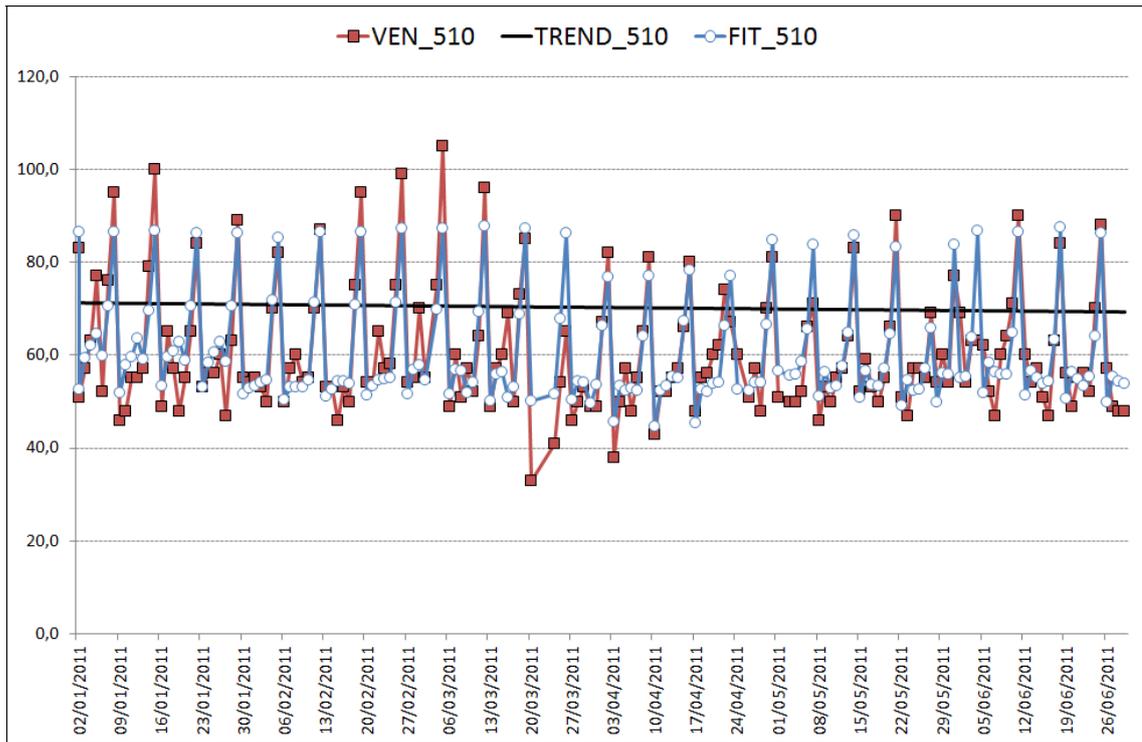


Figura 72. Vendite reali, trend e vendite previste per l'edicola 510

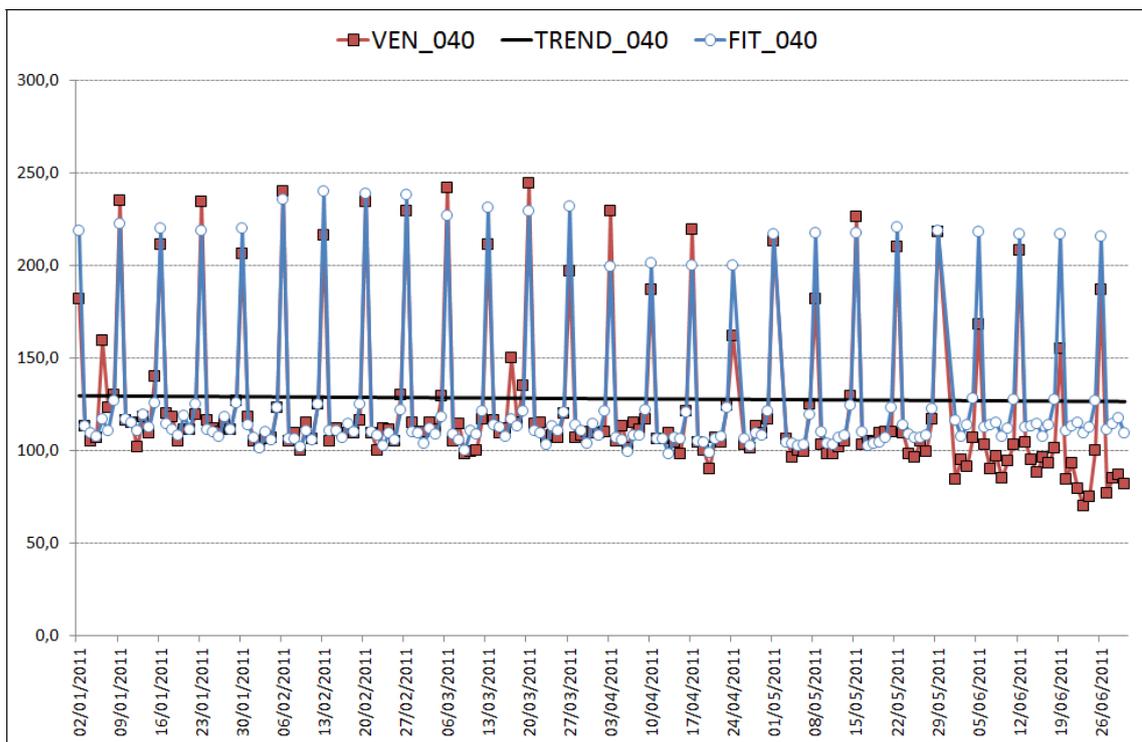


Figura 73. Vendite reali, trend e vendite previste per l'edicola 040

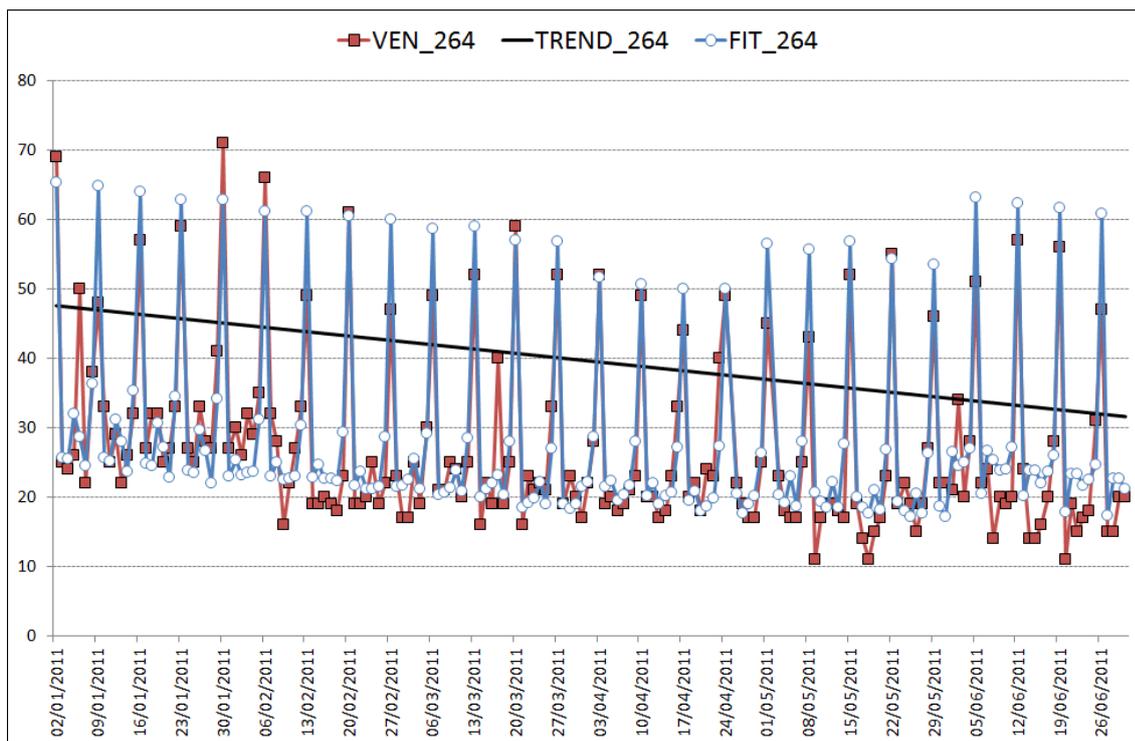


Figura 74. Vendite reali, trend e vendite previste per l'edicola 264

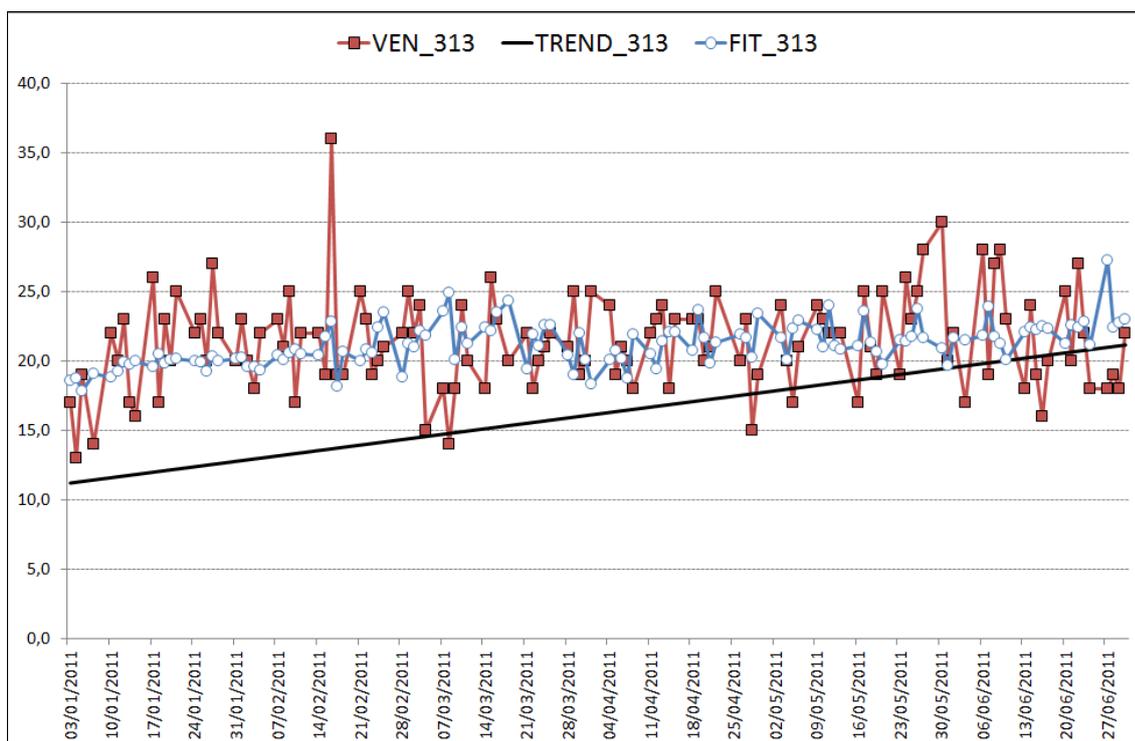


Figura 75. Vendite reali, trend e vendite previste per l'edicola 313

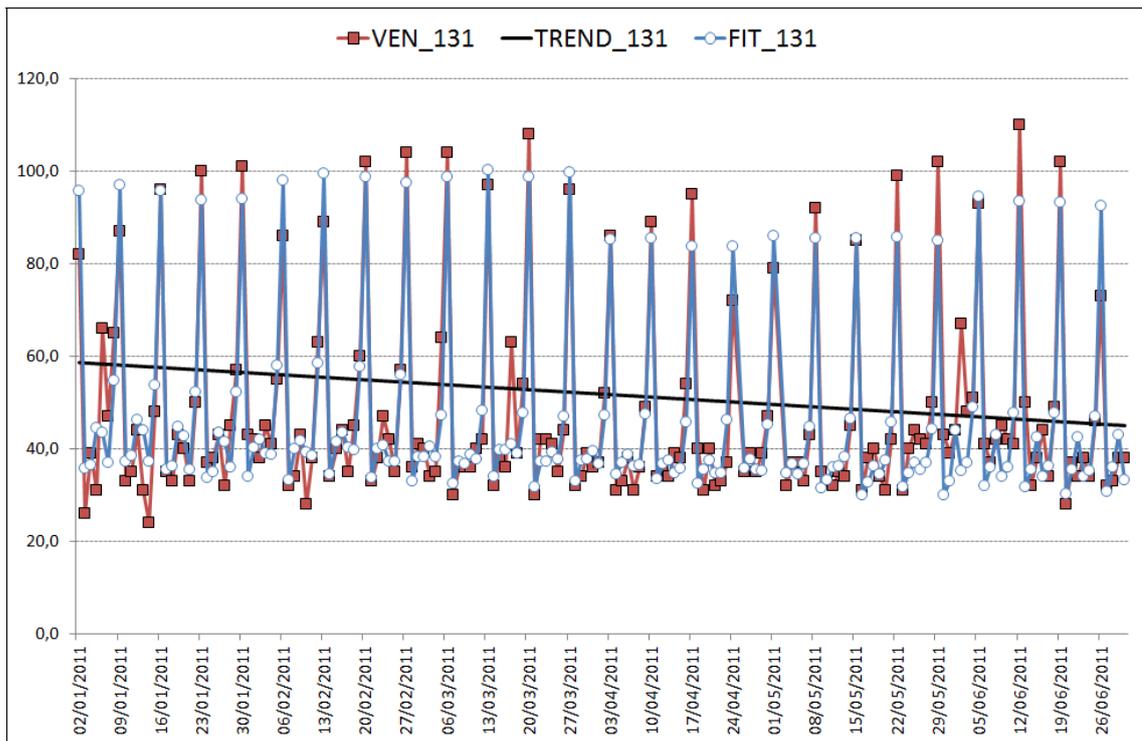


Figura 76. Vendite reali, trend e vendite previste per l'edicola 131

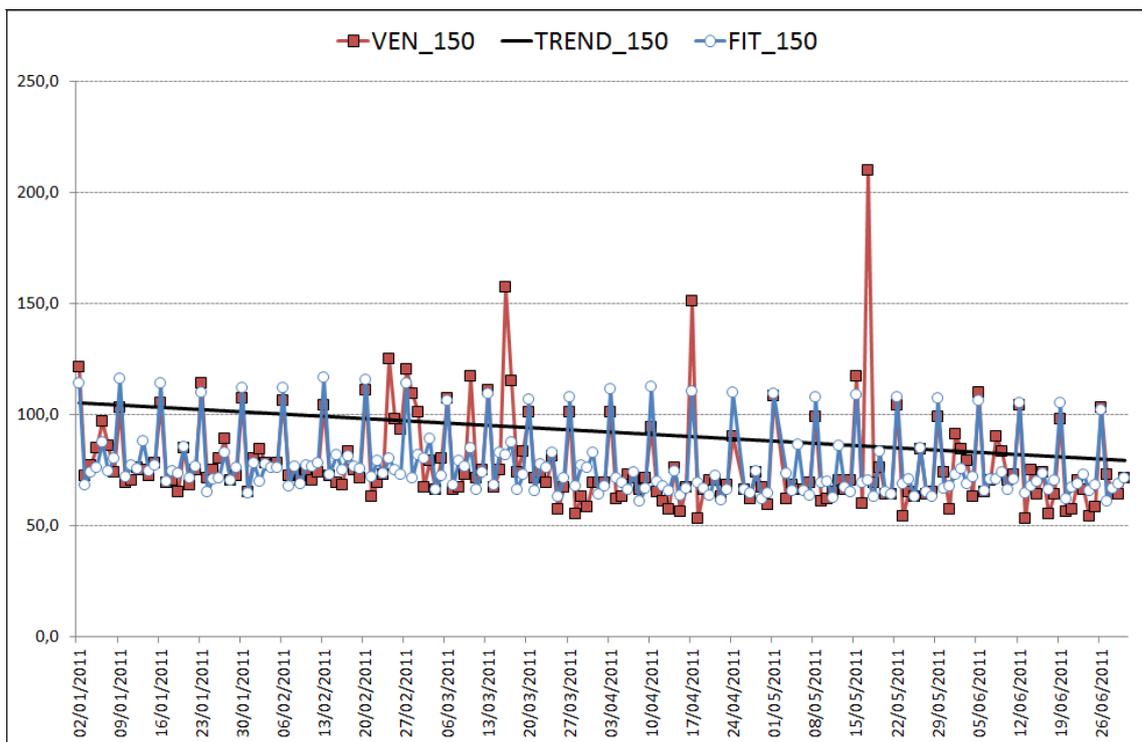


Figura 77. Vendite reali, trend e vendite previste per l'edicola 150

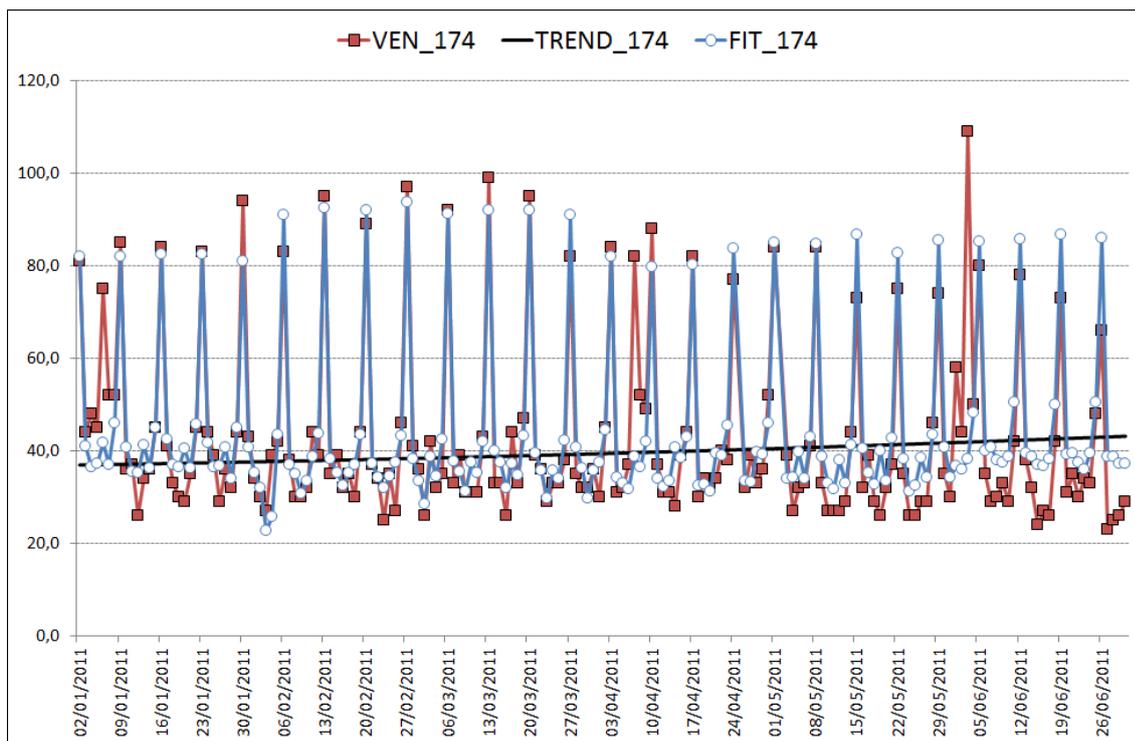


Figura 78. Vendite reali, trend e vendite previste per l'edicola 174

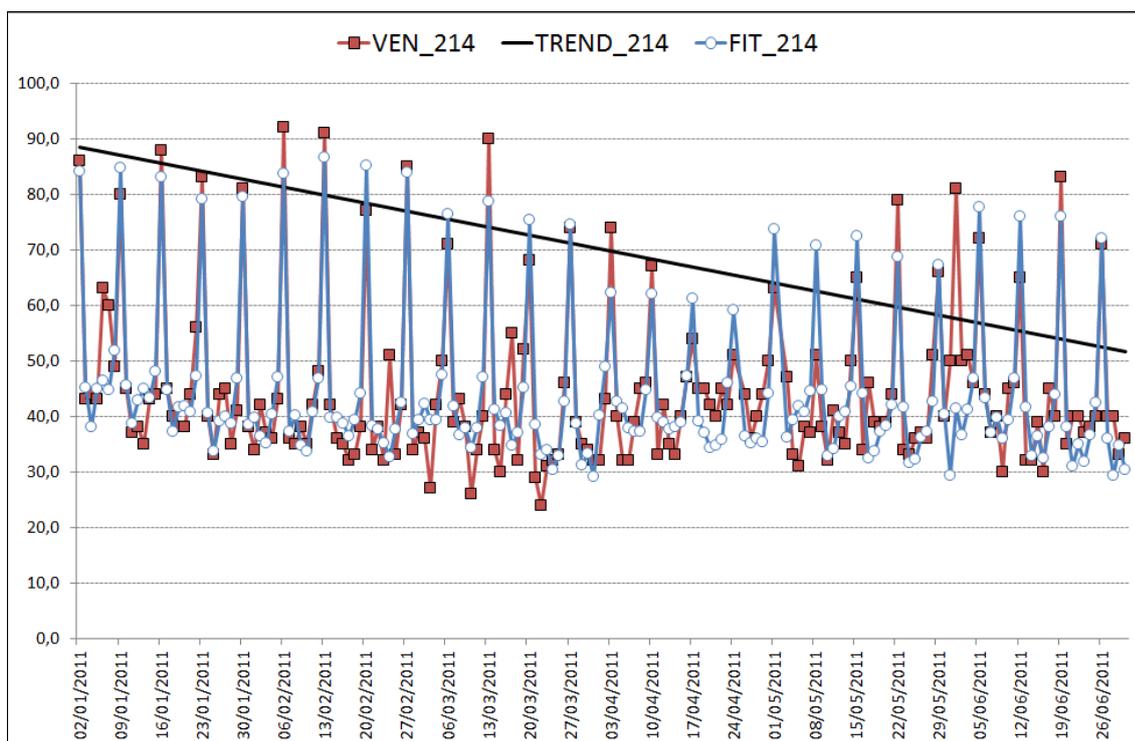


Figura 79. Vendite reali, trend e vendite previste per l'edicola 214

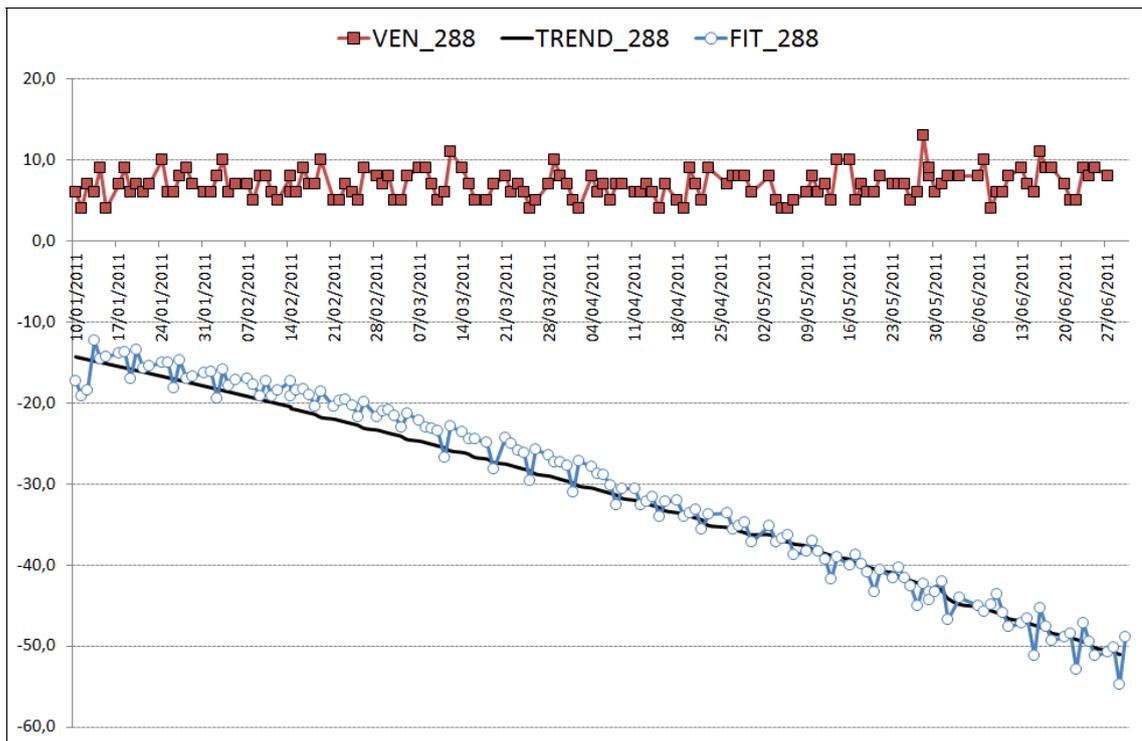


Figura 80. Vendite reali, trend e vendite previste per l'edicola 288

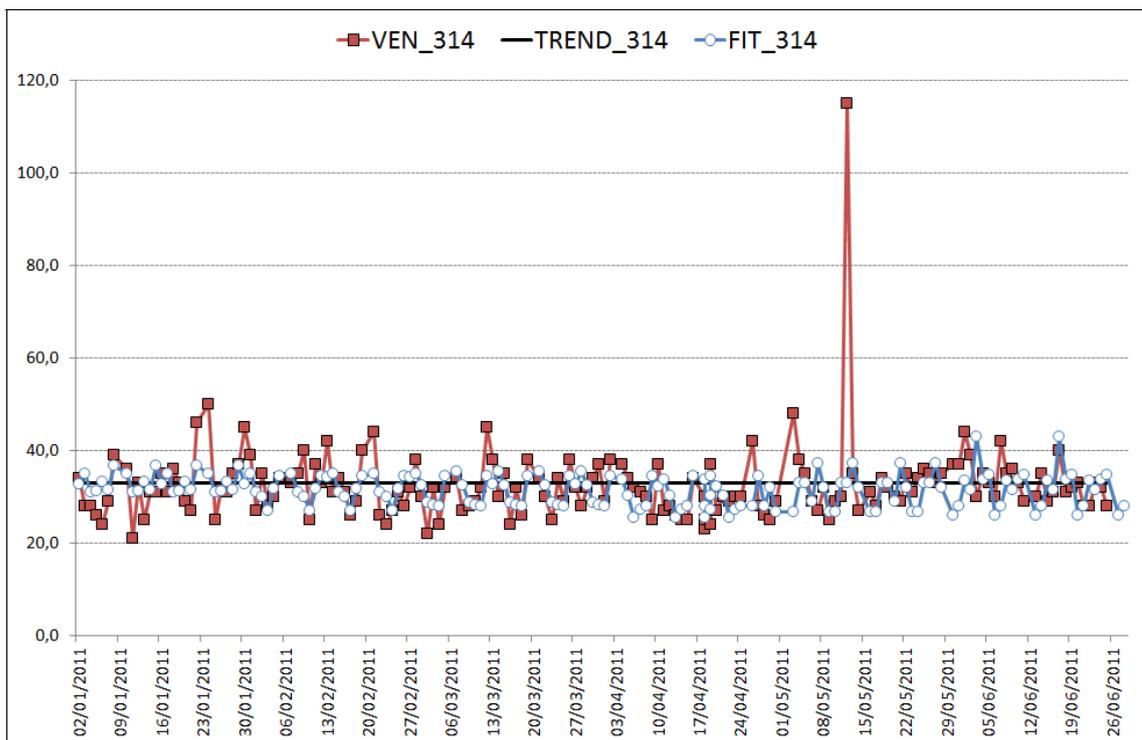


Figura 81. Vendite reali, trend e vendite previste per l'edicola 314

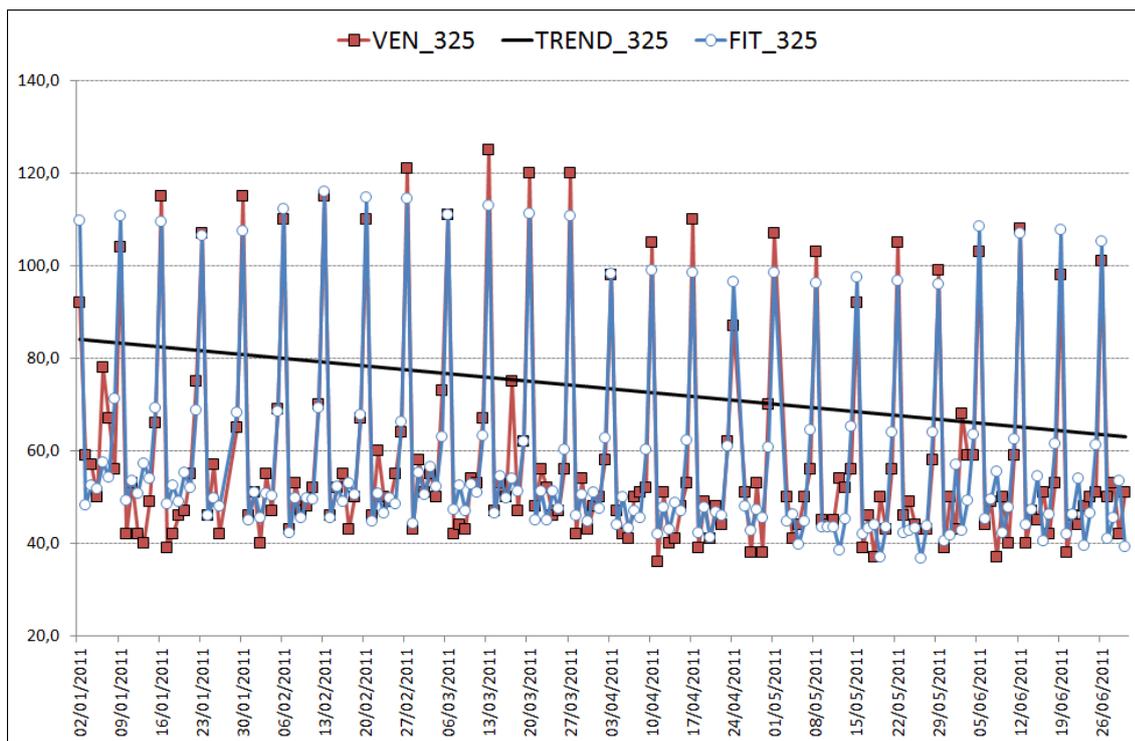


Figura 82. Vendite reali, trend e vendite previste per l'edicola 325

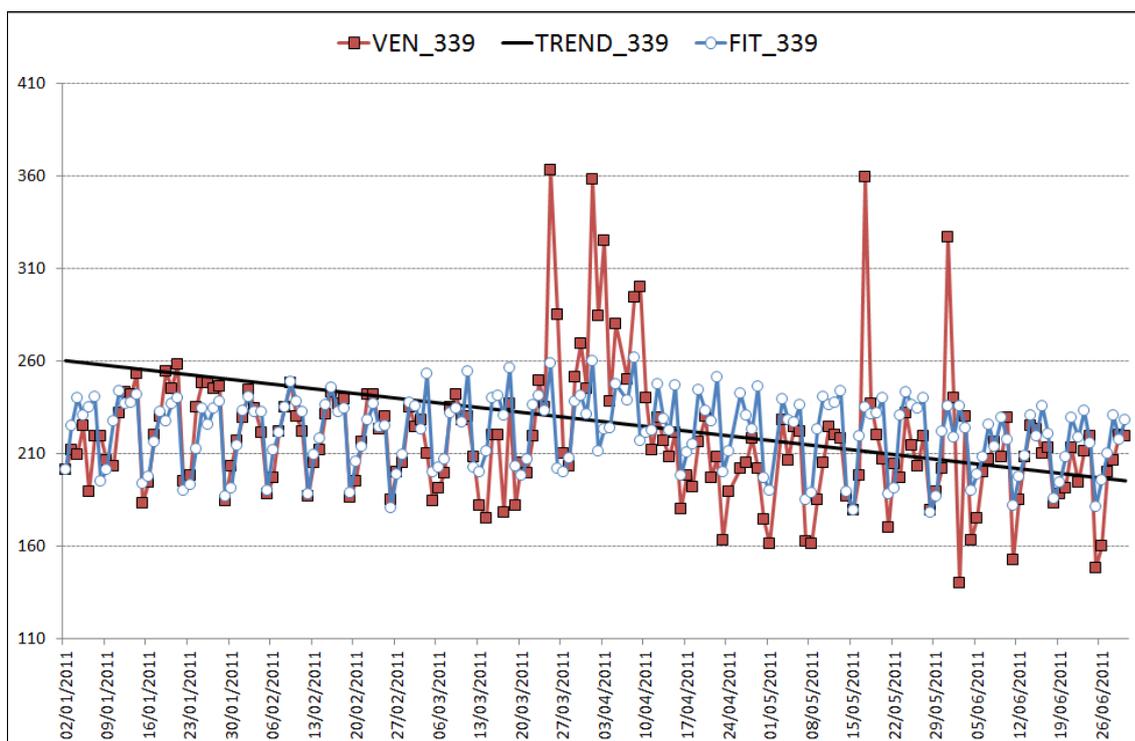


Figura 83. Vendite reali, trend e vendite previste per l'edicola 339

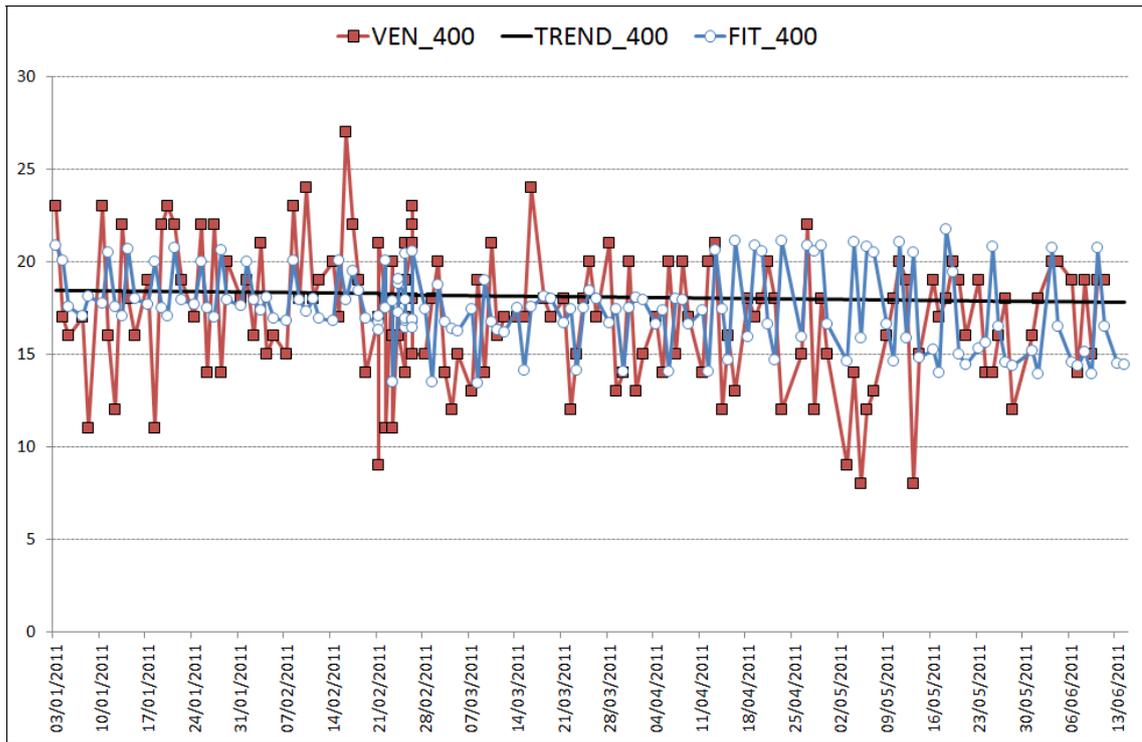


Figura 84. Vendite reali, trend e vendite previste per l'edicola 400

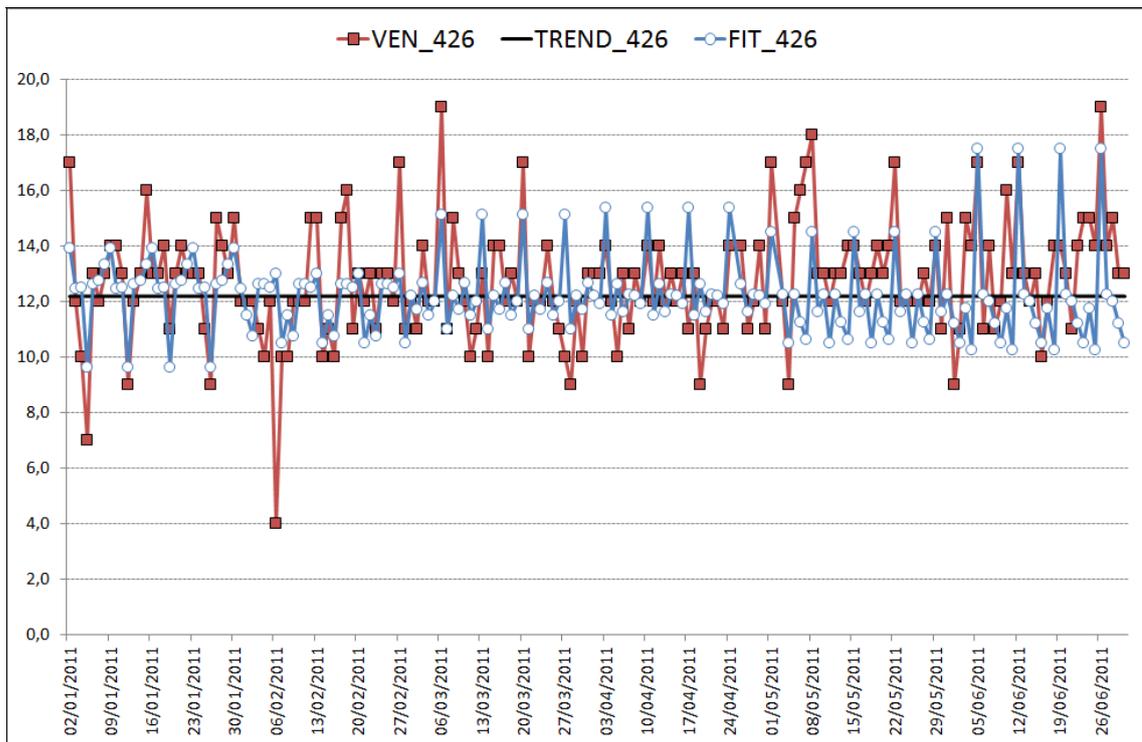


Figura 85. Vendite reali, trend e vendite previste per l'edicola 426

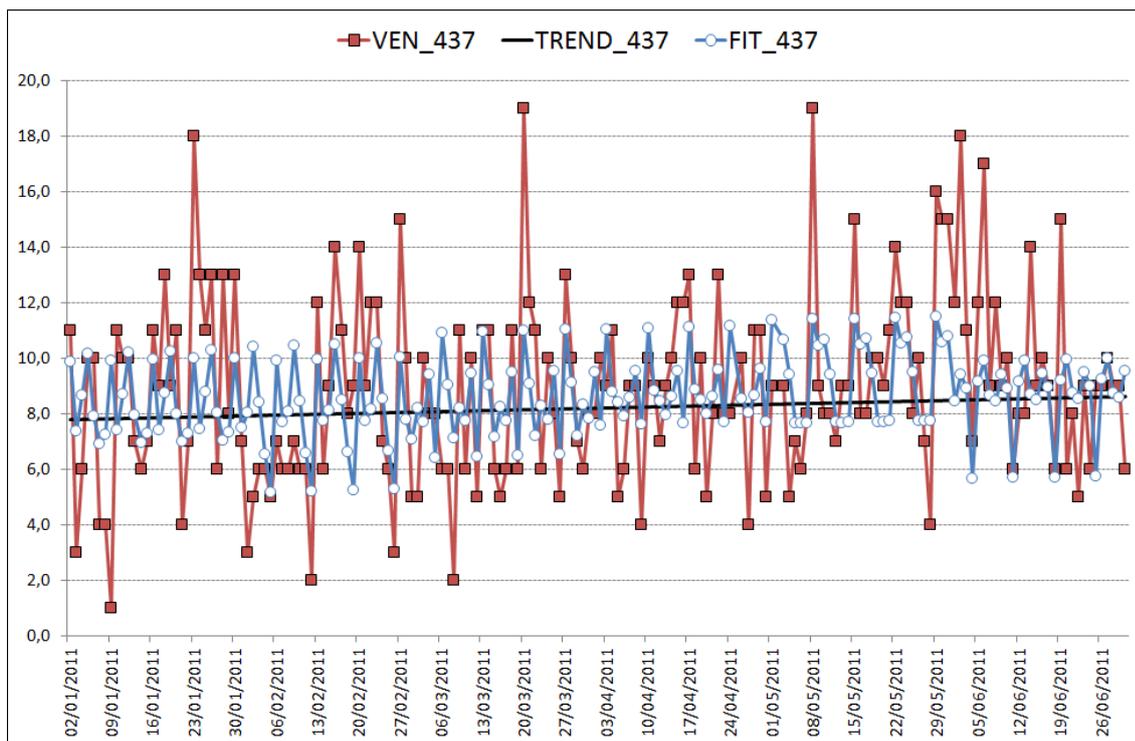


Figura 86. Vendite reali, trend e vendite previste per l'edicola 437

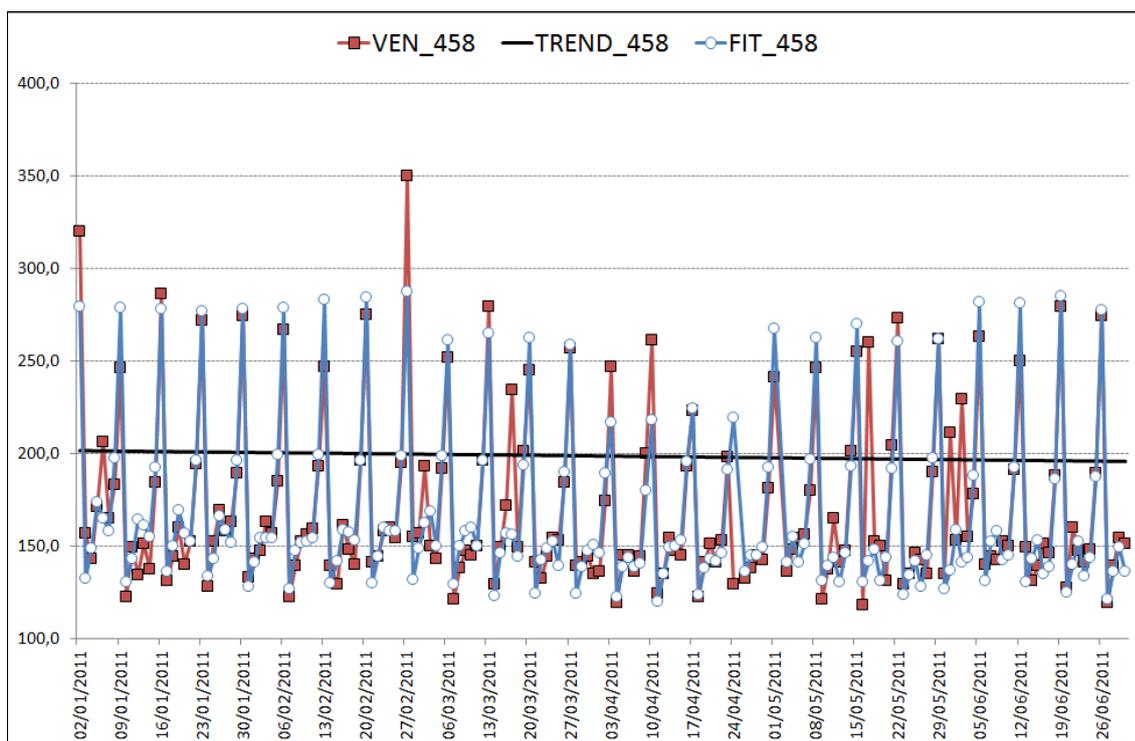


Figura 87. Vendite reali, trend e vendite previste per l'edicola 458

In linea generale, osservando i grafici delle 8 aree si nota che l'andamento delle vendite previste si avvicina a quello delle vendite reali. Il modello sviluppato risulta quindi affidabile e sicuramente più accurato di quello costruito per le singole edicole, in virtù della maggiore mole di vendite.

Per quanto riguarda le singole rivendite, si presentano varie situazioni: in alcuni casi il modello sembra funzionare e le previsioni si discostano poco dalle vendite reali. In altri è evidente che il modello non funziona, in particolare ciò si riscontra nelle edicole aventi quantità vendute piuttosto basse. Ciò è dovuto al coinvolgimento di altri fattori estranei al nostro modello, il quale quindi risulta in questo senso meno esauriente e dunque preferibilmente sostituibile con metodi più sofisticati. Inoltre, è possibile osservare casi in cui il modello risulta adeguato per gran parte dell'arco temporale considerato, ma che perde di efficacia in alcuni periodi o addirittura in giorni isolati. Ciò è spiegato dal fatto che possono accadere eventi particolari o casuali che, per forza di cose, non possono essere controllati.



## CONCLUSIONI

---

In generale possiamo oggettivamente ritenere che il modello sviluppato è accurato e attendibile soprattutto per elaborare previsioni laddove vengono venduti una maggior quantità di giornali quotidiani; facciamo quindi riferimento in particolare alle analisi svolte su ciascuna delle 8 aree in cui è stata suddivisa la Provincia di Padova, ma anche a quelle edicole che hanno registrato dei livelli di vendita giornalieri relativamente alti (circa 130 copie). A tale conclusione si è giunti confrontando l'andamento delle vendite reale con quanto previsto. Vi sono, però, delle singole edicole per le quali quanto progettato non appare coerente: un chiaro esempio è il caso dell'edicola 288 (Figura 80) per la quale il comportamento previsto è surreale. A sostegno di ciò possiamo prendere in considerazione l' $R^2$  che è piuttosto basso, indicandoci quindi che le variabili esplicative considerate hanno una debole proporzionalità con le vendite.

Ulteriori osservazioni interessanti riguardano i diversi picchi anomali (esempio Figura 81) che si verificano nell'andamento reale di alcune edicole: essi sono da imputare a fenomeni fisiologici occasionali privi di una causalità analizzabile. Inoltre i trend relativi alla serie storiche sono per la maggioranza lineari con pendenza negativa: questo fenomeno si può imputare all'aumento del prezzo a partire dall'inizio del 2011, proprio perché si osserva una inversione di tendenza in quel periodo (esempio Figura 18).

Relativamente alle condizioni meteorologiche (Figura 58 e Figura 59) utilizzate per l'analisi, possiamo notare come in alcuni casi laddove il coefficiente assume valori significativi ci possano essere delle tendenze opposte: questo ci suggerisce non che il modello sia incoerente ma piuttosto che sono necessari degli approfondimenti sul loro andamento stagionale. Viceversa la temperatura minima presenta una relazione di diretta proporzionalità con le vendite; ciò potrebbe indurre a una contraddizione dovuta all'effetto ciclico della variabile stessa, ma risulta coerente se si osserva l'arco di tempo analizzato: esso è prevalentemente costituito da basse temperature, avendo preso in considerazione solo i primi mesi del 2011. Inoltre possiamo constatare come alcune variabili risultino significative in rari casi: la radiazione solare perciò pur essendo un parametro climatico di rilievo non ha un reale impatto sulle vendite dei giornali, così come l'umidità relativa.



## BIBLIOGRAFIA

---

A.R.P.A.V. Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto,

(<http://www.arpa.veneto.it/home2/htm/home.asp>),

30 Marzo 2011.

Douglas C. Montgomery, 2005, *Progettazione e analisi degli esperimenti*, McGraw-Hill.

Levine David M., Krehbiel Timothy C., Berenson Mark L., 2010, *Statistica*, Pearson.

Ministero dell'Ambiente e della Tutela del Territorio e del Mare

([http://www.minambiente.it/home\\_it](http://www.minambiente.it/home_it)),

30 Marzo 2011

Minitab (<http://www.minitab.com/en-IT/default.aspx>),

24 Aprile 2011