

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

METODI STATISTICI PER LO STUDIO DELL'IMPATTO DELLA TIPOLOGIA DELLE NOTIZIE SULLE VENDITE DI UN GIORNALE QUOTIDIANO

Relatore: Ch.^{mo} Dott. Livio Corain

Correlatore: Ch.^{mo} Prof. Luigi Salmaso

Laureando: Sara Muraro

ANNO ACCADEMICO 2010-2011

INDICE

SOMMARIO	1
INTRODUZIONE	3
CAPITOLO 1. Analisi delle vendite di un giornale quotidiano	5
1.1 Obiettivi e modalità di svolgimento dell'attività di ricerca.....	5
1.2 Classificazione delle notizie e suddivisione del territorio in aree	5
1.3 Costruzione dei database.....	7
CAPITOLO 2. Aspetti statistici per l'analisi delle vendite del giornale	11
2.1 Definizione del modello statistico	11
2.2 Approfondimento metodologico: La stima.....	12
2.3 Procedura di stima del modello statistico.....	14
2.4 Il software statistico MINITAB.....	16
CAPITOLO 3. Risultati e discussione	19
3.1 Serie temporali delle vendite ed analisi del trend	19
3.2 Analisi statistica a 18 mesi	34
3.3 Analisi statistica previsionale.....	37
CONCLUSIONI	61
BIBLIOGRAFIA	63

SOMMARIO

L'obiettivo di questa tesi di Laurea è svolgere analisi statistiche lo studio dell'impatto della tipologia delle notizie sulle vendite di un giornale quotidiano.

Nel capitolo 1 "Analisi delle vendite di un giornale quotidiano" vengono illustrate le attività preliminari del progetto e gli obiettivi finali, nello specifico la ricerca di associazioni significative tra le vendite del quotidiano e la tipologia di notizie in esso contenute e lo sviluppo di un valido modello previsionale delle vendite.

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 modello grezzo di partenza, tenente conto dell'effetto ciclico e del prezzo del quotidiano oltre alla tipologia di notizie contenute, si opera con analisi statistiche al fine di scartare quelle variabili risultanti poco significative. Il fine è cioè quello di depurare il modello dalle variabili superflue.

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 precisione del progetto.

Il lavoro di tesi si conclude con alcuni commenti in merito ai risultati osservati dai grafici. Oltre a porre particolare evidenza ai casi in cui il modello sembra funzionare bene e risulta dunque uno strumento utile all'azienda, si fa anche riferimento ai casi dove non risulta abbastanza accurato, ai quali si cerca di dare una spiegazione.

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 la tipologia di notizie in essi contenute. A tal proposito, per ogni giorno dell'arco temporale considerato sono state catalogate in un file Excel tutte le notizie di dimensione superiore al quarto di pagina, indicandone per ciascuna la dimensione effettiva (180 corrisponde alla pagina intera), la rilevanza (locale, regionale, nazionale, internazionale) e la tipologia. I tipi di notizia considerati sono stati: cronaca, cronaca arresti, cronaca atti violenti, cronaca disastri ambientali, cronaca furti, cronaca giudiziaria, cronaca incidenti, cronaca incidenti mortali, cronaca morti, cronaca omicidi, cronaca scioperi, cronaca suicidi, cronaca violenza sessuale, economia, eventi, politica, sport, trasporti. A partire da questi dati è stato possibile costruire un dataset contenente per ogni giorno il numero di pagine del quotidiano appartenenti a ciascun tipo di notizia.

Una volta costruito il dataset, si è proseguito ad analizzare i dati raccolti 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. Il motivo di questo modo di operare risiede nel fatto che al diminuire del numero di variabili i test statistici diventano 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. In un primo momento si è costruito il trend delle vendite di tutte le edicole e aree, di modo da evidenziare l'andamento delle vendite dovuto solo ed esclusivamente al fattore ciclico e al prezzo, e non alle variabili riguardanti la tipologia delle notizie. 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. Il modello è stato dunque utilizzato al fine di ottenere delle previsioni di vendita, utilizzando i valori delle variabili individuate, ovvero di quanta parte del giornale del giorno i -esimo è stata dedicata a notizie di ciascuna prefissata categoria di rilevanza e di tipologia. 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 mostranti 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 previsivo 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:

definizione di una opportuna classificazione delle notizie del giornale sulla base delle loro caratteristiche principali (rilevanza e tipologia della notizia);

costruzione di un database ad hoc su base Excel per l'inserimento delle notizie;

inserimento nel suddetto database delle varie notizie, con una copertura finale di 18 mesi (gen. 2010 – giu. 2011);

collegamento del database notizie con il database vendite.

La ricerca si propone principalmente i due seguenti obiettivi:

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 delle notizie riportate nel giornale;

costruire e validare un modello statistico capace di fare delle previsioni possibilmente affidabili delle vendite in funzione delle caratteristiche delle notizie riportate nel giornale e di tutti i possibili fattori potenzialmente impattanti sulle vendite.

1.2 Classificazione delle notizie e suddivisione del territorio in aree

Di intesa con la direzione del giornale, è stata formulata una opportuna classificazione delle notizie sulla base delle loro caratteristiche principali, ovvero la rilevanza e la tipologia della notizia, rispettivamente definite rispettivamente secondo le seguenti categorie:

– **RILEVANZA DELLA NOTIZIA**

- internazionale (INT)
- nazionale (NAZ)
- regionale (REG)
- locale (LOC)

– **TIPOLOGIA DELLA NOTIZIA**

- Cronaca - Omicidi (CR_OM)
- Cronaca - Suicidi (CR_SU)
- Cronaca - Morti (non violenti) (CR_MO)
- Cronaca - Furti/rapine (CR_FU)
- Cronaca - Violenza sessuale (CR_VI_SES)
- Cronaca - Atti violenti (altri, rispetto a furti/rapine) (CR_AV)
- Cronaca - Incidenti mortali (CR_IN_M)
- Cronaca - Incidenti (non mortali) (CR_IN)
- Cronaca - Disastri ambientali (CR_DS)
- Cronaca - Arresti (CR_ARR)
- Cronaca - Scioperi, proteste, manifestazioni (CR_SC)
- Cronaca giudiziaria (CR_GIU)
- Cronaca - Altro (CR)
- Trasporti, viabilità (TRA)
- Politica (POL)
- Economia (ECO)
- Eventi (palio, fiera di ..., ecc.) (EVE)
- Gossip (GOS)
- Insetto culturale (mostre, musei, ecc.) (CUL)
- Approfondimenti di "servizio" (orari autobus, mappe autovelox, ecc.) (APP)
- Altro (ALT)
- Sport (SPO)

1.3 Costruzione dei database

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

Database notizie;

Database quotidiano;

Database vendite;

Database previsioni.

Per la costruzione del database notizie (Figura 1), finora popolato con 536 quotidiani per un totale di 17526 notizie, i criteri adottati di inclusione delle notizie sono stati i seguenti:

notizia di dimensione almeno 90 moduli (mezza pagina), per le notizie di rilevanza internazionale, nazionale e regionale;

notizia di dimensione almeno 45 moduli (un quarto pagina), per le notizie di rilevanza locale.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Data	Pag	Sez	Dim	Ril	Luogo	Luogo_int	Luogo_int_2	Tipo	Val_diff	Note			
2	02/01/2010	3	PP		90 NAZ	PADOVA			ECO					
3	02/01/2010	3	PP		45 REG	PADOVA			ECO		notizia che riguarda i trasporti: aumento dei prezzi dei biglietti			
4	02/01/2010	4	AT		90 NAZ	ROMA			ECO					
5	02/01/2010	4	AT		40 NAZ	ROMA			ECO					
6	02/01/2010	4	AT		40 NAZ	ROMA			ECO					
7	02/01/2010	5	AT		180 NAZ	ROMA			POL					
8	02/01/2010	6	AT		90 INT	THERAN			CR					
9	02/01/2010	7	AT		90 INT	ISLAMABAD			CR_OM		terrorismo			
10	02/01/2010	7	AT		45 INT	AFGHANISTAN			CR_OM		sequestro e omicidi			
11	02/01/2010	7	AT		45 INT	WASHINGTON			POL					
12	02/01/2010	9	AT		90 NAZ	PALERMO			CR		mafia			
13	02/01/2010	12	RE		70 NAZ	ROMA	VENEZIA		EVE		candidatura olimpici 2020			
14	02/01/2010	13	RE		90 NAZ	PADOVA			POL		senatore del PD parla di sanità			
15	02/01/2010	14	ECO		90 REG	VENETO			ALT		fonti alternative			
16	02/01/2010	14	ECO		45 NAZ	VENEZIA	TRIESTE		TRA					
17	02/01/2010	19	PD		90 LOC	PADOVA			CR_FU					
18	02/01/2010	20	PD		180 LOC	PADOVA			CR_ARR		polizia spegne rave party a sorpresa			
19	02/01/2010	21	PD		180 LOC	PADOVA			CR_ARR		polizia spegne rave party a sorpresa			
20	02/01/2010	23	PD		90 LOC	PADOVA			CR_MO					
21	02/01/2010	24	PD		90 LOC	PADOVA			ALT		il primo nato d'Italia è di Busa di Vigonza			
22	02/01/2010	25	PD		45 LOC	PADOVA			CR_SC					
23	02/01/2010	27	LR-RJ-SC		60 LOC	LIMENA			CR		azienda salvata in extremis			
24	02/01/2010	29	AB-MON_T-CL		80 LOC	MONTEGROTTO			CR_AV					
25	02/01/2010	30	CI-CA		80 LOC	CITTEDELLA			CR					
26	02/01/2010	31	CI-CA		80 LOC	CITTEDELLA			CR_IN					
27	02/01/2010	31	CI-CA		45 LOC	VIGONZA			CR					
28	02/01/2010	32	PS-SA		55 LOC	PIOVE DI SACCO			TRA					
29	02/01/2010	32	PS-SA		60 LOC	PIOVE DI SACCO			CR_IN					
30	02/01/2010	33	MONS-ES		50 LOC	ESTE			CR		Sequestrato un quintale di botti			
31	02/01/2010	35	SP		100 NAZ	ROMA			SPO					
32	02/01/2010	36	SP		130 LOC	PADOVA			SPO					
33	02/01/2010	36	SP		45 LOC	PADOVA			SPO					

Figura 1. Alcuni record del database notizie.

In base a tali criteri, il numero medio di notizie che sono state classificate ed archiviate per i 536 quotidiani considerati è stato pari a 32.2 (per quotidiano).

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

Il database notizie è stato quindi sintetizzato in un dataset denominato "database quotidiano", dove ciascun record indica lo spazio (numero di pagine) che in un dato quotidiano è stato assegnato alle notizie con diversa rilevanza e tipologia (Figura 2). Infine, il database quotidiano è stato collegato al "database vendite" (Fig. 3) in modo da costituire il "database previsioni", che comprende due versioni: la prima utile ad una analisi aggregata per area (Fig. 4), la seconda finalizzata allo studio delle vendite per singola edicola.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Data	INT	LOC	NAZ	REG	CR_TOT	CR_NERA	CR_GRIGIA	CR_NC	CR_ARR	CR_AV	CR_DS	CR_FU	CR_GIU	CR_IN	CR_IN M	CR_MO	CR_OM	CR_SC
2	02/01/2010	1.5	8	4.638889	0.75	5.777778	0.944444	3.277778	1.305556	2	0.444444	0	0.5	0	0.777778	0	0.5	0	0.25
3	03/01/2010	2.138889	11.02778	2	3.194444	4.027778	0.805556	1.5	1.722222	0.444444	0.25	0	0.277778	0	0	0.277778	1.055556	0	0
4	04/01/2010	1.277778	13.33333	3	0.444444	4.083333	3.277778	0	0.805556	0	0	0	0	0	0	1.277778	0	0	0
5	05/01/2010	1.611111	9.444444	2.583333	2.055556	3.194444	1.5	1.444444	0.25	0.361111	0	0	0	0.25	1.111111	0.833333	0.388889	0	0
6	06/01/2010	0.388889	14.44444	3.194444	1.444444	3.861111	2.416667	0.25	1.194444	0	0	1.194444	0	0	1.222222	0.25	0	0	0
7	07/01/2010	0.444444	12.94444	3.694444	3.194444	5.388889	1.083333	1.111111	3.194444	0	0	0	0	0	0.555556	0.694444	0.555556	0	0
8	08/01/2010	0.25	12.75	4.666667	2.333333	4.833333	2.138889	0	2.194444	0	0	0	0.5	0	1.888889	0	0	0	0
9	09/01/2010	0.25	12	4.722222	1.25	3.888889	2	0.555556	0.833333	0.305556	0	0	0.5	0.25	2	0	0	0	0
10	10/01/2010	0.916667	3.083333	3.166667	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	11/01/2010	0.555556	16.33333	2.472222	3.444444	4.861111	2.722222	0	2.138889	0	0.444444	0	0.444444	0	0	1.833333	0	0	0
12	12/01/2010	0.5	10.08333	3.805556	3.166667	3.333333	1.638889	0.25	0.75	0	0	0	0	0.694444	0.25	1.638889	0	0	0
13	13/01/2010	0.5	11.11111	3.416667	2.027778	6.222222	3.75	1.472222	0.75	0.25	0.25	0	1.722222	0.25	2.5	1.777778	0.972222	0	0
14	14/01/2010	1.722222	9.638889	6.361111	0.944444	3.611111	2.722222	0.25	0.638889	0	0	0	1.916667	0	0.25	0.805556	0	0	0
15	15/01/2010	2.138889	13.58333	2.5	1.638889	9.027778	2.805556	0.888889	1.75	0.388889	0	0.944444	0.305556	2.638889	0.5	0.25	0	0	0
16	16/01/2010	0.833333	9.166667	2.555556	2.305556	3.166667	1.388889	0.277778	1.5	0	0.305556	1.083333	0	0	0.277778	0	0	0	0
17	17/01/2010	1.972222	14.55556	1	3.75	5.333333	1.25	0.888889	2.083333	0.25	0	1.25	1.111111	0.25	0	0.388889	0	0	0
18	18/01/2010	1	16.25	2.722222	1	2.666667	0.888889	1.277778	0.5	0	0.25	0.25	0	0	0.388889	1.277778	0	0	0
19	19/01/2010	1.25	9.027778	4.277778	1.694444	4.444444	0.472222	1	2.722222	0.25	0	0	0	0	0	0.75	0.472222	0.25	0
20	20/01/2010	0.833333	12.61111	3.972222	1.527778	5.75	2.194444	0.444444	2.861111	0	0.25	0	0.694444	0.25	0	0.25	0.444444	0.25	0
21	21/01/2010	1.722222	10.75	2.638889	1.583333	5.722222	3.75	0	1.972222	0	0.5	0	0.583333	0	1.166667	0	0.5	0	0
22	22/01/2010	0.972222	10.5	5.111111	1.75	3.75	2.75	0.25	0.5	0.25	0	0.5	0.25	0	2	0	0.25	0	0
23	23/01/2010	2.5	8.083333	3.888889	1.083333	5.472222	2.75	0.444444	2.277778	0	1.027778	0	0.5	0	0.388889	0.444444	0.833333	0	0
24	24/01/2010	0.444444	15.02778	4.277778	1.333333	5.027778	1.527778	0.75	2.25	0.75	0	0.527778	0.25	0	0	0.5	0.25	0	0
25	25/01/2010	1.833333	14.86111	3	1	2.305556	0.277778	0.5	0.694444	0.5	0	0	0	0	0.277778	0	0	0.833333	0
26	26/01/2010	0.5	9.111111	5.111111	1.583333	2.138889	1.083333	0	0.5	0	0.25	0	0	0	0.833333	0	0	0.666667	0
27	27/01/2010	0.5	11.16667	2.833333	0.75	4.444444	2.583333	0	0.777778	0	0	0	0	0	1.583333	0	0	0.416667	0
28	28/01/2010	0.611111	9.388889	4.194444	1.611111	2.166667	0.583333	0.333333	0.75	0.333333	0	0	0.25	0	0	0	0.333333	0.5	0
29	31/01/2010	0	9.388889	5.583333	1.611111	2.138889	0.638889	0.388889	0.861111	0	0	0	0.638889	0	0	0	0.388889	0.25	0
30	01/02/2010	0	11.52778	6.805556	0	1.944444	0	0.25	0	0	0	0	0	0.25	0	0.694444	0	0	0
31	02/02/2010	0	9.138889	5.25	1.416667	2.777778	0.25	0.5	1.194444	0.5	0	0.833333	0.25	0	0	0	0	0	0
32	03/02/2010	1.666667	9.805556	3.805556	1	4.111111	2	0.583333	0.444444	0.416667	0	0	0.805556	0.25	0.777778	0.333333	0	0.833333	0
33	04/02/2010	0.777778	9.027778	6.25	2.555556	1.305556	0	1.027778	0	0	0	0	0.277778	0	0	0	0	0	0

Figura 2. Alcuni record del database quotidiano.

Dati di Vendita 2010.xlsx - Microsoft Excel uso non commerciale

A	B	C	D	E	G	H	I	J	K	L	M	N	O	P
1	Data	Comune	CAP	Località/Quartiere	Cod. Edicola	Inviato	Reso	Vendita	Esaurito	Gior_set	Settimana	Mese	Zona	
2	02/01/2010	Padova	35100	PADOVA	2040001	60	15	45	0	7	1	1	1 PADOVA	
3	03/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	2	1	1 PADOVA	
4	04/01/2010	Padova	35100	PADOVA	2040001	40	1	39	0	2	2	1	1 PADOVA	
5	05/01/2010	Padova	35100	PADOVA	2040001	40	1	39	0	3	2	1	1 PADOVA	
6	06/01/2010	Padova	35100	PADOVA	2040001	40	26	14	0	4	2	1	1 PADOVA	
7	07/01/2010	Padova	35100	PADOVA	2040001	75	38	37	0	5	2	1	1 PADOVA	
8	08/01/2010	Padova	35100	PADOVA	2040001	60	23	37	0	6	2	1	1 PADOVA	
9	09/01/2010	Padova	35100	PADOVA	2040001	50	6	44	0	7	2	1	1 PADOVA	
10	10/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	3	1	1 PADOVA	
11	11/01/2010	Padova	35100	PADOVA	2040001	45	16	29	0	2	3	1	1 PADOVA	
12	12/01/2010	Padova	35100	PADOVA	2040001	43	2	41	0	3	3	1	1 PADOVA	
13	13/01/2010	Padova	35100	PADOVA	2040001	50	8	42	0	4	3	1	1 PADOVA	
14	14/01/2010	Padova	35100	PADOVA	2040001	50	4	46	0	5	3	1	1 PADOVA	
15	15/01/2010	Padova	35100	PADOVA	2040001	49	11	38	0	6	3	1	1 PADOVA	
16	16/01/2010	Padova	35100	PADOVA	2040001	50	0	50	1	7	3	1	1 PADOVA	
17	17/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	4	1	1 PADOVA	
18	18/01/2010	Padova	35100	PADOVA	2040001	35	11	24	0	2	4	1	1 PADOVA	
19	19/01/2010	Padova	35100	PADOVA	2040001	45	13	32	0	3	4	1	1 PADOVA	
20	20/01/2010	Padova	35100	PADOVA	2040001	45	1	44	0	4	4	1	1 PADOVA	
21	21/01/2010	Padova	35100	PADOVA	2040001	40	0	40	1	5	4	1	1 PADOVA	
22	22/01/2010	Padova	35100	PADOVA	2040001	40	1	39	0	6	4	1	1 PADOVA	
23	23/01/2010	Padova	35100	PADOVA	2040001	55	7	45	0	7	4	1	1 PADOVA	
24	24/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	5	1	1 PADOVA	
25	25/01/2010	Padova	35100	PADOVA	2040001	45	19	26	0	2	5	1	1 PADOVA	
26	26/01/2010	Padova	35100	PADOVA	2040001	43	4	39	0	3	5	1	1 PADOVA	
27	27/01/2010	Padova	35100	PADOVA	2040001	50	9	41	0	4	5	1	1 PADOVA	
28	28/01/2010	Padova	35100	PADOVA	2040001	40	9	31	0	5	5	1	1 PADOVA	
29	29/01/2010	Padova	35100	PADOVA	2040001	45	10	35	0	6	5	1	1 PADOVA	
30	30/01/2010	Padova	35100	PADOVA	2040001	58	11	47	0	7	5	1	1 PADOVA	
31	31/01/2010	Padova	35100	PADOVA	2040001	1	1	0	0	1	6	1	1 PADOVA	
32	02/02/2010	Padova	35100	PADOVA	2040003	65	27	38	0	7	1	1	1 PADOVA	
33	03/02/2010	Padova	35100	PADOVA	2040003	1	1	0	0	1	2	1	1 PADOVA	
34	04/02/2010	Padova	35100	PADOVA	2040003	35	1	34	0	2	2	1	1 PADOVA	

Figura 3. Alcuni record del database vendite.

Dataset_Vendite AREE + 40 Edicole.xlsx - Microsoft Excel uso non commerciale

A	B	C	D	E	F	G	H	I	AA	AB	AC	AD	AE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP			
1	Data	VEN_CAMP	VEN_CINT	VEN_ESTI	VEN_MONS	VEN_PADD	VEN_PIOV	VEN_TERM	VEN_510	VEN_040	VEN_063	VEN_246	VEN_264	Prezzo	Giorno	Giorno*	Giorno**	Mese	Gior_set	INT	LOC	MA2	REG	ECO			
2	02/01/2010	2.598	5.943	1.096	1.095	2.261	8.549	2.084	3.695	76	140	163	42	50	1	1	1	1	7	15	8	4.638895	0.75	0			
3	03/01/2010	3.022	6.333	2.189	2.268	3.695	8.933	2.972	4.029	93	201	169	39	72	1	2	4	8	1	2	13.9889	10.0278	2	3.944444	0.88889	0	
4	04/01/2010	2.084	4.735	1.769	1.663	2.405	7.881	1.626	2.762	60	124	139	26	33	1	3	9	27	1	2	12.7778	13.3333	3	0.444444	0	0	
5	05/01/2010	2.253	5.012	1.895	1.841	2.475	8.231	1.803	2.391	63	119	140	33	25	1	4	16	64	1	3	16.1111	9.444444	2.583333	2.055556	0.916667	0	
6	06/01/2010	2.981	5.897	2.056	2.089	3.071	9.027	2.244	3.651	76	139	163	38	65	1	5	25	125	1	4	8.388889	14.444444	3.944444	14.444444	0.444444	0	
7	07/01/2010	2.858	5.744	2.232	2.165	2.997	9.169	2.020	3.420	76	121	165	43	36	1	6	36	26	1	5	0.444444	12.944444	3.944444	3.944444	1	0	
8	08/01/2010	2.282	5.007	1.892	1.774	2.553	8.347	1.795	2.679	70	124	161	34	30	1	7	49	343	1	6	0.26	12.75	4.666667	2.333333	0.277778	0	
9	09/01/2010	2.446	5.759	1.899	2.805	2.923	8.950	2.066	3.309	66	127	143	32	45	1	8	64	512	1	7	0.25	12.472222	1.25	1.333333	0	0	
10	10/01/2010	3.294	7.029	2.405	2.595	3.977	9.394	2.910	4.412	84	250	208	76	1	9	81	729	1	1	0.666667	13.861111	1.955556	0	1.966667	0	0	
11	11/01/2010	2.152	4.328	1.661	1.621	2.536	7.602	1.579	2.690	52	109	140	31	29	1	10	100	1000	1	2	0.805556	16.333333	2.472222	3.444444	0	0	
12	12/01/2010	2.056	4.580	1.612	2.364	2.545	7.934	1.619	2.866	65	115	125	36	1	11	121	1331	1	3	0.5	10.083333	3.805556	3.866667	1.083333	0	0	
13	13/01/2010	2.099	4.612	1.590	1.673	2.574	8.165	1.596	2.888	65	114	145	35	32	1	12	144	1728	1	4	0.5	11.1111	3.486667	2.027778	0	0	
14	14/01/2010	2.259	4.910	1.659	1.984	2.492	8.281	2.787	2.866	67	109	142	44	24	1	13	169	2197	1	5	1.722222	9.638889	6.361111	0.944444	15.27778	0	0
15	15/01/2010	2.122	4.651	1.657	1.747	2.450	8.114	2.096	2.522	70	106	152	29	30	1	14	196	2744	1	6	2.138889	13.583333	2.5	15.38889	0.88889	0	0
16	16/01/2010	2.464	5.650	1.840	2.096	2.996	8.952	2.193	3.224	70	104	169	42	37	1	15	225	3375	1	7	0.633333	9.866667	2.555556	2.905556	0.25	0	0
17	17/01/2010	3.416	7.967	2.404	2.649	4.170	9.989	3.081	4.418	98	212	220	42	70	1	16	256	4096	1	1	1.972222	14.55556	1	3.75	0.333333	0	0
18	18/01/2010	2.266	4.420	1.636	1.703	2.371	7.501	1.614	2.587	62	120	144	35	28	1	17	289	4913	1	2	1	16.25	2.722222	1	1	0	0
19	19/01/2010	2.075	4.619	1.616	1.619	2.254	7.884	1.606	2.767	66	119	140	34	34	1	18	324	5932	1	3	1.25	9.02778	4.277778	16.94444	0.486667	0	0
20	20/01/2010	2.077	4.554	1.602	1.695	2.454	8.084	1.626	2.714	65	105	150	33	24	1	19	361	6859	1	4	0.633333	12.61111	3.972222	15.27778	15	0	0
21	21/01/2010	2.293	4.760	1.659	1.695	2.415	8.391	1.651	2.906	69	118	145	37	31	1	20	400	8000	1	5	1.722222	10.75	2.638889	1.593333	0.25	0	0
22	22/01/2010	2.110	4.780	1.673	1.693	2.434	8.233	1.696	2.738	68	120	170	30	26	1	21	441	9261	1	6	0.972222	10.5	5.11111	1.75	1.083333	0	0
23	23/01/2010	2.903	5.957	1.837	2.023	2.819	9.130	2.089	3.249	75	126	163	37	39	1	22	484	10648	1	7	2.5	8.883333	3.888889	1.083333	0.944444	0	0
24	24/01/2010	3.450	7.264	2.379	2.595	4.063	10.004	2.361	4.402	82	262	208	74	1	23	529	12667	1	1	0.444444	15.02778	4.277778	13.33333	1	0	0	
25	25/01/2010	2.225	4.411	1.673	1.683	2.402	7.447	1.596	2.596	63	109	148	30	20	1	24	576	13824	1	2	1.833333	9.861111	2	1	0	0	0
26	26/01/2010	2.067	5.204	1.640	1.660	2.345	8.136	1.612	2.827	68	111	162	38	27	1	25	625	15625	1	3	0.5	9.11111	5.11111	1.593333	2.444444	0	0
27	27/01/2010	2.148	4.691	1.571	1.640	2.307	7.930	1.637	2.848	65	111	161	36	32	1	26	676	17576	1	4	0.5	11.66667	2.833333	0.75	0.777778	0	0
28	28/01/2010	2.091	4.767	1.619	1.649	2.253	7.849	1.621	2.632	72	117	160	34	30	1	27	729	19683	1	5	0.611111	9.388889	4.944444	16.11111	0.5	0	0
29	29/01/2010	2.095	4.716	1.649	1.703	2.400	8.034	1.706	2.926	70	104	169	33	24	1	28	784	21952	1	6	0.5	10.933333	2.666667	15	1	0	0
30	30/01/2010	2.445	5.657	1.838	2.027	2.832	8.960	2.151	3.500	83	130	170	35	35	1	29	841	24389	1	7	0	9.277778	2.138889	15	0.75	0	0
31	31/01/2010	3.354	7.143	2.512	2.391	3.835	9.678	2.883	4.411	77	223	200	38	76	1	30	900	27000	1	1	0	9.388889	5.583333	16.11111	0.5	0	0
32	02/02/2010	2.222	4.307	1.759	1.652	2.304	7.129	1.593	2.646	62	108	147	35	31													

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 delle notizie presenti nel quotidiano 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 notizie. 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 5.

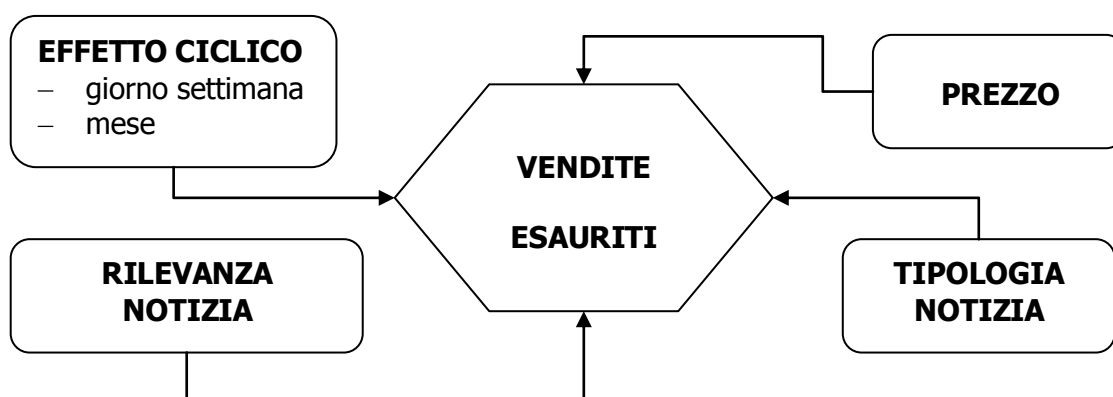


Figura 5. Rappresentazione grafica del modello statistico per lo studio delle associazioni tra le caratteristiche delle notizie presenti nel quotidiano e le vendite/esauriti.

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

$$V_i = \text{Trend}_i + \text{GSett}_i + \text{Mese}_i + \text{GSett}_i * \text{Mese}_i + \text{Prezzo}_i + \text{RilNot}_i + \text{TipoNot}_i + \varepsilon_i$$

dove V_i indica le vendite osservate il giorno i -esimo, $Trend_i$ l'eventuale trend presente nelle serie temporale delle vendite (effetto di lungo periodo), $GSett_i$ e $Mese_i$ indicano l'effetto ciclico del giorno della settimana e del mese di riferimento dello specifico giorno i -esimo, $Prezzo_i$ il prezzo di vendita del quotidiano (aumentato, a partire dal 1° gennaio 2011, dal valore di 1 ad 1,20 Euro), $RilNot_i$ e $TipoNot_i$ indicano rispettivamente la rilevanza della notizia (INT = internazionale, LOC = locale, NAZ = nazionale, REG = regionale) e la tipologia della notizia (per questa ultima classificazione, si sono adottate le categorie indicate nel paragrafo 2). Più precisamente, $RilNot_i$ e $TipoNot_i$ si riferiscono a coefficienti che misurano l'effetto di quanta parte del giornale del giorno i -esimo è stata dedicata a notizie di ciascuna prefissata categoria di rilevanza e di tipologia. Infine, ε_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 stima

La teoria della regressione lineare multipla risponde all'obiettivo di studiare la dipendenza di una variabile quantitativa Y da un insieme di m variabili esplicative quantitative X_1, \dots, X_m , dette regressori, mediante un modello lineare.

$$Y = f(X_1, \dots, X_m) + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon$$

La funzione f dipende da parametri che determinano l'influenza di ogni singolo regressore sul valore di Y . Nella formulazione del modello di regressione multipla *la linearità vale rispetto ai parametri*.

- 1) vi sono infiniti fattori che non è possibile né rilevare né considerare nella funzione f ;
- 2) nei fenomeni reali vi è un elemento imprevedibile di casualità;
- 3) i valori di Y possono essere rilevati e/o misurati con errore.

La relazione che lega Y a (X_1, \dots, X_m) non è quindi esprimibile mediante una funzione matematica, pertanto nell'equazione (2.1) viene aggiunta una variabile aleatoria ε che riassume l'effetto su Y di tutti quei fattori non inclusi nella funzione f .

Equazione del modello

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon$$

il termine $\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m$ rappresenta *la componente sistematica* del modello, la variabile casuale ε è la *componente d'errore* del modello.

I parametri (non noti) del modello sono: β_0 (*l'intercetta*), e $\beta_1, \beta_2, \dots, \beta_m$ (*i coefficienti di regressione*).

Ipotesi sui regressori (X_1, \dots, X_m)

(X_1, \dots, X_m) sono variabili deterministiche, ovvero misurate senza errore; (nell'analisi di alcuni fenomeni reali tale ipotesi non è realistica; può essere resa meno restrittiva).

Ipotesi su ε

1. L'effetto su Y di tutti i fattori non rilevati e/o non rilevabili può essere positivo o negativo; e non dipende dai valori dei regressori;
2. La variabilità dell'effetto di tutti i fattori non rilevati e/o non rilevabili non dipende dai valori dei regressori;
3. Gli effetti su Y dei fattori non rilevati per la famiglia i non dipendono da quelli relativi alla famiglia j .

Lo stimatore dei minimi quadrati ordinari dei parametri del modello

Si suppone che il modello valga nella popolazione oggetto di interesse. Poiché nella maggior parte dei casi la popolazione nel suo complesso non è direttamente rilevabile, al fine di stimare i parametri del modello ci si deve basare sulle informazioni contenute su un campione causale di n unità, su ciascuna delle quali vengono rilevati i valori della variabile Y e degli m regressori:

($Y_i, X_{i1}, \dots, X_{im}$) per $i = 1, \dots, n$.

Se il campione viene estratto con criterio casuale semplice allora gli n vettori di variabili dell'equazione sono indipendenti.

Posto il modello, per la generica osservazione campionaria vale la seguente relazione:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{im} + \varepsilon_i$$

che, formulata per ciascuna delle n unità del campione, dà luogo al seguente sistema di n equazioni in $m+1$ incognite:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \dots + \beta_m X_{1m} + \varepsilon_1$$

...

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_m X_{im} + \varepsilon_i$$

...

$$Y_n = \beta_0 + \beta_1 X_{n1} + \dots + \beta_m X_{nm} + \varepsilon_n$$

Indicando con:

y il vettore $n \times 1$ dei valori della variabile dipendente per le n unità del campione;
 X la matrice $n \times (m+1)$ dei valori degli m regressori per le n unità del campione. La matrice contiene, oltre ai valori dei regressori, una colonna supplementare composta da n valori tutti pari a 1 in corrispondenza dell'intercetta del modello.
 β il vettore $(m+1) \times 1$ dei parametri del modello; ε il vettore $n \times 1$ dei termini d'errore; il sistema può essere riscritto in maniera compatta e semplificata nella forma di un'equazione matriciale:

$$y = X\beta + \varepsilon$$

2.3 Procedura di stima del modello statistico

Nella prima analisi l'obiettivo era quello di individuare le variabili significative e dunque escludere quelle variabili che non lo erano. 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, giorno², giorno³ e numero di pagine di ogni tipo di notizia; 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 precisi. Per quanto riguarda le variabili giorno, giorno², giorno³, non sono state toccate 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 tipo di notizia e vendita di giornali) esiste una relazione lineare significativa si può verificare se β (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.

2.2 Nella seconda analisi l'obiettivo era quello di testare la significatività del modello sui 16 mesi (da gennaio 2010 a aprile 2011). Per testare la significatività del modello sui 16 mesi è necessario effettuare 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 6

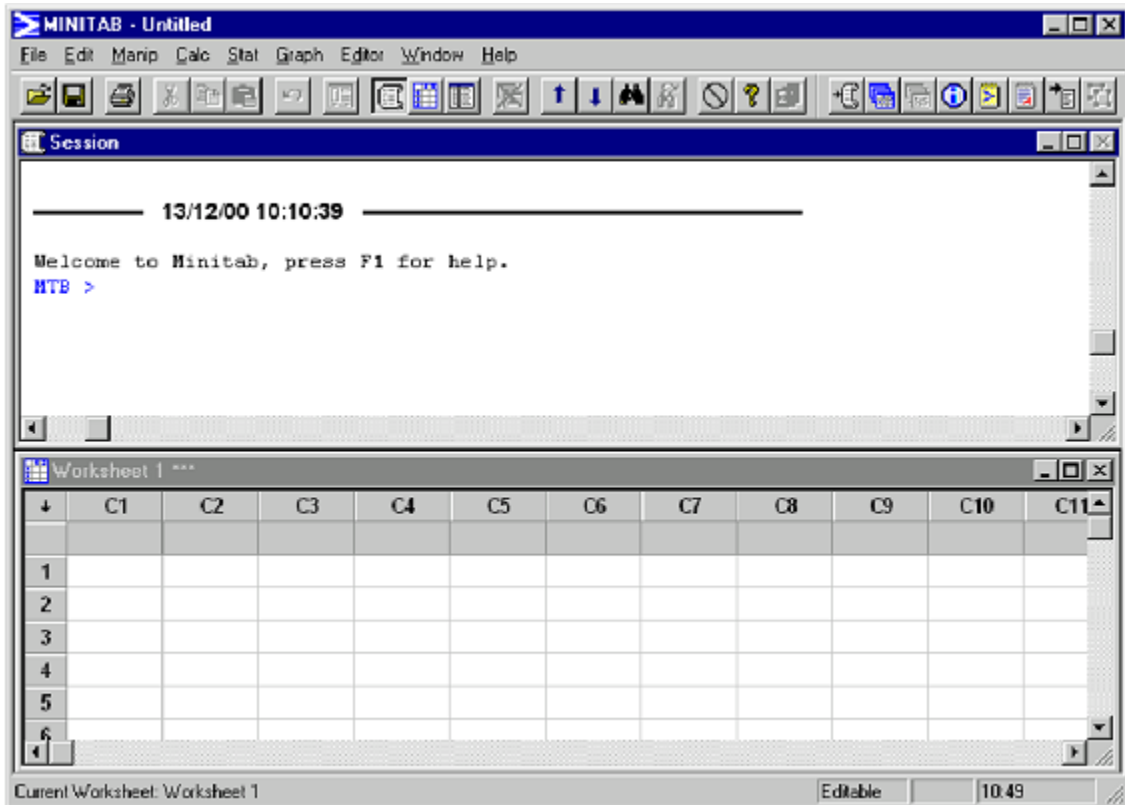


Figura 6. Schermata iniziale Minitab

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 che mostrano i dati contenuti nel Worksheet; la finestra della sessione di lavoro contenente i risultati delle elaborazioni; le finestre grafiche contenenti i grafici ad alta risoluzione prodotti durante le elaborazioni; 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

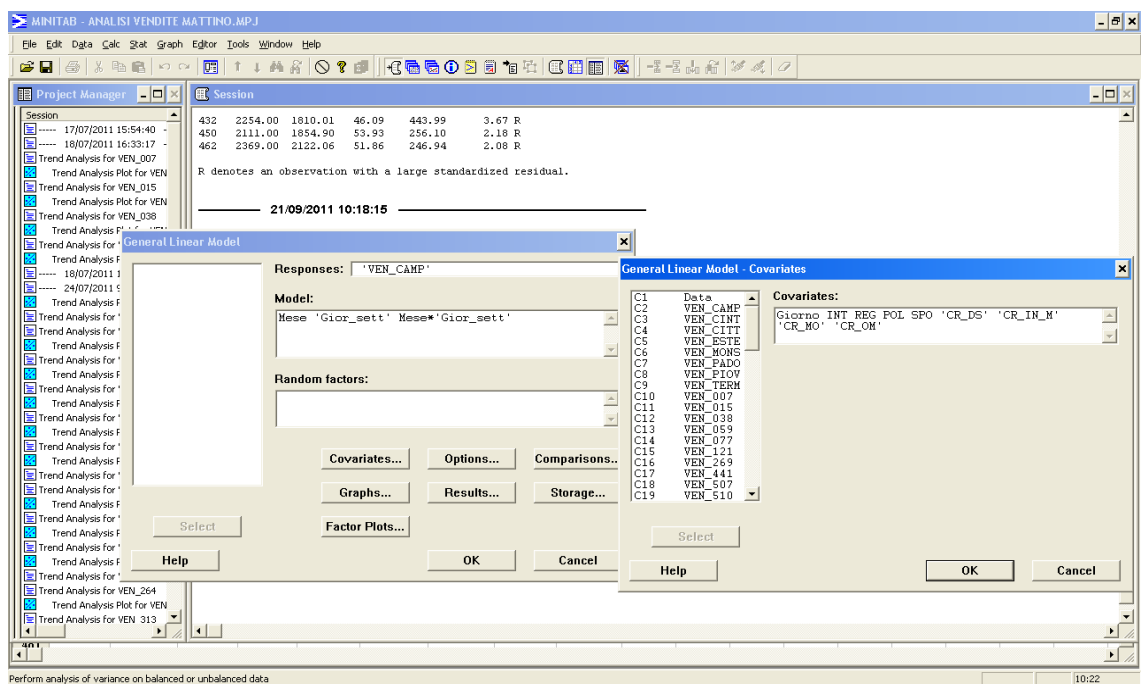


Figura 7. Processo di ricerca delle variabili significative

CAPITOLO 3. Risultati e discussione

In questo capitolo vengono presentati i risultati ottenuti in questo lavoro di tesi. In particolare, si evidenziano i grafici mostrandoti il trend delle vendite (dovuto alla variabile prezzo e al fattore ciclico) e quelli confrontanti l'andamento previsto delle vendite con quello reale.

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:

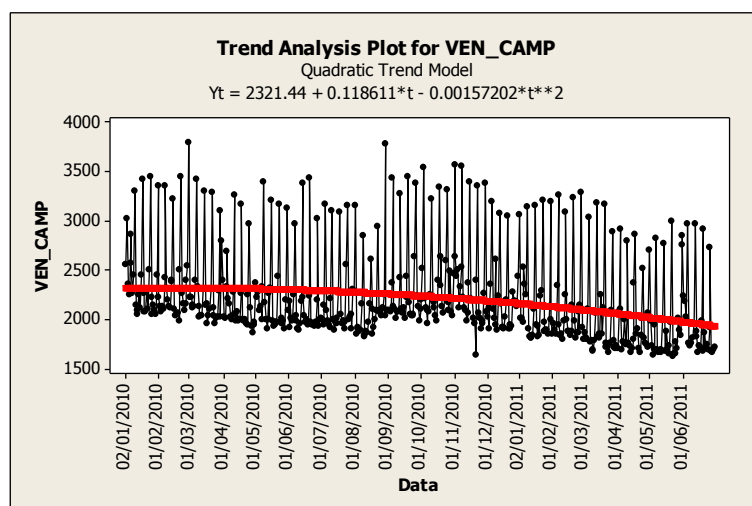


Figura 8. Serie temporale e trend analysis per le vendite dell'area Camposampierense

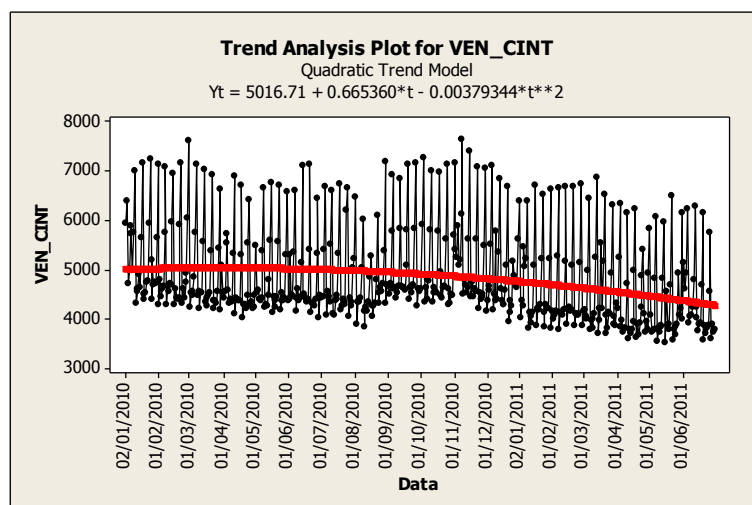


Figura 9. Serie temporale e trend analysis per le vendite della Cintura urbana

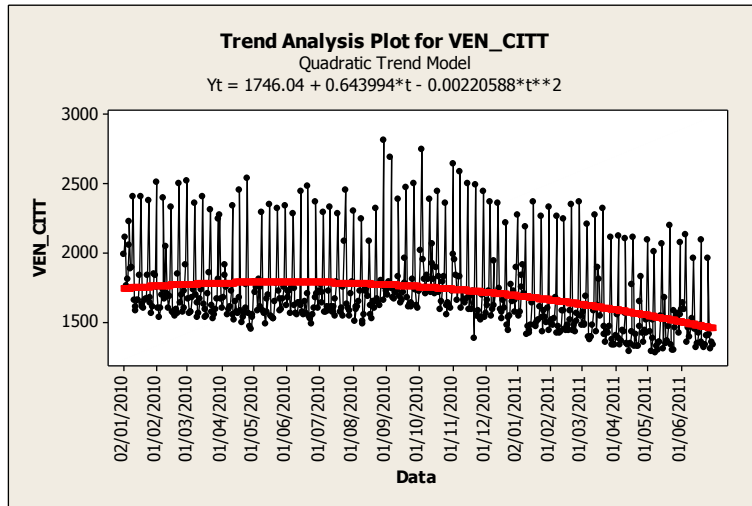


Figura 10. Serie temporale e trend analysis per le vendite dell'area Cittadense

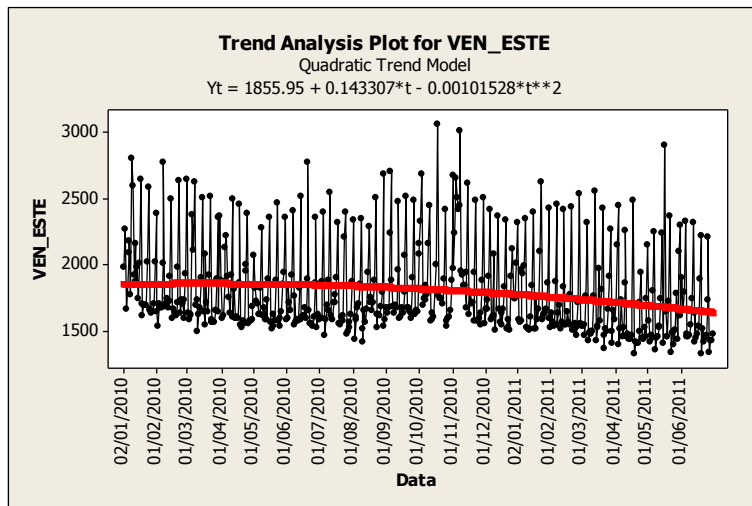


Figura 11. Serie temporale e trend analysis per le vendite dell'area Estense

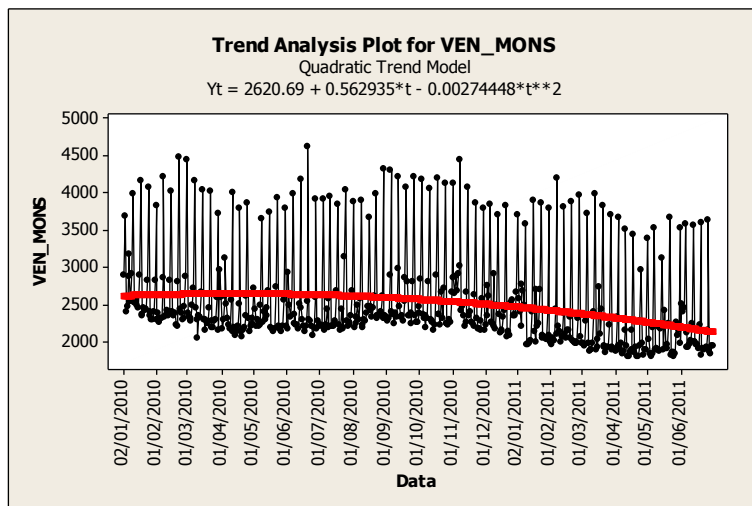


Figura 12. Serie temporale e trend analysis per le vendite dell'area Monselicense

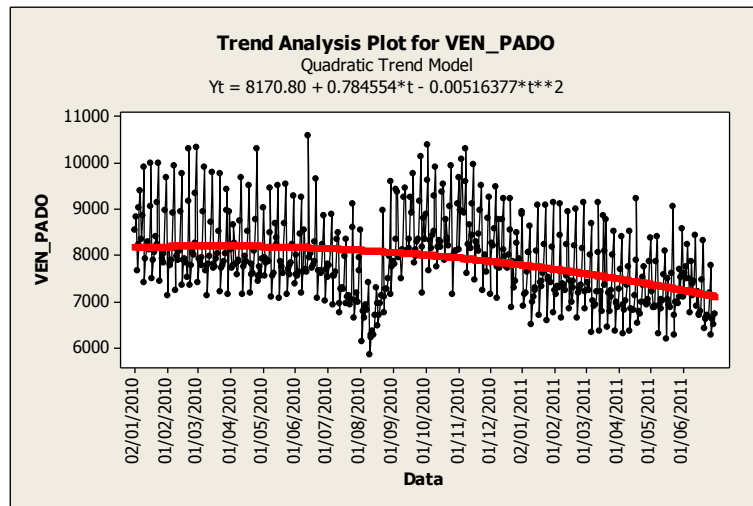


Figura 13. Serie temporale e trend analysis per le vendite di Padova

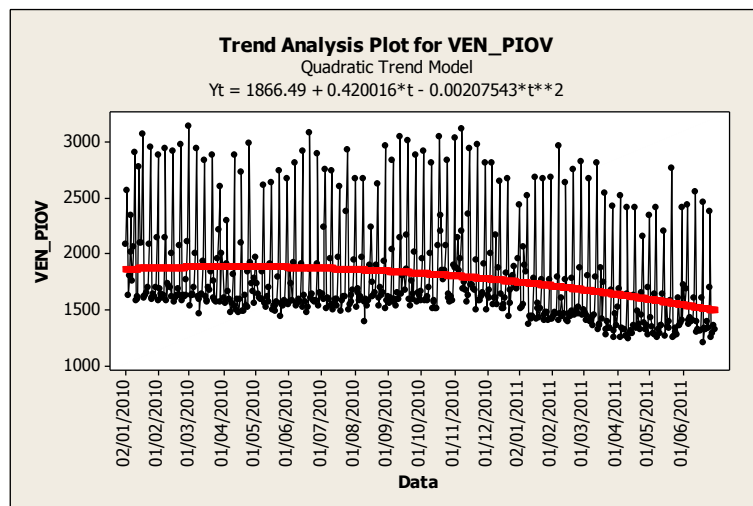


Figura 14. Serie temporale e trend analysis per le vendite di Piovese

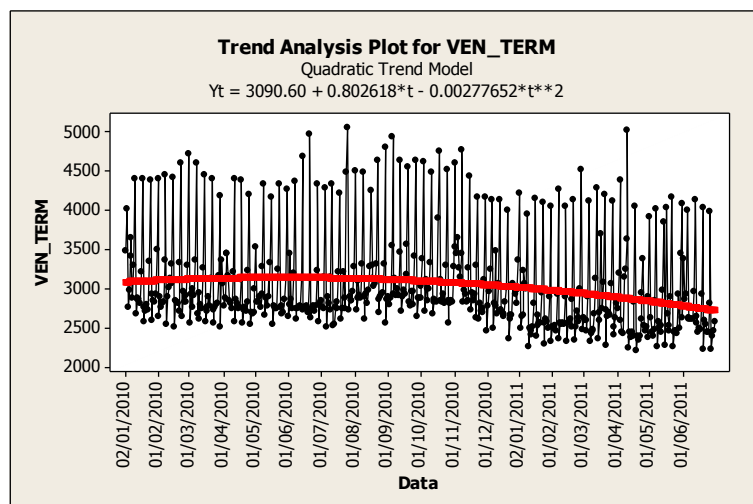


Figura 15. Serie temporale e trend analysis per le vendite di Terme Euganee

Questi sono i trend delle edicole:

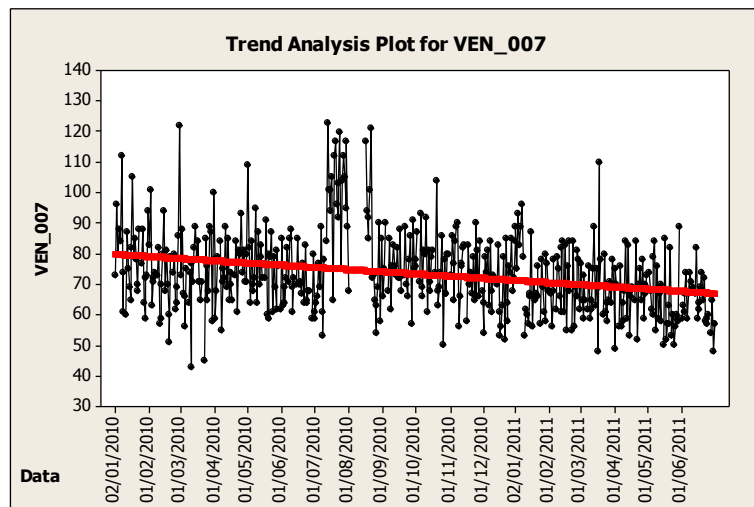


Figura 16. Serie temporale e trend analysis per le vendite dell'edicola 007

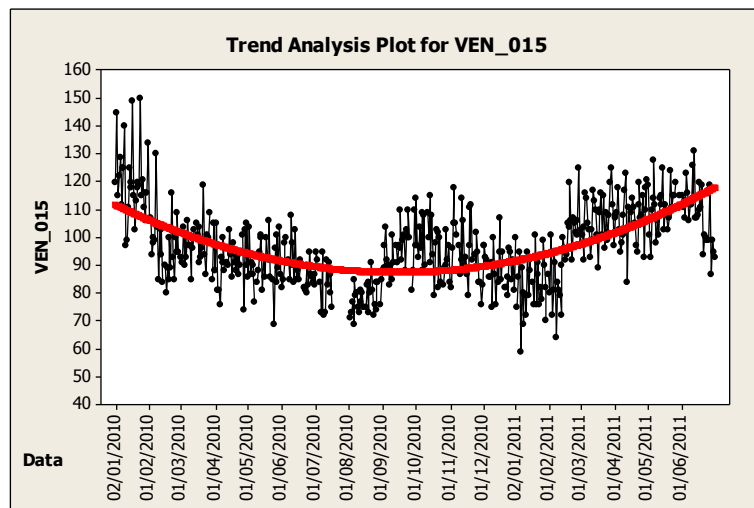


Figura 17. Serie temporale e trend analysis per le vendite dell'edicola 015

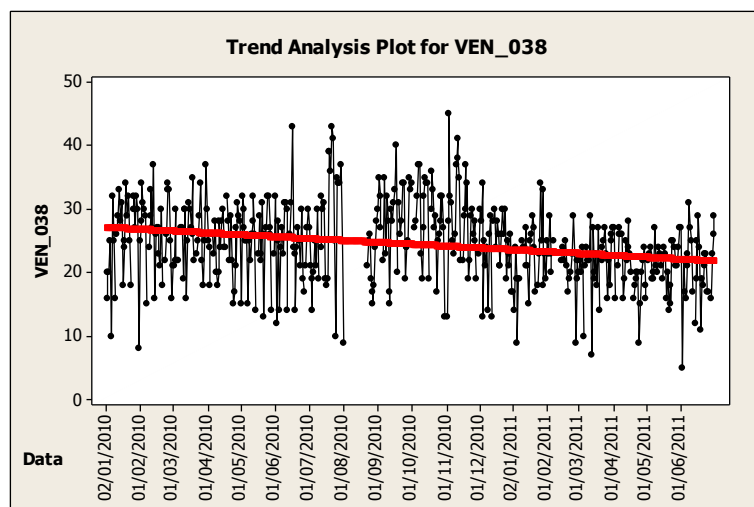


Figura 18. Serie temporale e trend analysis per le vendite dell'edicola 038

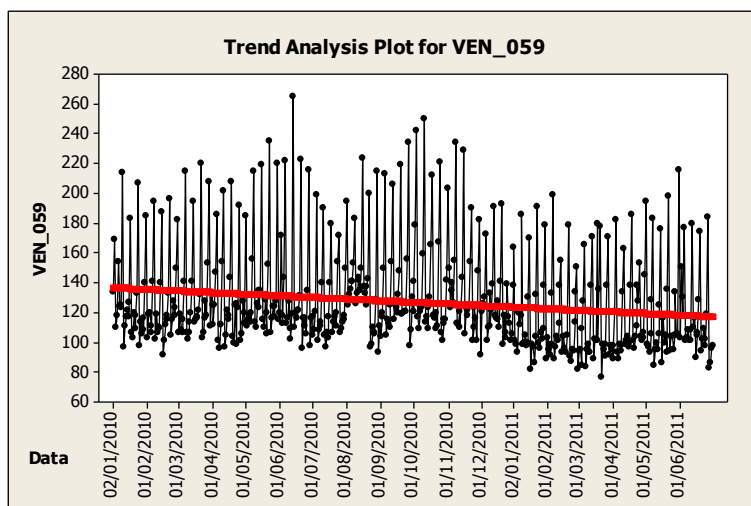


Figura 19. Serie temporale e trend analysis per le vendite dell'edicola 059

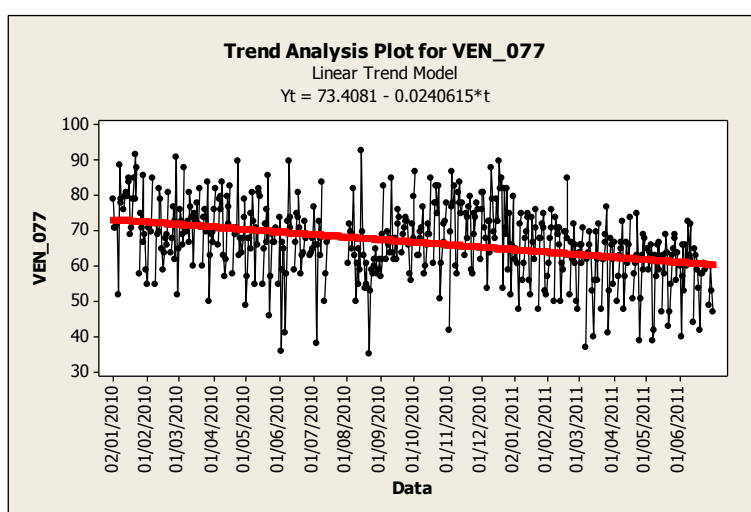


Figura 20. Serie temporale e trend analysis per le vendite dell'edicola 077

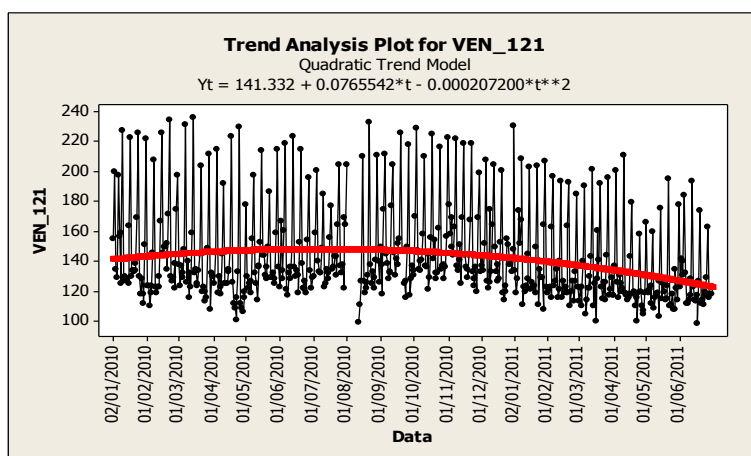


Figura 21. Serie temporale e trend analysis per le vendite dell'edicola 121

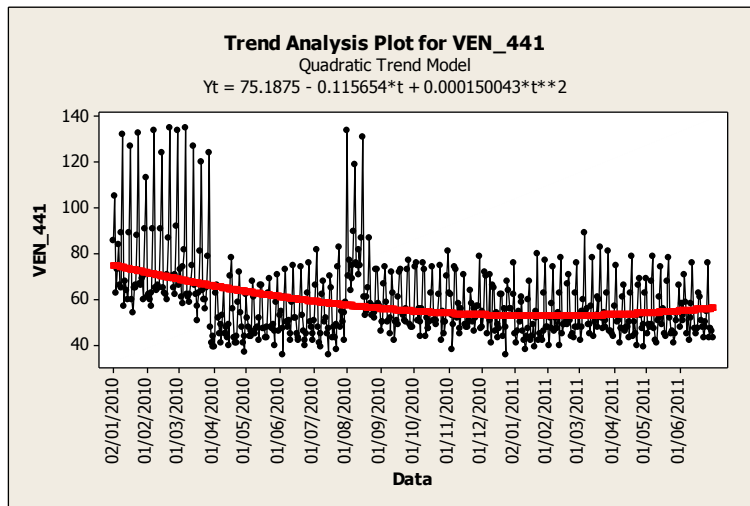


Figura 22. Serie temporale e trend analysis per le vendite dell'edicola 441

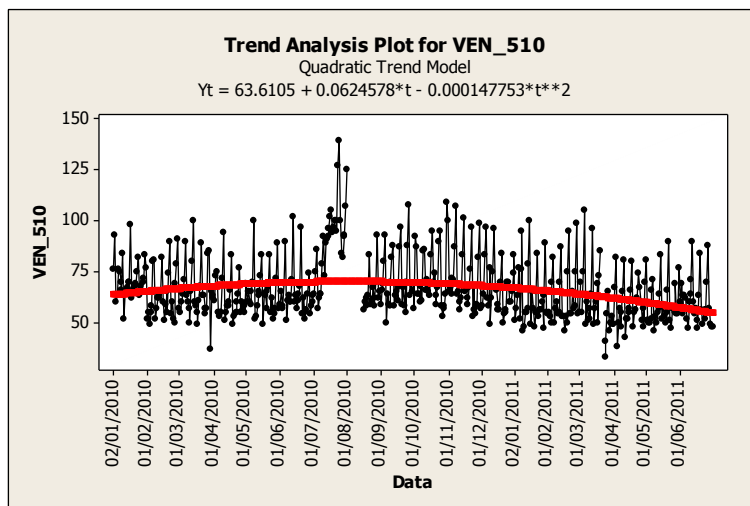


Figura 23. Serie temporale e trend analysis per le vendite dell'edicola 510

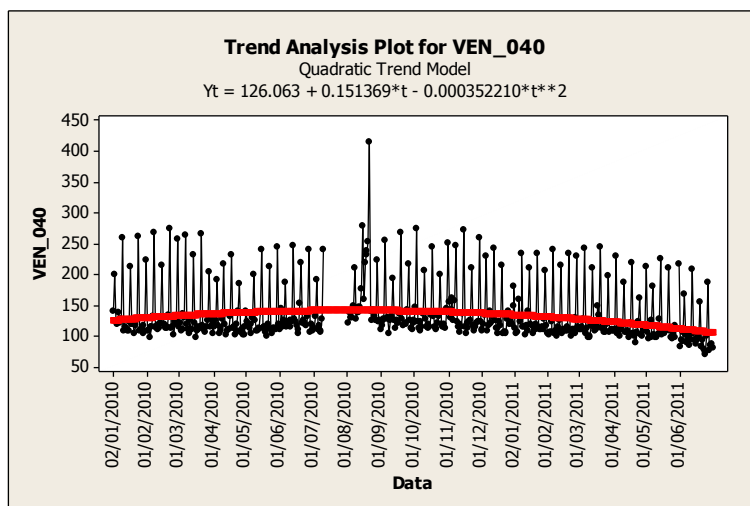


Figura 24. Serie temporale e trend analysis per le vendite dell'edicola 040.

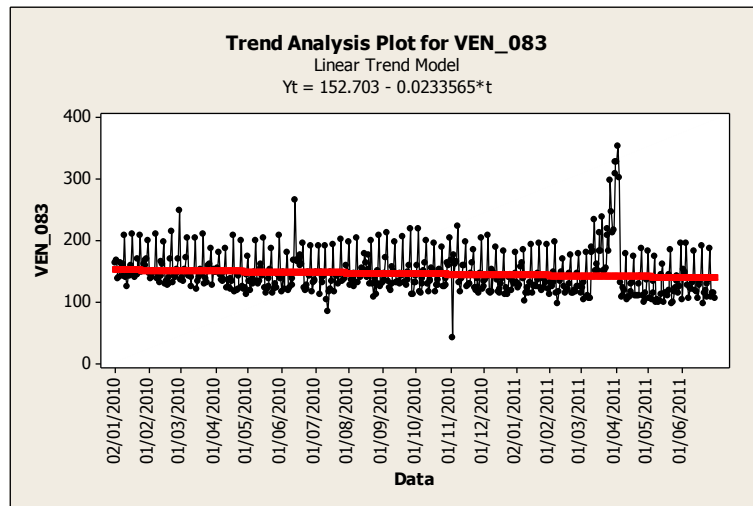


Figura 25. Serie temporale e trend analysis per le vendite dell'edicola 083

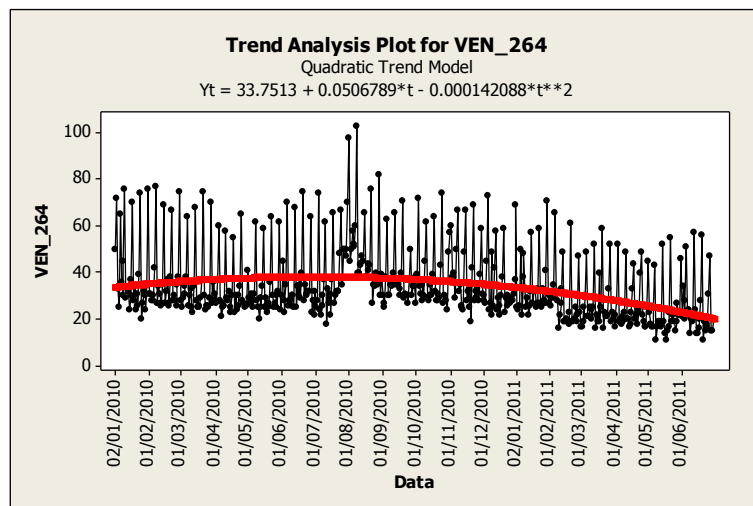


Figura 26. Serie temporale e trend analysis per le vendite dell'edicola 264

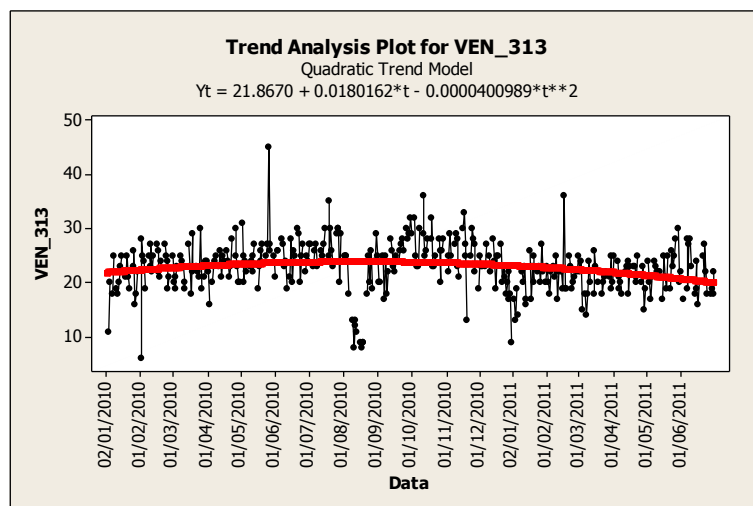


Figura 27. Serie temporale e trend analysis per le vendite dell'edicola 313

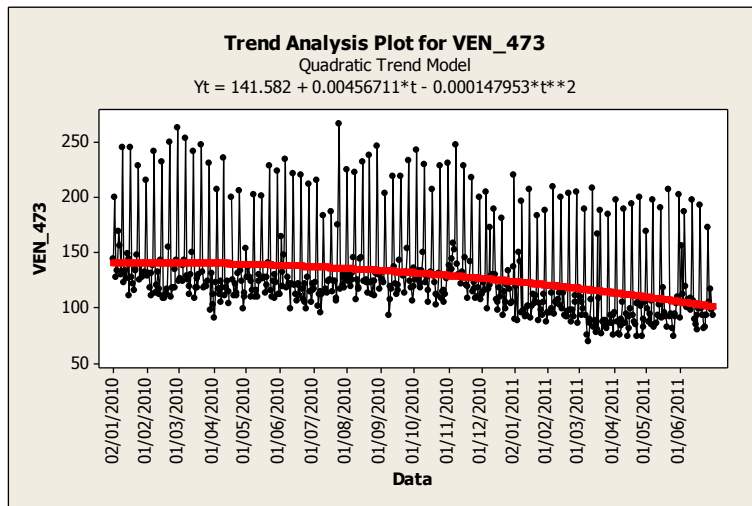


Figura 28. Serie temporale e trend analysis per le vendite dell'edicola 473

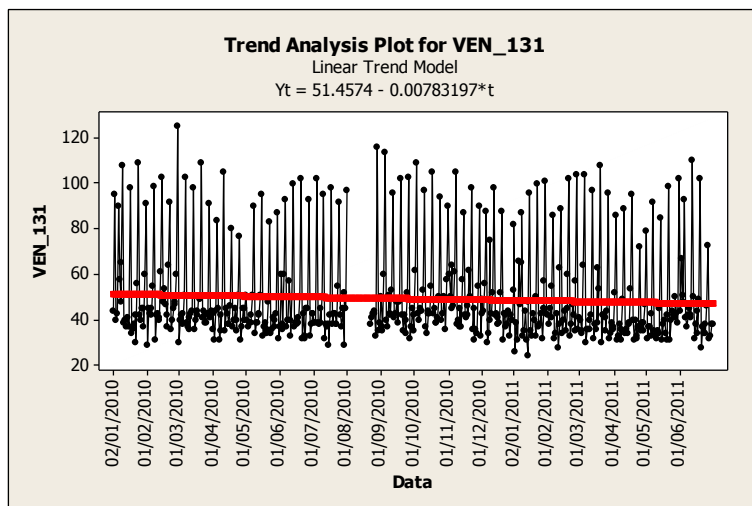


Figura 29. Serie temporale e trend analysis per le vendite dell'edicola 131

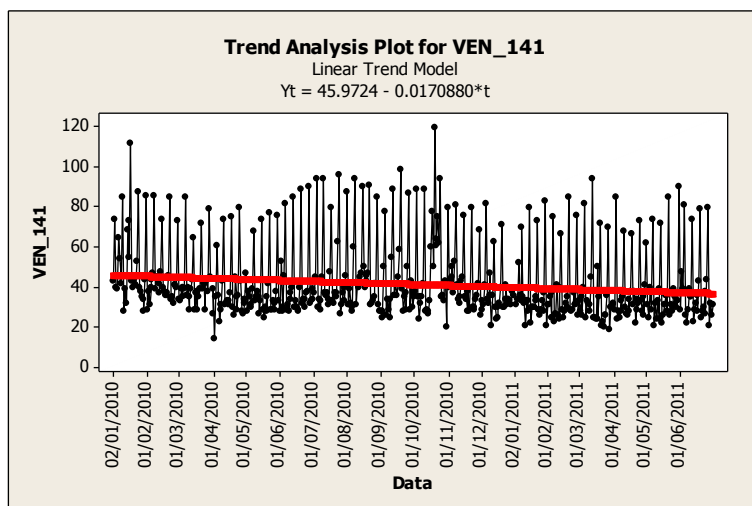


Figura 30. Serie temporale e trend analysis per le vendite dell'edicola 141

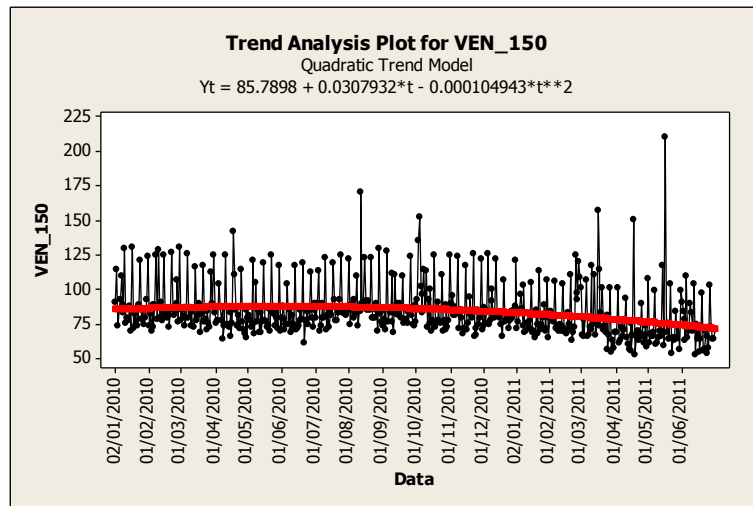


Figura 31. Serie temporale e trend analysis per le vendite dell'edicola 150

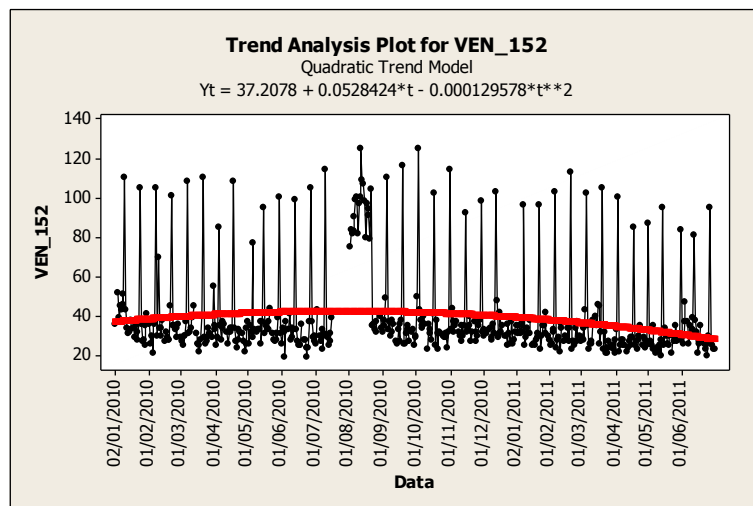


Figura 32. Serie temporale e trend analysis per le vendite dell'edicola 152

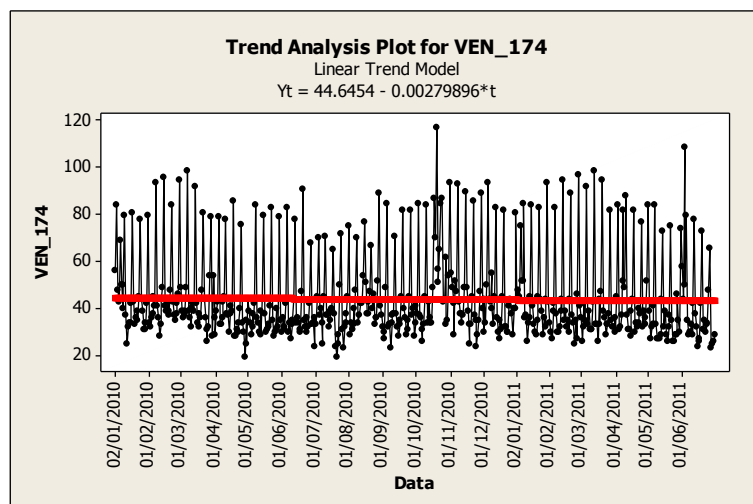


Figura 33. Serie temporale e trend analysis per le vendite dell'edicola 174

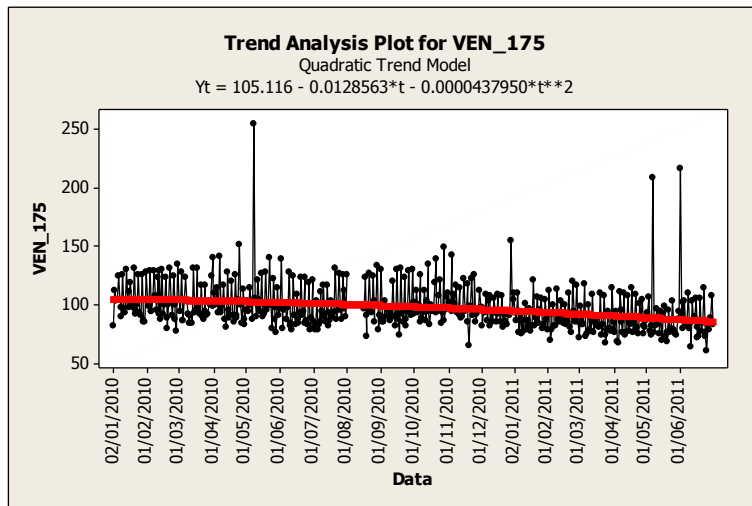


Figura 34. Serie temporale e trend analysis per le vendite dell'edicola 175

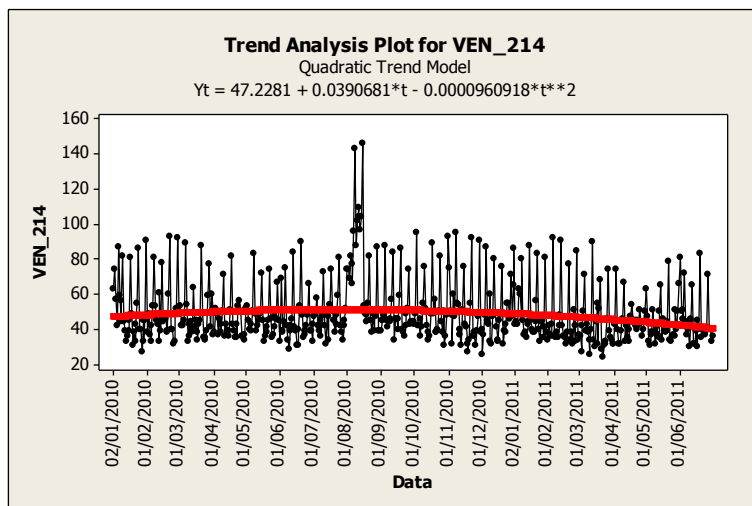


Figura 35. Serie temporale e trend analysis per le vendite dell'edicola 214

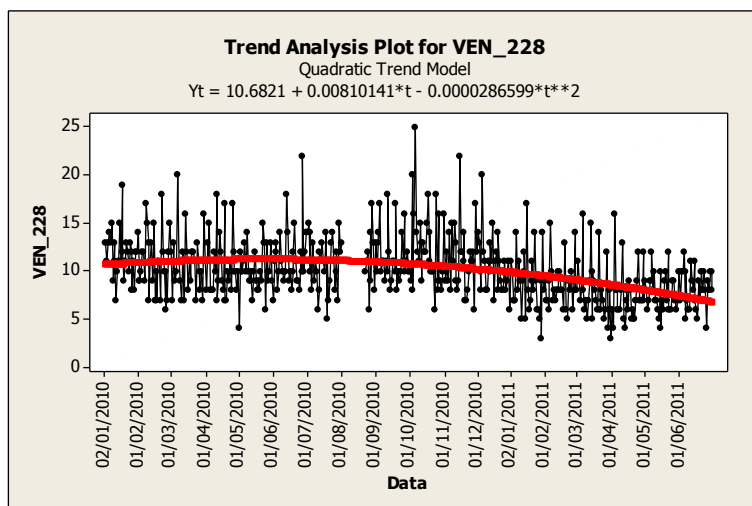


Figura 36. Serie temporale e trend analysis per le vendite dell'edicola 228

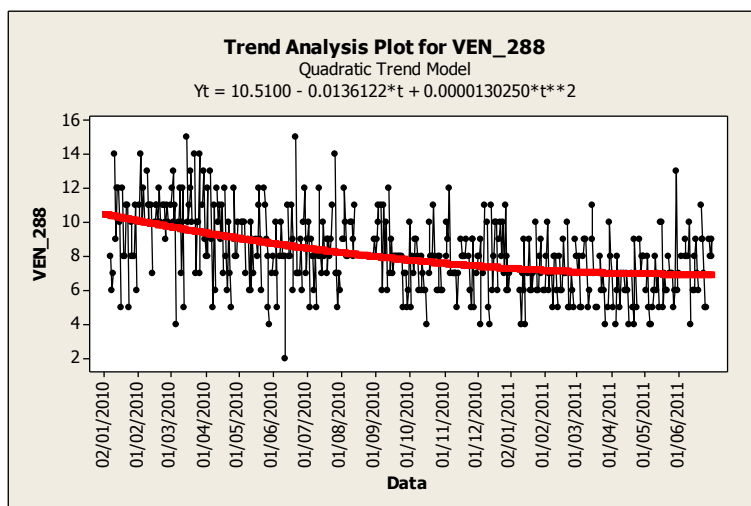


Figura 37. Serie temporale e trend analysis per le vendite dell'edicola 288

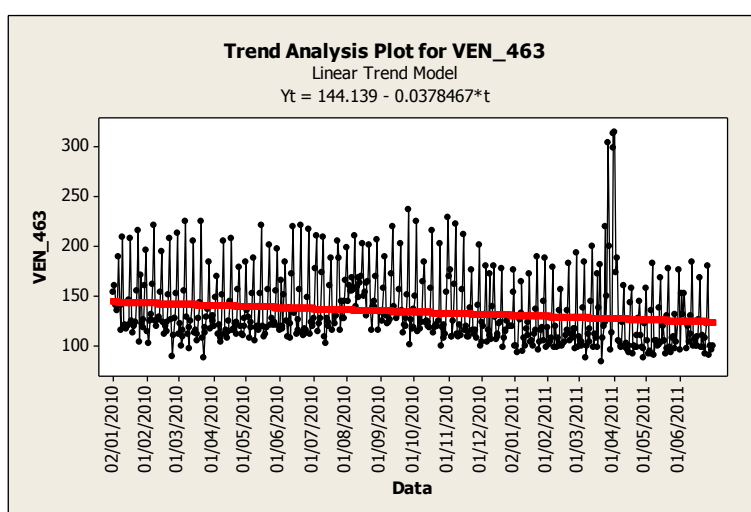


Figura 38. Serie temporale e trend analysis per le vendite dell'edicola 463

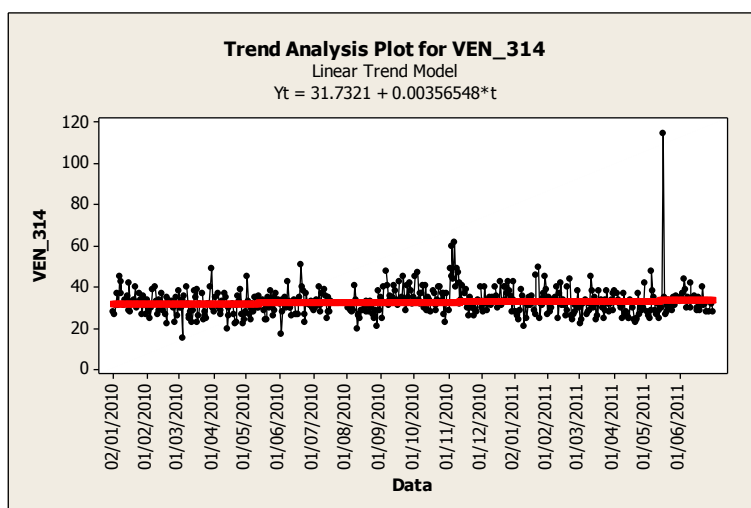


Figura 39. Serie temporale e trend analysis per le vendite dell'edicola 314

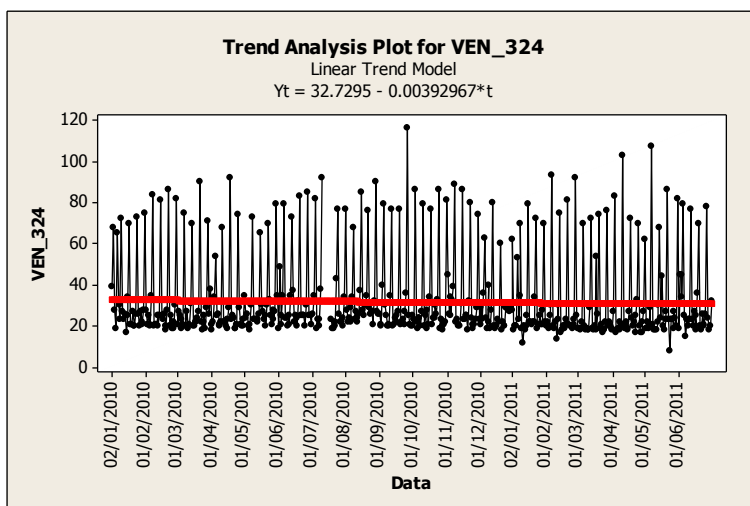


Figura 40. Serie temporale e trend analysis per le vendite dell'edicola 324

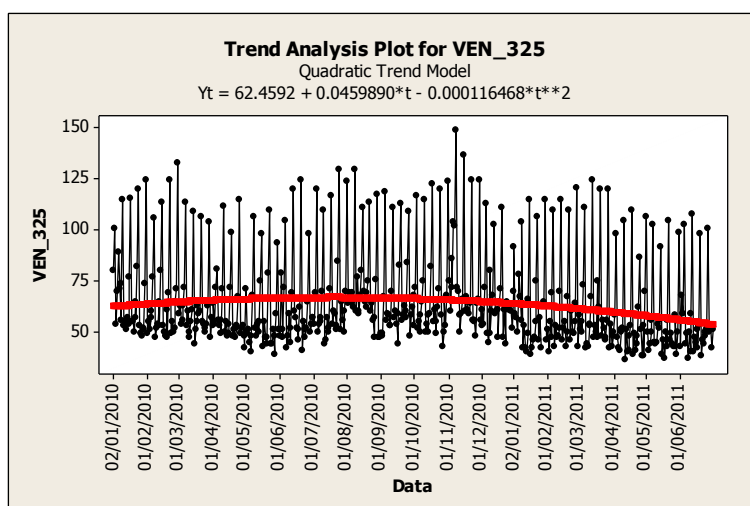


Figura 41. Serie temporale e trend analysis per le vendite dell'edicola 325

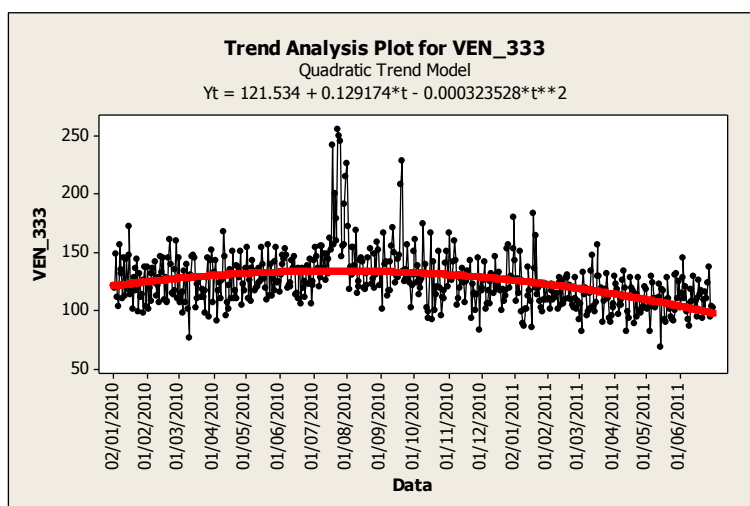


Figura 42. Serie temporale e trend analysis per le vendite dell'edicola 333

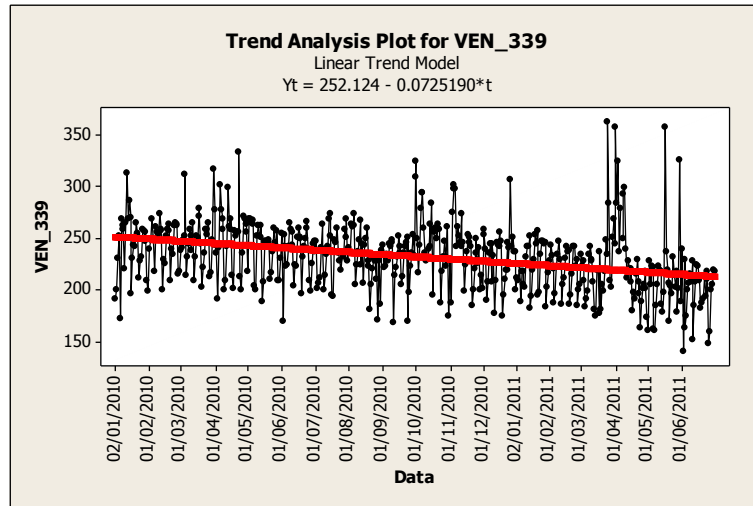


Figura 43. Serie temporale e trend analysis per le vendite dell'edicola 339

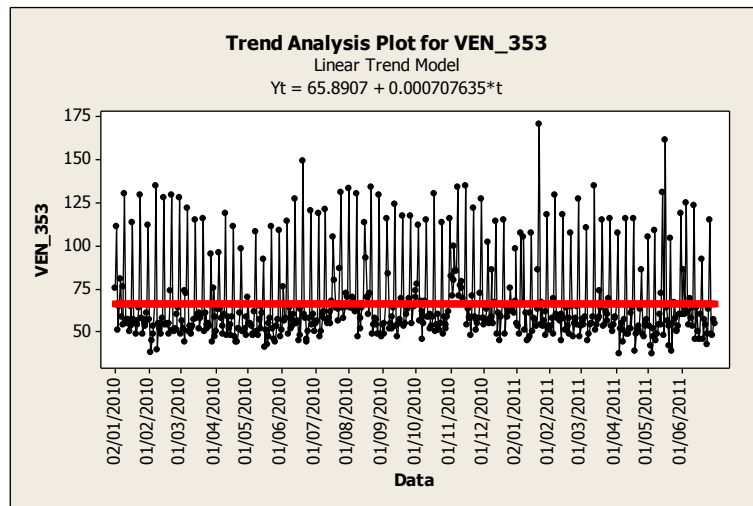


Figura 44. Serie temporale e trend analysis per le vendite dell'edicola 353

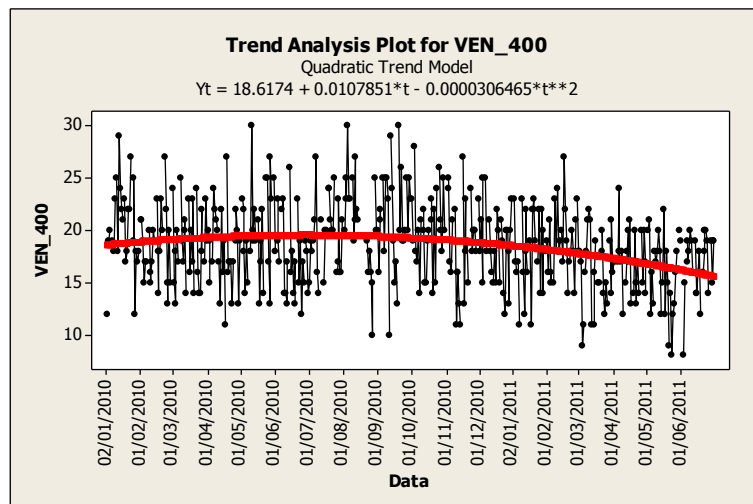


Figura 45. Serie temporale e trend analysis per le vendite dell'edicola 400

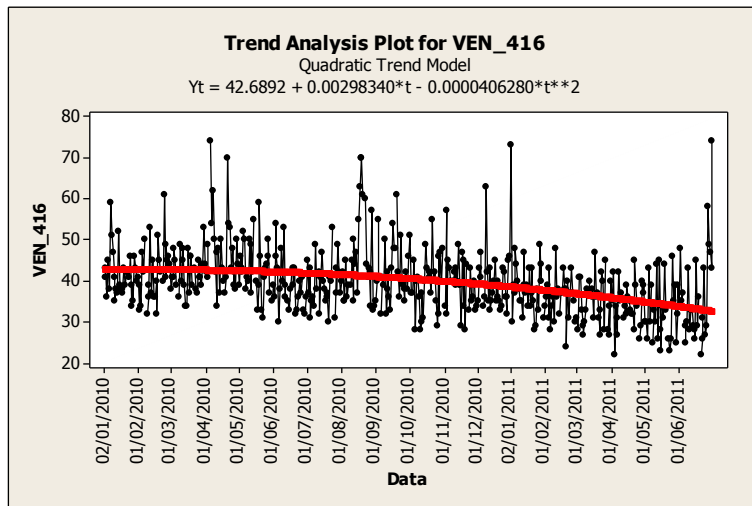


Figura 46. Serie temporale e trend analysis per le vendite dell'edicola 416

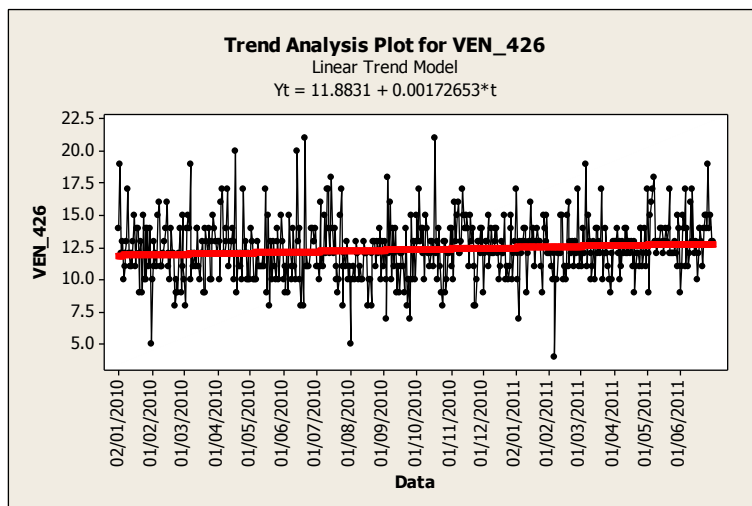


Figura 47. Serie temporale e trend analysis per le vendite dell'edicola 426

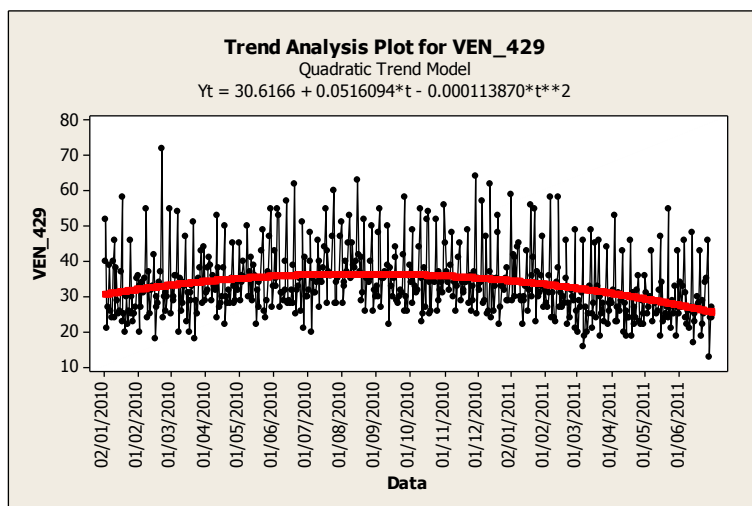


Figura 48. Serie temporale e trend analysis per le vendite dell'edicola 429

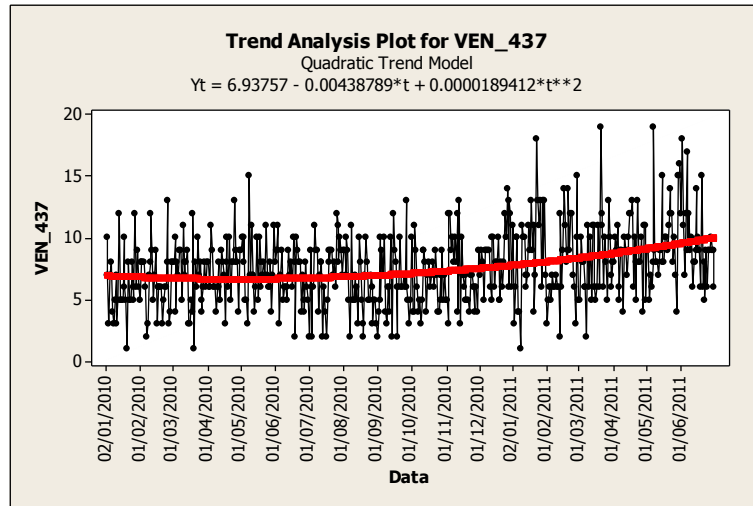


Figura 49. Serie temporale e trend analysis per le vendite dell'edicola 437

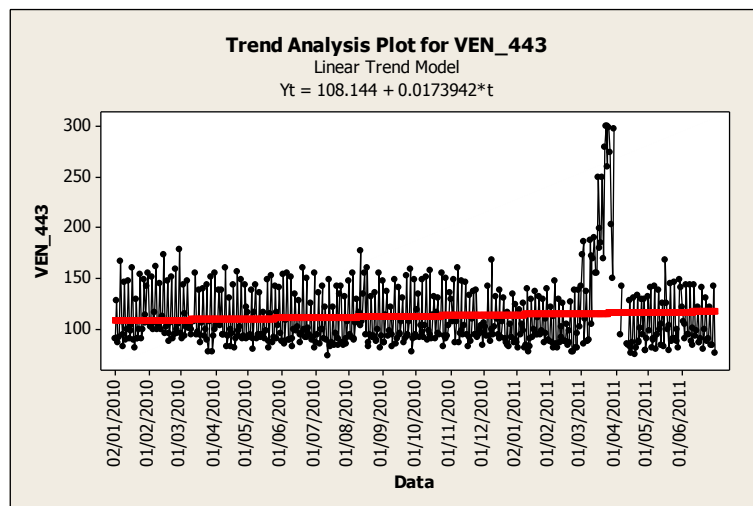


Figura 50. Serie temporale e trend analysis per le vendite dell'edicola 443

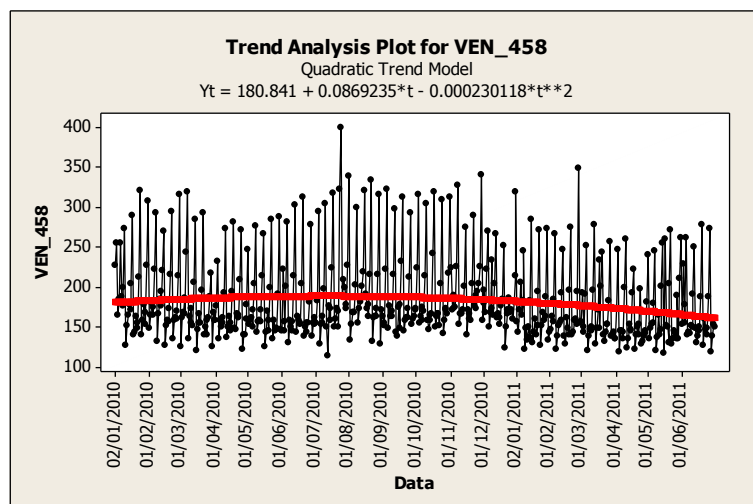


Figura 51. Serie temporale e trend analysis per le vendite dell'edicola 458

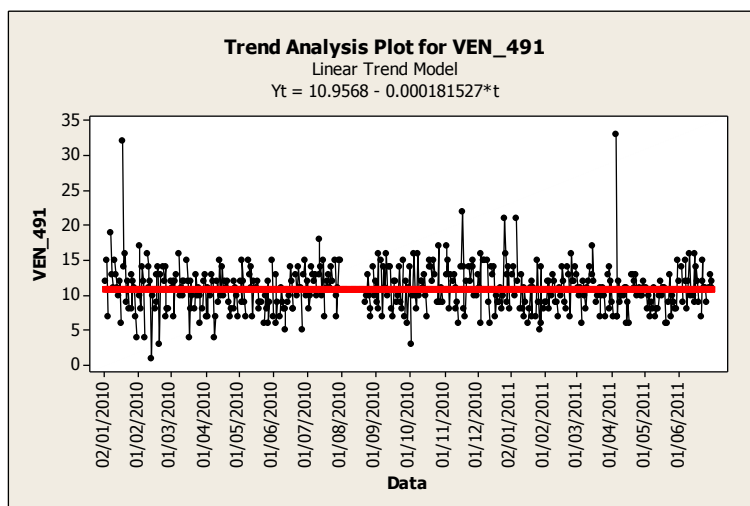


Figura 52. Serie temporale e trend analysis per le vendite dell'edicola 491

Come si può notare dai grafici, il trend 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, come c'era da aspettarsi il cambio di pendenza si verifica agli inizi del 2011, ovvero quando il prezzo del quotidiano aumenta. In alcuni casi isolati 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.

Est_Coeff	INT	LOC	NAZ	REG	ECO	EVE	POL	SPO	TRA	CR_NC	CR_ARR
VEN_CAMP	-10.727	0	0	35.458	0	0	23.953	17.818	0	0	0
VEN_CINT	0	0	0	32.17	0	0	14.29	33.02	0	0	0
VEN_CITT	0	-11.718	-7.421	15.776	0	0	23.744	28.721	0	11.602	25.01
VEN_ESTE	0	0	0	25.1	0	0	25.231	0	0	0	0
VEN_MONS	0	0	0	17.23	0	0	25.604	11.074	0	0	0
VEN_PADO	0	0	0	46.45	0	0	46.17	66.97	0	0	0
VEN_PIOV	0	0	0	11.333	0	0	24.896	17.843	0	0	17.47
VEN_TERM	-20.26	7.006	0	30.37	0	0	12.36	0	0	0	0
VEN_007	-1.8019	0	0	0	0	0	1.3133	1.4025	0	0	0
VEN_015	0	0	0	0	0	0	0	1.1299	0	0	0
VEN_038	0	0	0.4491	0	0	0	0	0	1.053	0	0

Figura 53. Tabella dei coefficienti di regressione

Est_Coeff	CR_AV	CR_DS	CR_FU	CR_GIU	CR_IN	CR_IN_M	CR_MO	CR_OM	CR_SC	CR_SU	CR_VI_SES
VEN_CAMP	0	32.656	0	0	0	30.145	22.31	29.35	0	0	0
VEN_CINT	0	72.68	0	0	0	0	0	0	-41.57	0	0
VEN_CITT	29.38	29.337	0	0	42.78	17.548	39.85	22.14	0	0	0
VEN_ESTE	0	62.221	0	0	0	43.535	0	54.85	0	0	0
VEN_MONS	31.82	55.595	0	0	0	30.59	0	47.64	-27.07	0	0
VEN_PADO	0	104.29	0	0	0	55.59	0	0	0	0	0
VEN_PIOV	0	34.813	0	0	0	55.637	0	0	0	0	0
VEN_TERM	0	57.51	0	0	0	0	0	58.59	0	0	0
VEN_007	2.495	1.4271	0	0	0	0	6.448	0	0	0	0
VEN_015	1.843	1.3724	0	0	0	0	0	0	0	0	0
VEN_038	0	0.6372	-1.2865	0	0	0	0	0	0.7243	0	0

Figura 54. Tabella dei coefficienti di regressione

Est_Coeff	INT	LOC	NAZ	REG	ECO	EVE	POL	SPO	TRA	CR_NC	CR_ARR
VEN_059	-1.3028	0	0	0	0	1.25	0.9852	1.298	1.79	0	3.084
VEN_077	0	0	0	0	0	-1.1516	0	0	0	0	0
VEN_121	-1.8382	0	0	0	0	3.6298	0	0	0	0	0
VEN_441	-1.2893	0	0	0	0	0	0	0	0	0	0
VEN_510	-1.5709	0.3883	-0.9992	0	0	0	0	0	-1.796	-0.5728	0
VEN_040	0	0	0	0	0	0	1.536	1.3733	0	0	0
VEN_083	0	0	0	2.358	0	0	0	0	0	0	0
VEN_264	-0.8878	0	0	0	0	0	0	0	0	0	0
VEN_313	0	0	0	0	0.7134	0	0	0.2751	0	0	0
VEN_473	-1.8375	0	0	0.9479	0	0	0	0.8329	0	0	2.013
VEN_131	0	0.6199	0	0.5394	0	0	0	0	0	0	0

Figura 55. Tabella dei coefficienti di regressione

Est_Coeff	CR_AV	CR_DS	CR_FU	CR_GIU	CR_IN	CR_IN_M	CR_MO	CR_OM	CR_SC	CR_SU	CR_VI_SES
VEN_059	3.544	1.6617	0	0	0	0	2.998	0	0	0	0
VEN_077	0	0.7093	0	0	0	0	-2.096	0	0	0	0
VEN_121	0	1.9524	0	0	0	0	2.189	0	0	0	0
VEN_441	0	0	0	0	0	0	0	0	0	0	0
VEN_510	0	1.163	0	0	0	0	0	-1.5754	0	-0.906	0
VEN_040	5.492	2.4863	0	-4.077	0	0	0	0	0	0	0
VEN_083	0	0	0	0	0	0	0	0	-3.373	0	0
VEN_264	0	0.5957	0	0	0	0	0	0	-1.0398	0	0
VEN_313	0	0	0	0	0	0	0	0	1.2011	0	0
VEN_473	5.49	3.1734	0	0	0	0	0	0	0	0	7.616
VEN_131	0	0	-1.5401	0	0	0	0	0	-1.355	0	0

Figura 56. Tabella dei coefficienti di regressione

Est_Coeff	INT	LOC	NAZ	REG	ECO	EVE	POL	SPO	TRA	CR_NC	CR_ARR
VEN_141	0	0.5864	0	0	0	0	0	0	0	0	0
VEN_150	0	0	0	0	0	0	1.0955	0	0	0	2.306
VEN_152	0	0	-0.712	0	0	0	1.0691	0	0	0	1.7747
VEN_174	0	0.2972	0	0	0	0	0	0	0	0	-1.7579
VEN_175	-0.9519	0.9245	0	2.2509	0	0	0	0	0	0	0
VEN_214	0	0.6979	0	1.1909	0	0	0	-0.4865	0	0	0
VEN_228	0	0	0	0	0	0	0	0	0	0	0
VEN_288	0	0	0	0	0	0	0	0	0	0	0
VEN_463	0	0	0	0	0	0	0	0	0	0	0
VEN_314	0	0	0	0	0	0	0.3549	0	0	0	0
VEN_324	-0.6804	0	0	0.8058	0	0	0	0.4796	0	0	0

Figura 57. Tabella dei coefficienti di regressione

Est_Coeff	CR_AV	CR_DS	CR_FU	CR_GIU	CR_IN	CR_IN_M	CR_MO	CR_OM	CR_SC	CR_SU	CR_VI_SES
VEN_141	0	0	0	0	0	3.4096	0	0	-1.9361	0	0
VEN_150	0	0	0	0	0	0	0	1.8837	0	0	0
VEN_152	0	0	0	0	0	0	0	0	0	0	0
VEN_174	0	0	0	0	0	2.9992	0	0	-1.3582	0	0
VEN_175	0	0	0	0	3.419	0	0.438	0	0	0	0
VEN_214	0	0	0	0	-5.036	0	0	0	-2.0341	0	0
VEN_228	0	0	0	0	0	0	0	0	0	0	0
VEN_288	0	0	0	0.4953	0	0	0	0	0	0	-1.3061
VEN_463	0	0	-6.3	0	0	0	0	0	0	0	0
VEN_314	0	1.9749	0	0	0	0	0	0	0	0	0
VEN_324	0	0.6669	0	0	1.535	0	0	0	0	0	0

Figura 58. Tabella dei coefficienti di regressione

Est_Coeff	INT	LOC	NAZ	REG	ECO	EVE	POL	SPO	TRA	CR_NC	CR_ARR
VEN_325	0	0	0	0	0	0	0	0.8554	0	0	1.2472
VEN_333	-1.725	1.2245	0	0	0	3.749	0	-1.8514	-4.809	0	0
VEN_339	0	0	0	2.68	0	0	1.773	0	0	0	3.061
VEN_353	0	0	0	0	0	0	0.7654	0.4355	0	0	1.6762
VEN_400	0	0	0	0	0	0	0	0	-0.9368	0	0
VEN_416	0	0	0	0	-0.9752	0	0	0	0	0	0
VEN_426	0	0	0	0	-0.473	0	0	0	0	0	-0.2832
VEN_429	-0.8679	0	-0.8532	0.4815	0	0.7668	0	0	0	0	0
VEN_437	0	0.11436	0	0	-0.6164	-0.3321	0	0	0	-0.3432	0
VEN_443	7.626	1.6489	0	0	0	0	0	0	0	0	0
VEN_458	0	0	0	0	0	0	-0.856	0	0	0	0
VEN_491	0	0	0	0	0	0	0	0.5205	0	0	0

Figura 59. Tabella dei coefficienti di regressione

Est_Coeff	CR_AV	CR_DS	CR_FU	CR_GIU	CR_IN	CR_IN_M	CR_MO	CR_OM	CR_SC	CR_SU	CR_VI_SES
VEN_325	0	2.6262	0	0	0	0	0	0	-0.8802	0	0
VEN_333	0	3.1246	0	0	0	0	0	0	-2.898	0	0
VEN_339	0	5.557	0	0	0	0	0	5.108	0	0	0
VEN_353	0	2.6826	0	0	0	0	1.728	2.7889	-1.598	0	0
VEN_400	0	0	0	0	0	0	0	0	0	0	0
VEN_416	1.9765	0	0	0	0	0	0	1.4455	0	0	-5.07
VEN_426	0.6082	0	0	0	0	0	0	0	0	-0.9254	0
VEN_429	0	0	0	0	0	0	0	1.9416	0	0	0
VEN_437	0	0.2849	0	0	0	0	0	0	0	0	0
VEN_443	0	0	0	0	0	-2.821	0	0	-4.97	0	0
VEN_458	0	2.238	0	0	0	0	0	0	0	0	13.734
VEN_491	0	0	0	0	0	0	1.1091	0	0	0	0

Figura 60. Tabella dei coefficienti di regressione

In questa tabella sono presenti sulle colonne tutte le tipologie di notizie considerate e sulle righe ogni area e edicola esaminata. Ogni qualvolta si trova uno zero significa che in quell'edicola/area il tipo di notizia considerata non è significativa per le vendite di giornali. Se si riscontra un coefficiente negativo significa che la tipologia di notizia considerata 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:

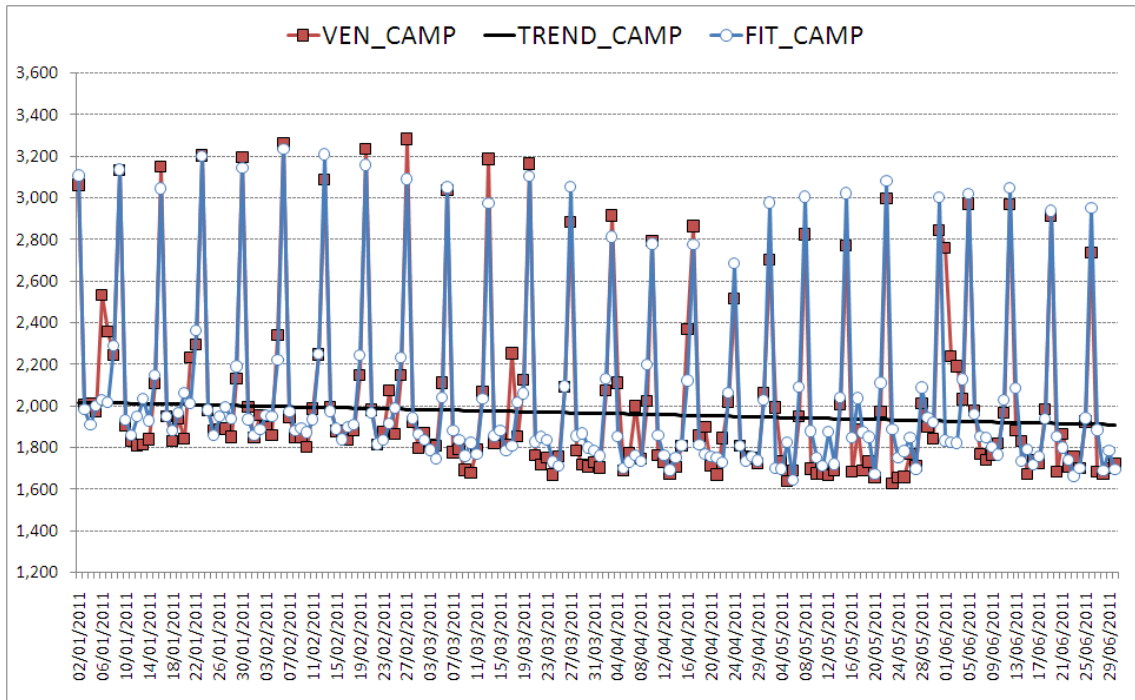


Figura 61. Vendite reali, trend e vendite previste per l'area Camposampierese

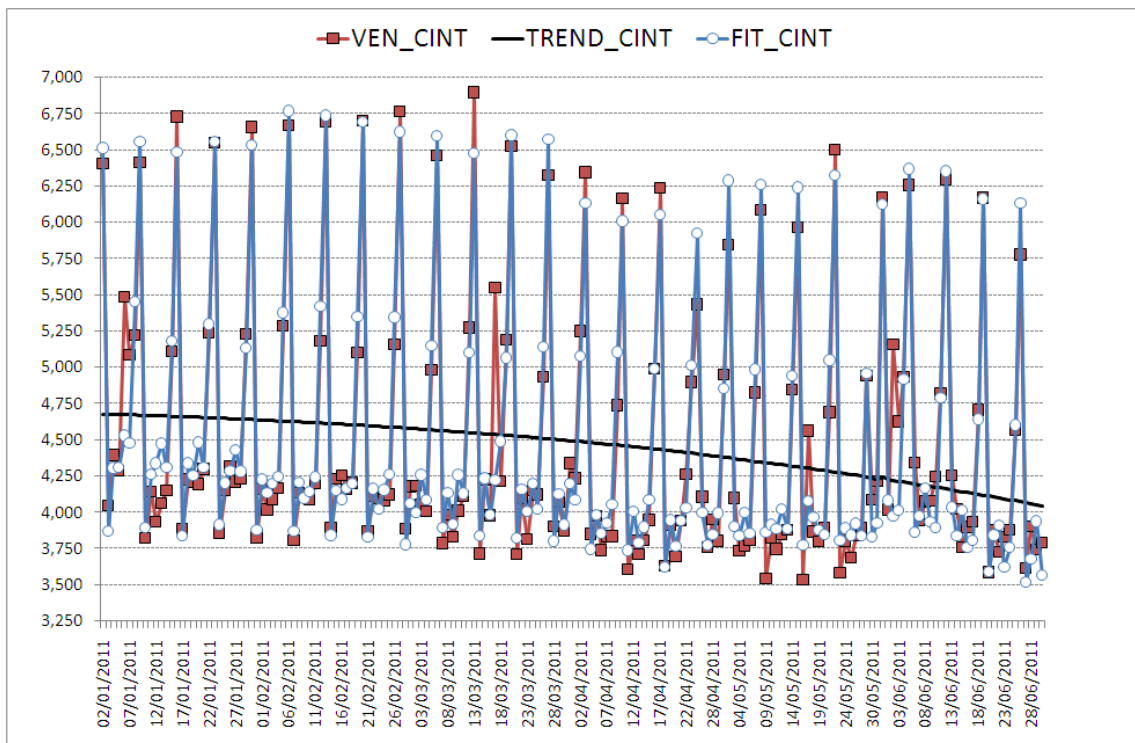


Figura 62. Vendite reali, trend e vendite previste per la Cintura urbana

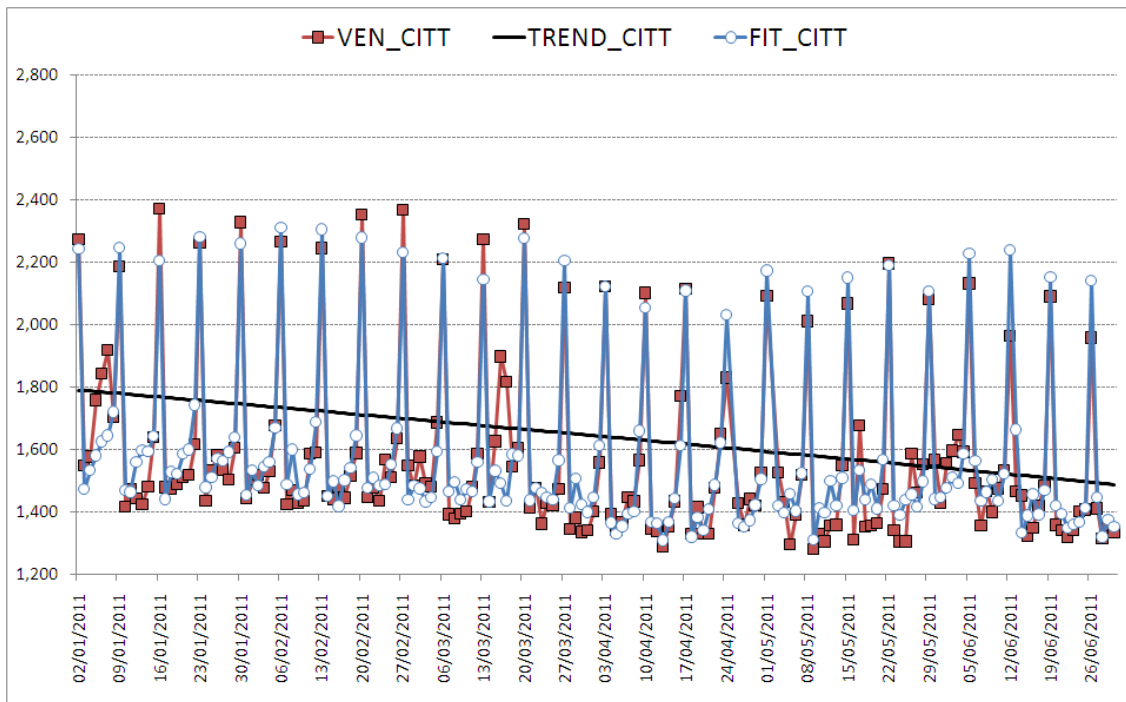


Figura 63. Vendite reali, trend e vendite previste per l'area Cittadense

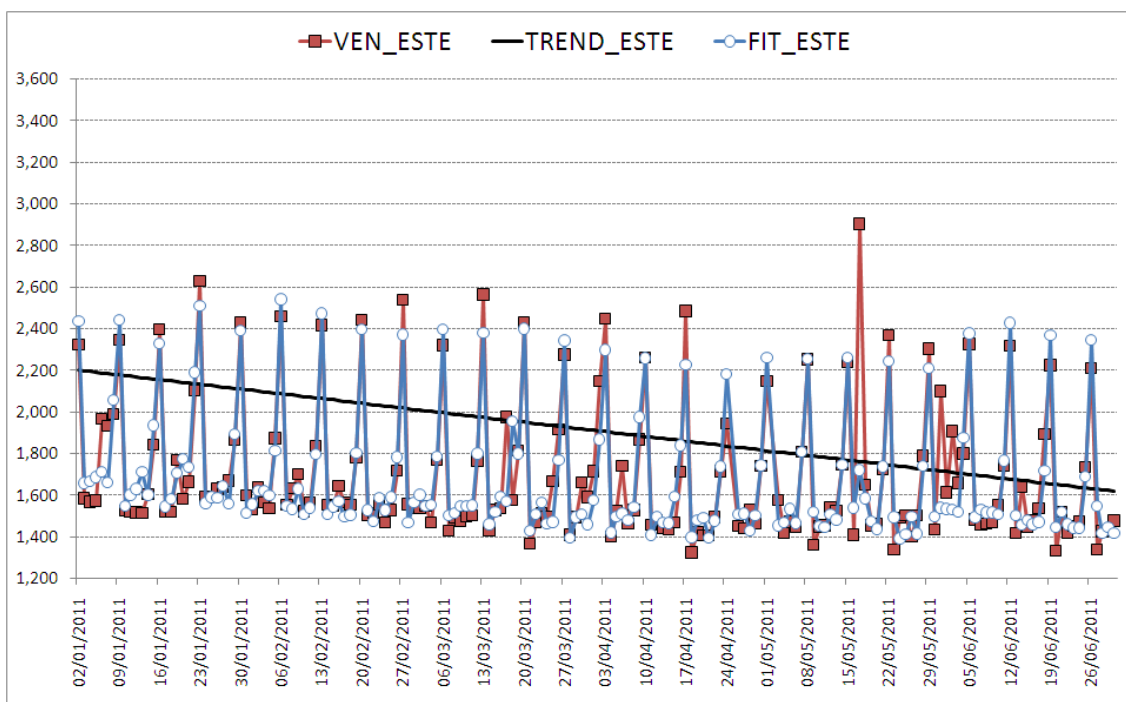


Figura 64. Vendite reali, trend e vendite previste per l'area Estense

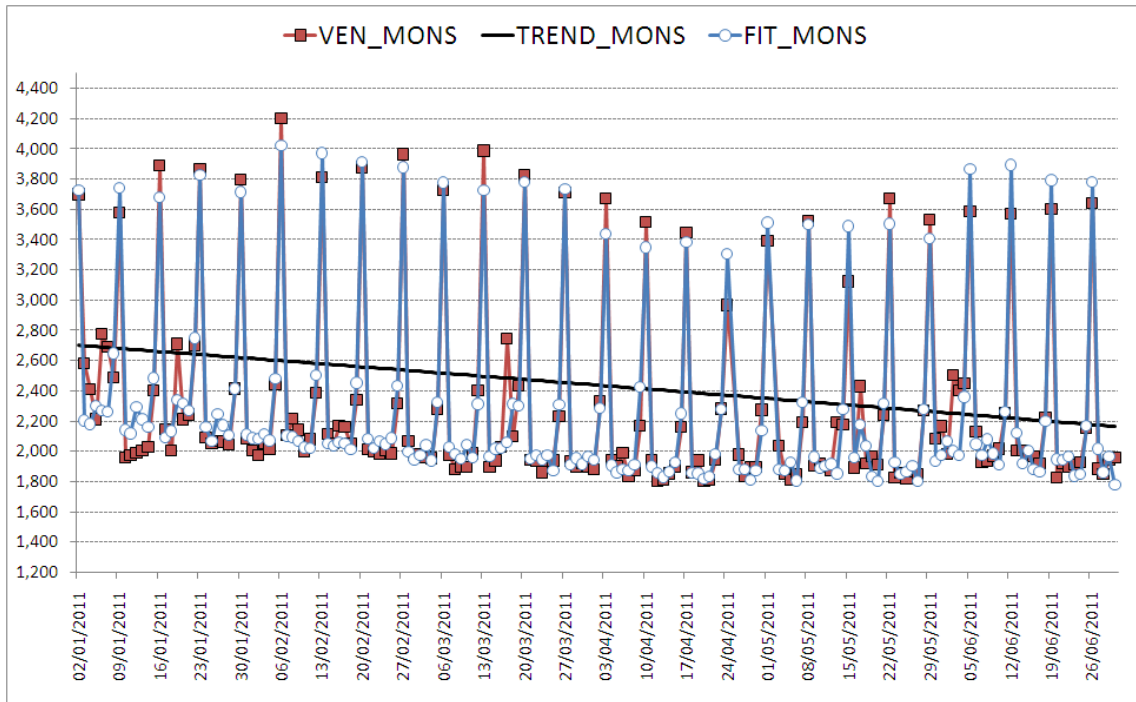


Figura 65. Vendite reali, trend e vendite previste per l'area Monselicense

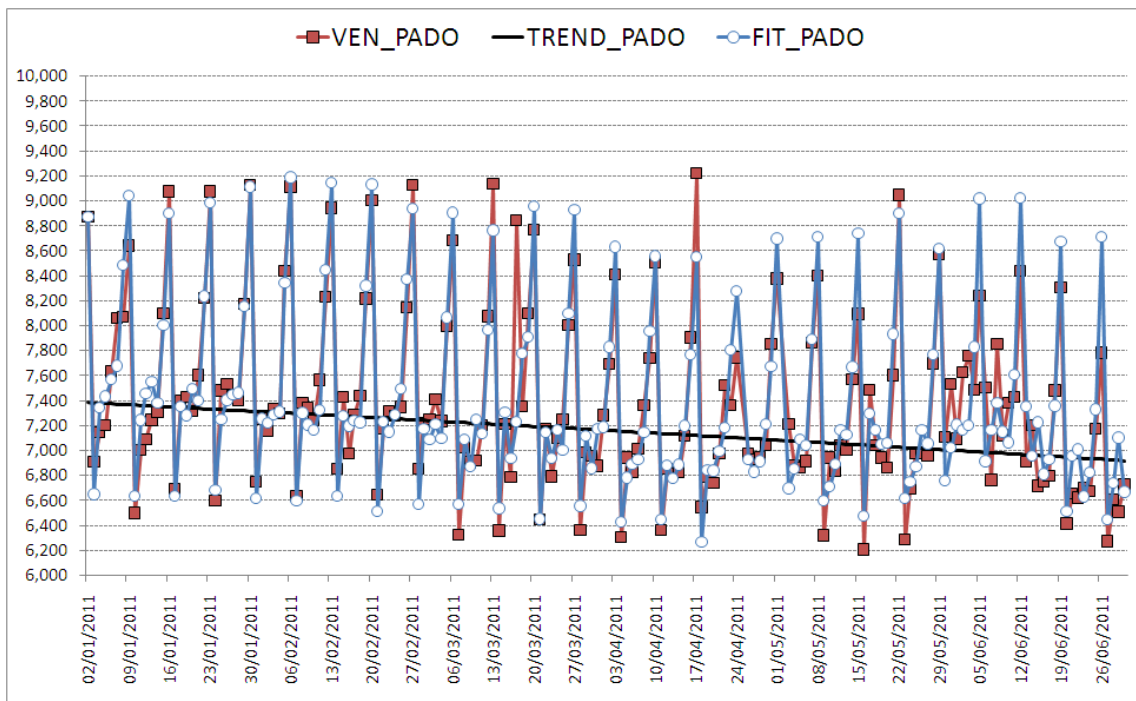


Figura 66. Vendite reali, trend e vendite previste per Padova

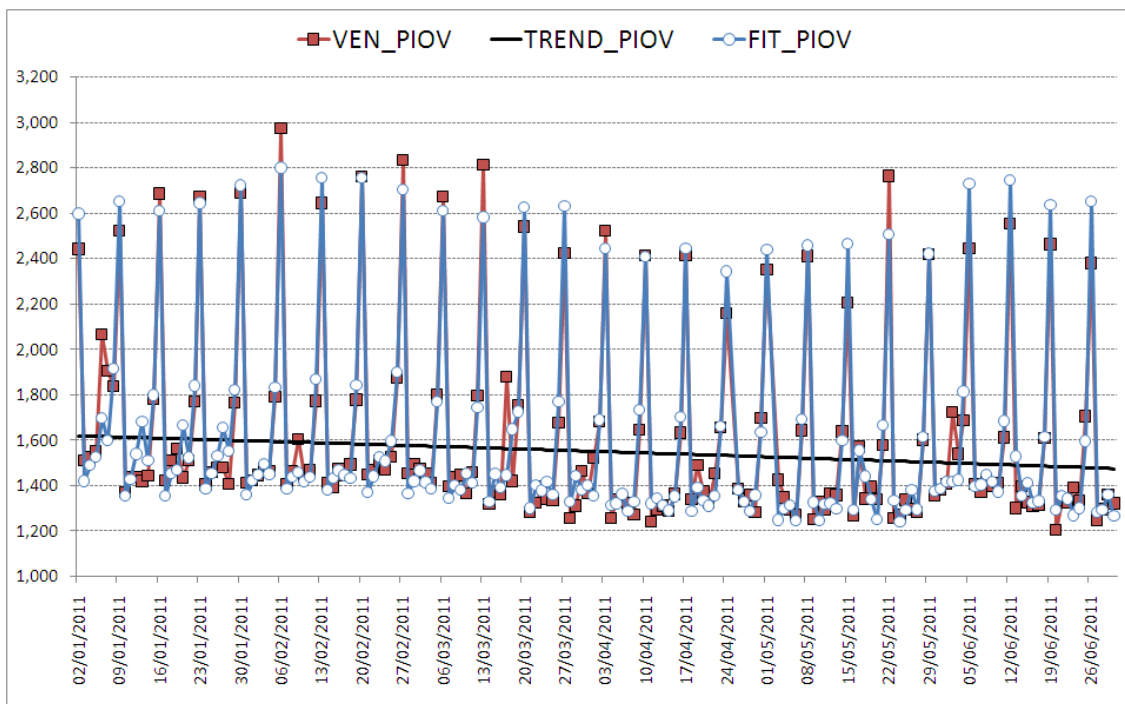


Figura 67. Vendite reali, trend e vendite previste per Piovese

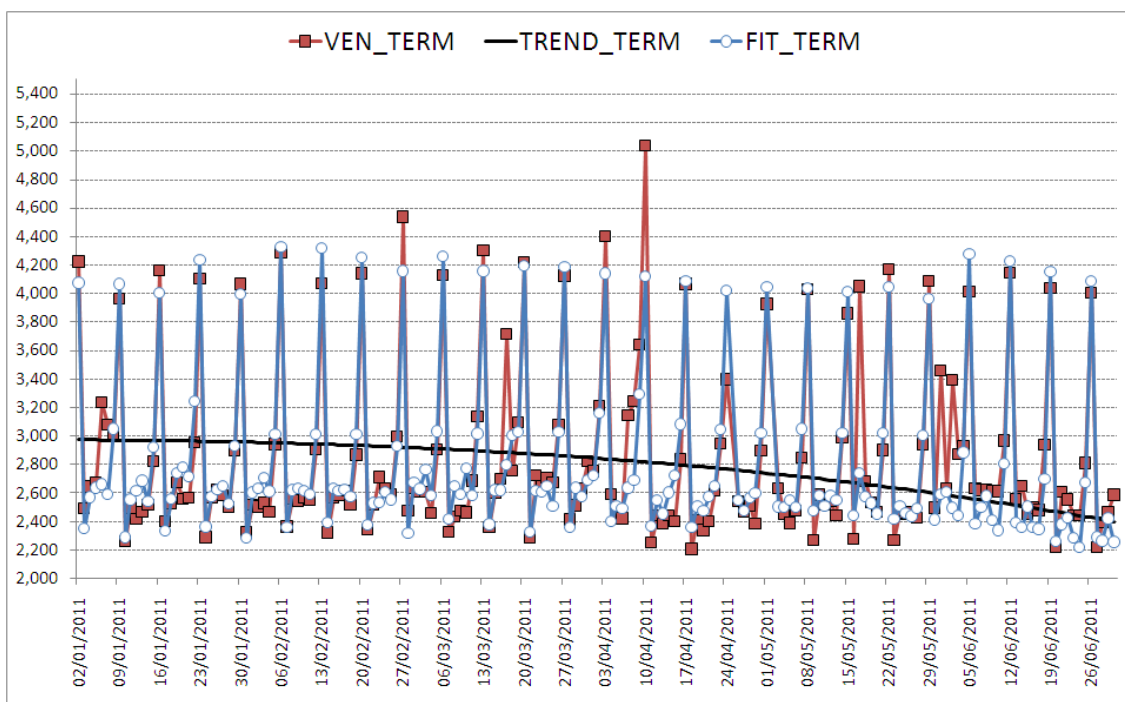


Figura 68. Vendite reali, trend e vendite previste per Terme Euganee

Questi sono i grafici ottenuti per le edicole:

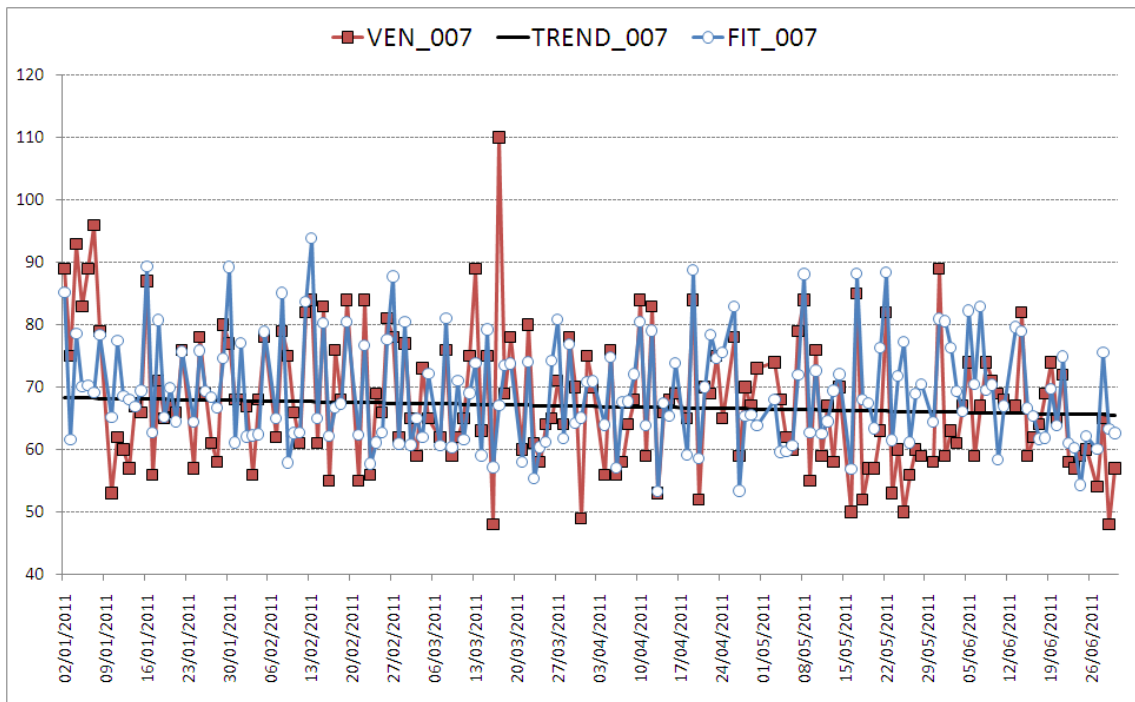


Figura 69. Vendite reali, trend e vendite previste per l'edicola 007

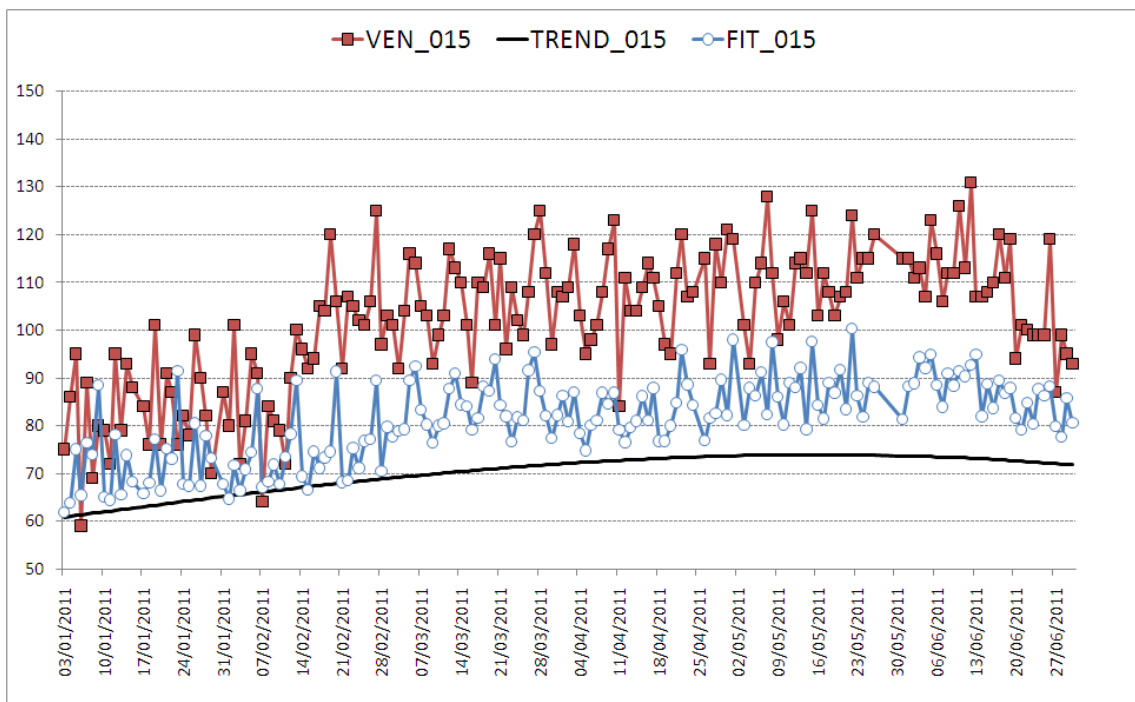


Figura 70. Vendite reali, trend e vendite previste per l'edicola 015

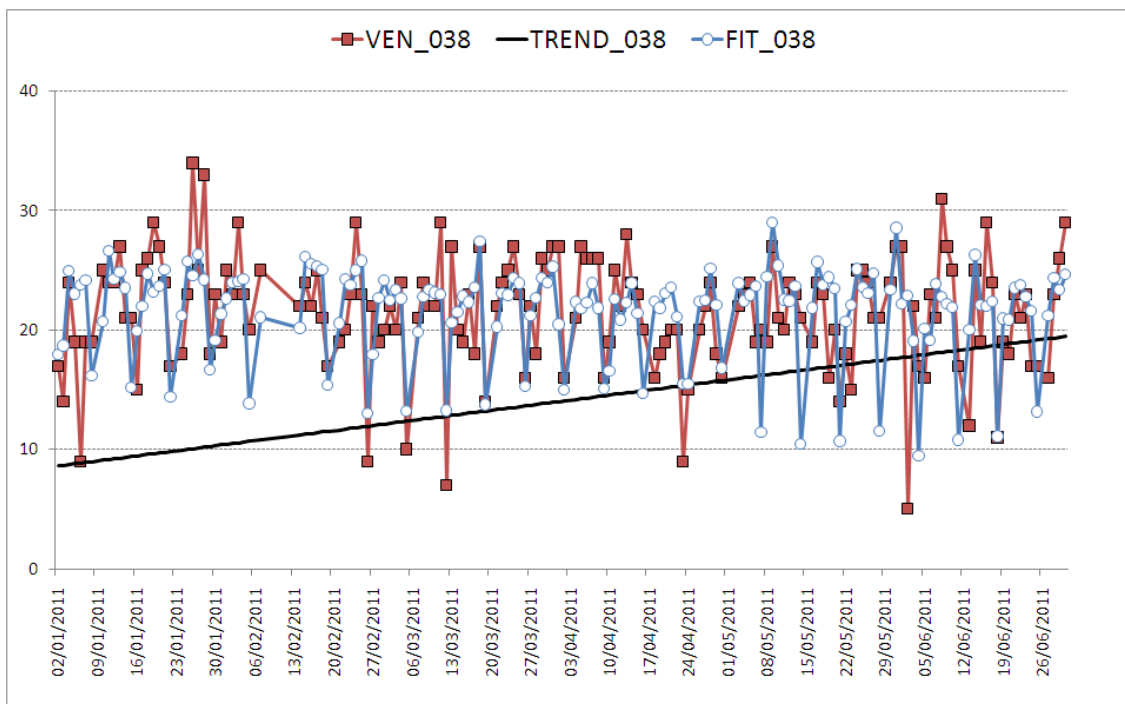


Figura 71. Vendite reali, trend e vendite previste per l'edicola 038

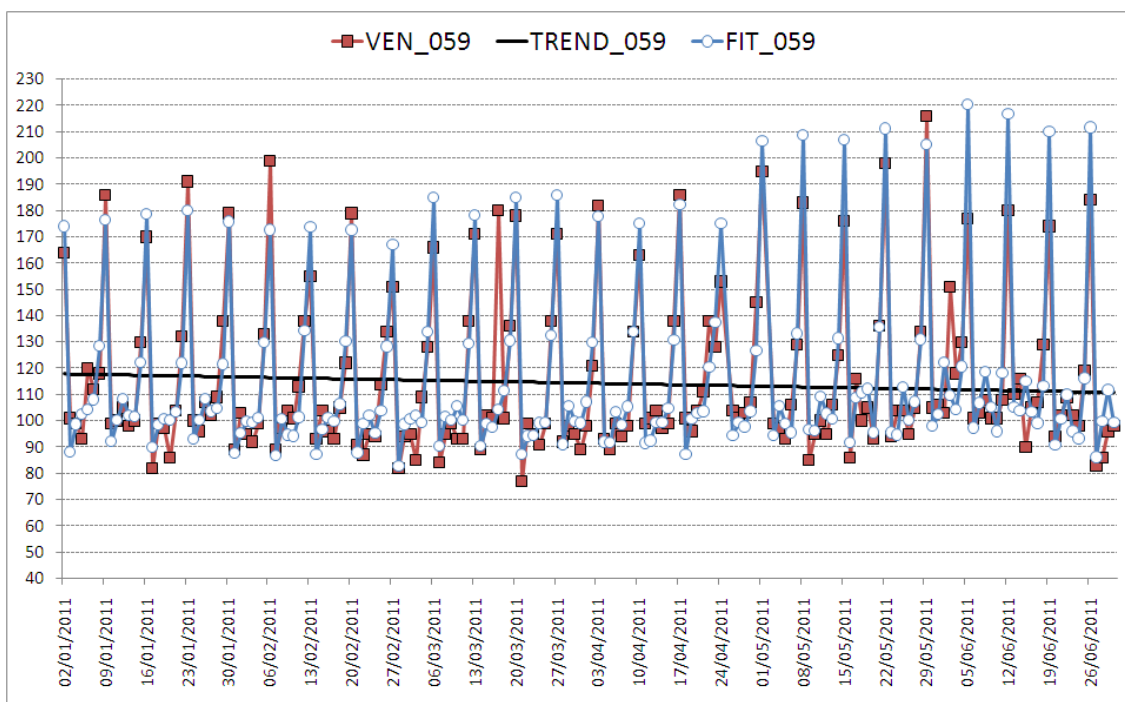


Figura 72. Vendite reali, trend e vendite previste per l'edicola 059

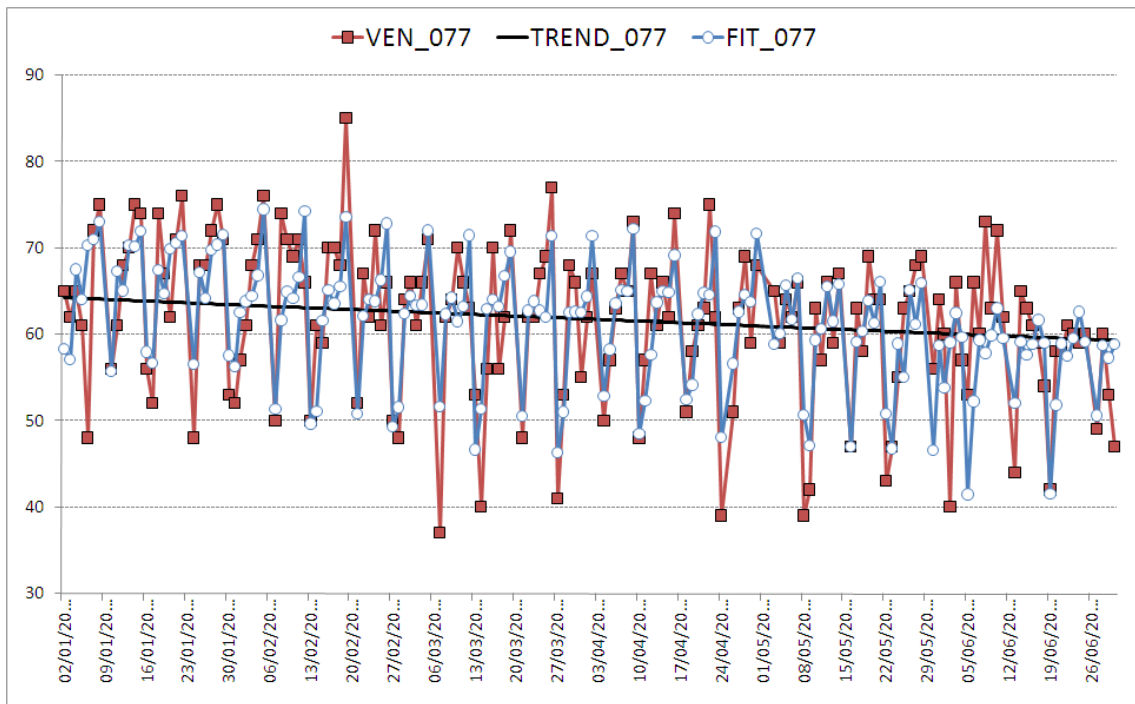


Figura 73. Vendite reali, trend e vendite previste per l'edicola 077

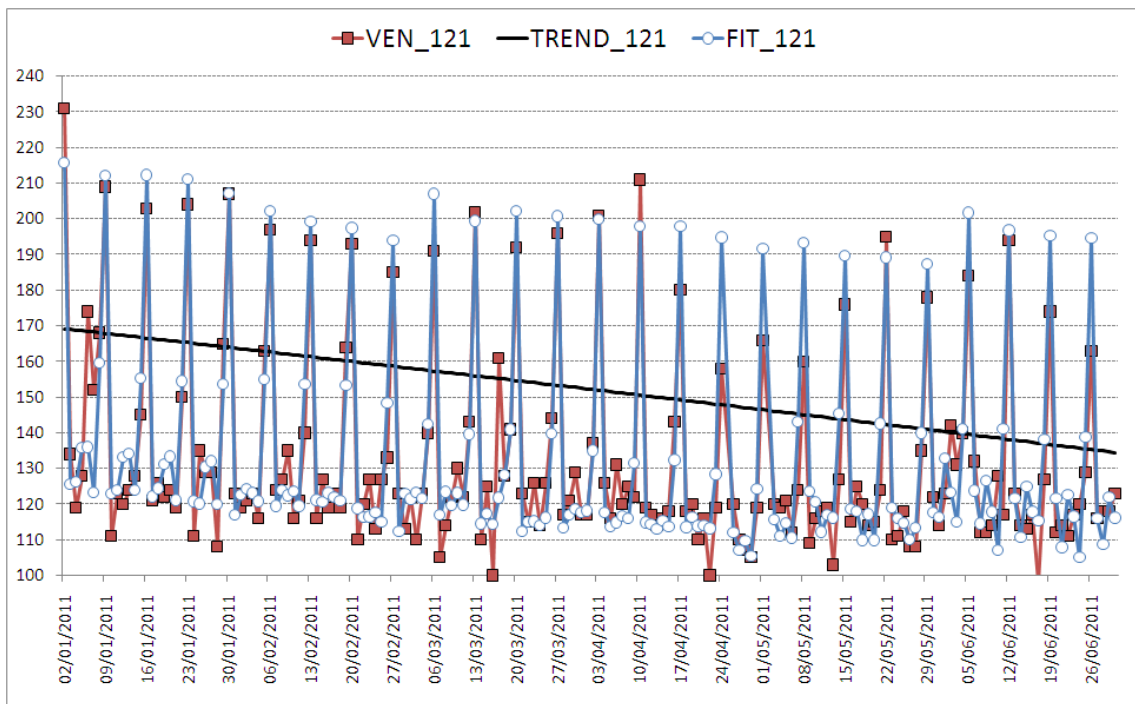


Figura 74. Vendite reali, trend e vendite previste per l'edicola 121

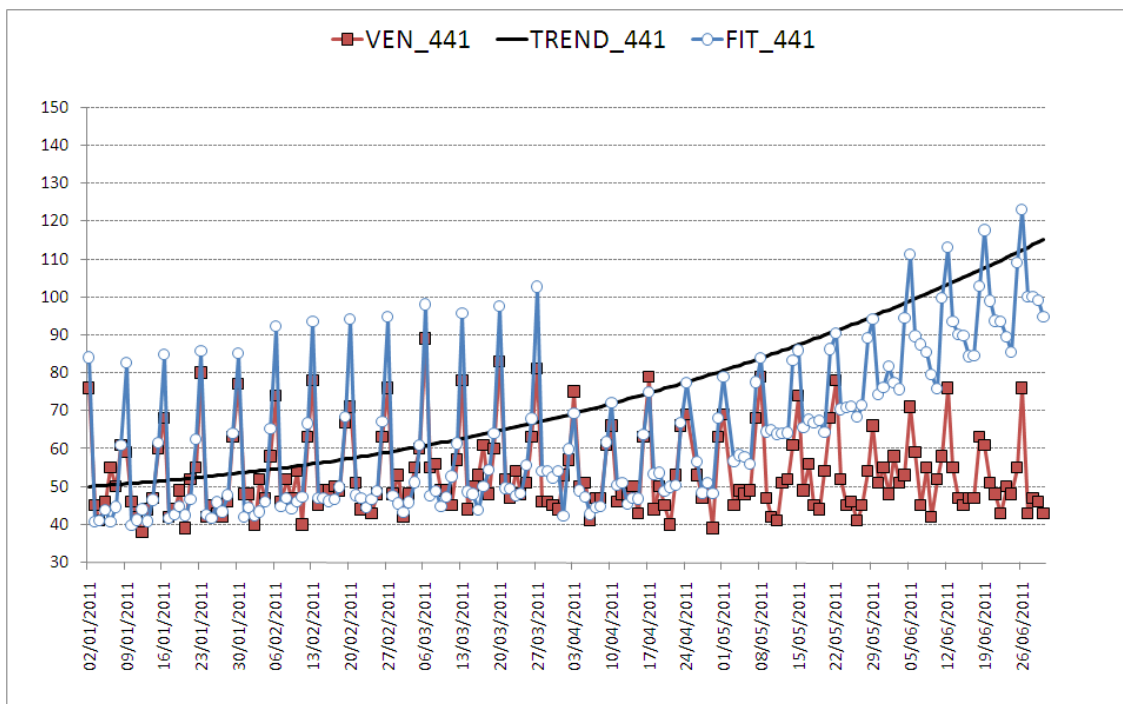


Figura 75. Vendite reali, trend e vendite previste per l'edicola 441

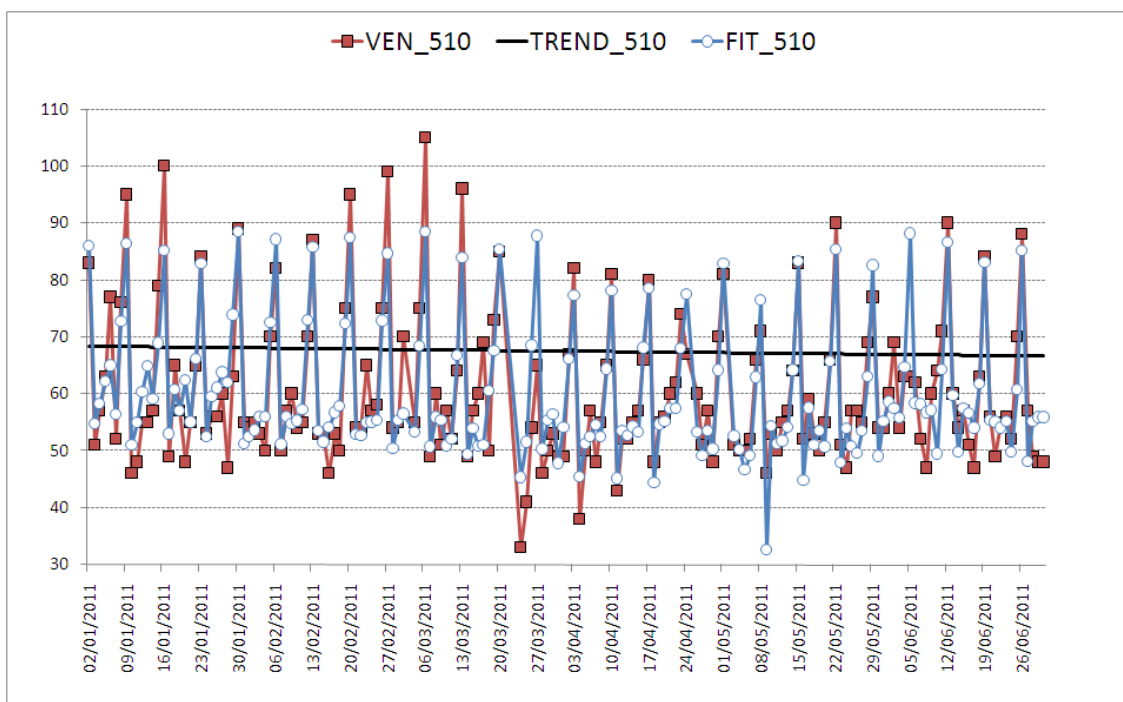


Figura 76. Vendite reali, trend e vendite previste per l'edicola 510

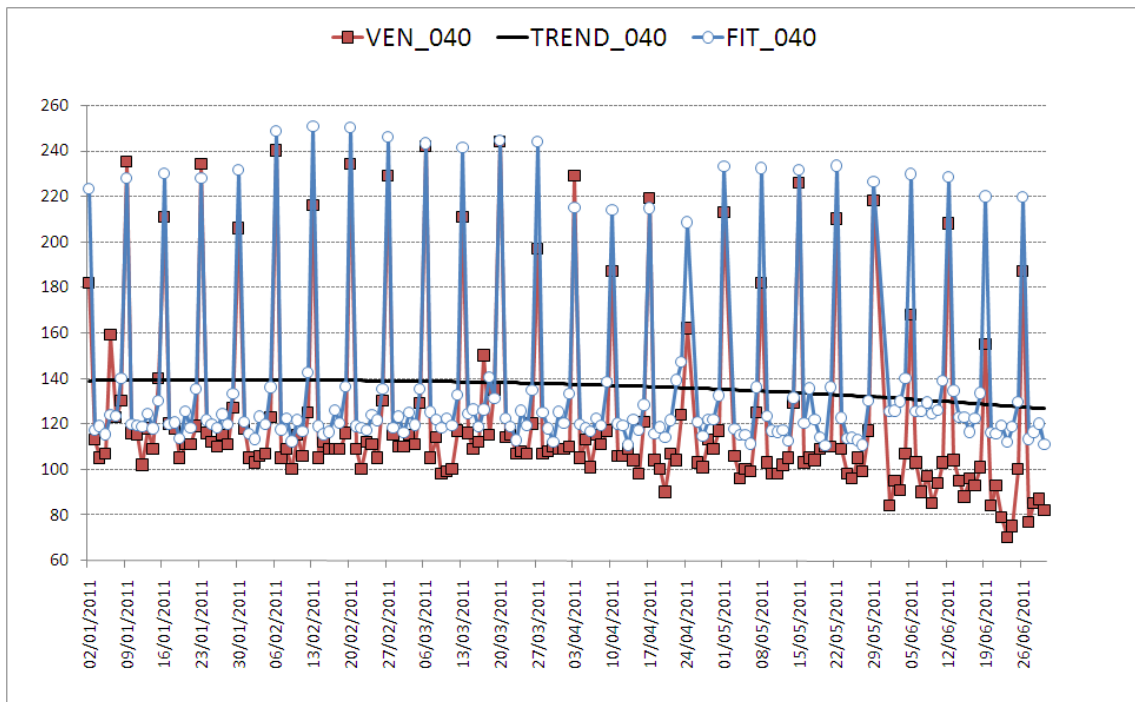


Figura 77. Vendite reali, trend e vendite previste per l'edicola 040

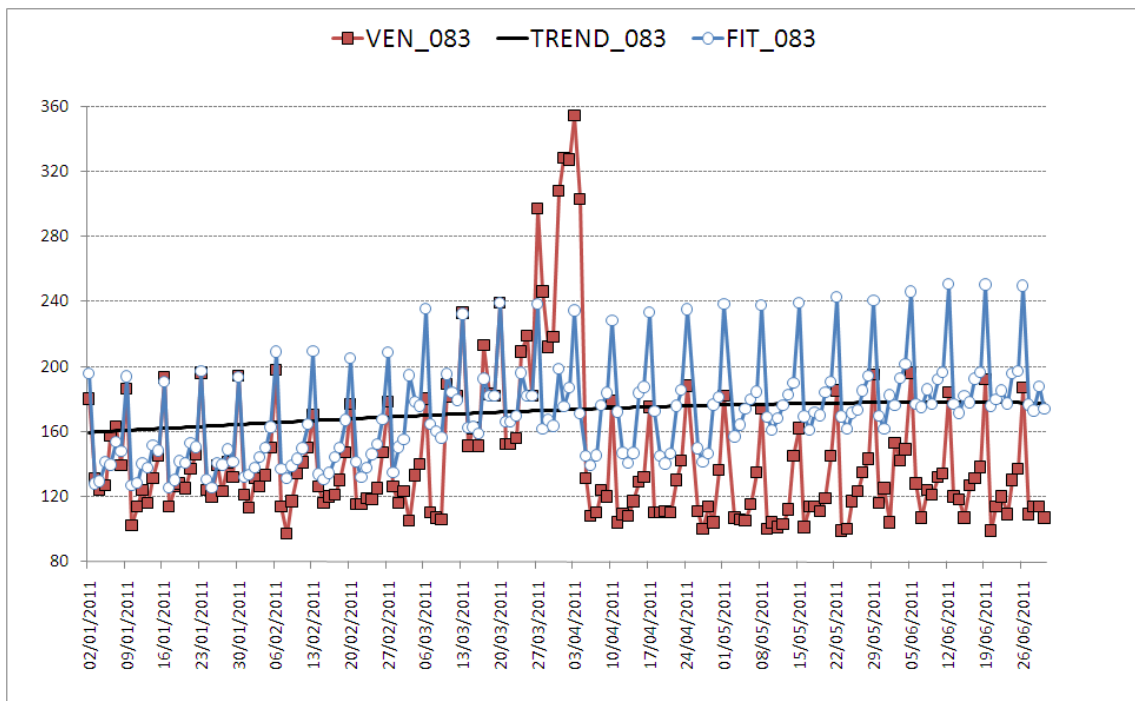


Figura 78. Vendite reali, trend e vendite previste per l'edicola 083

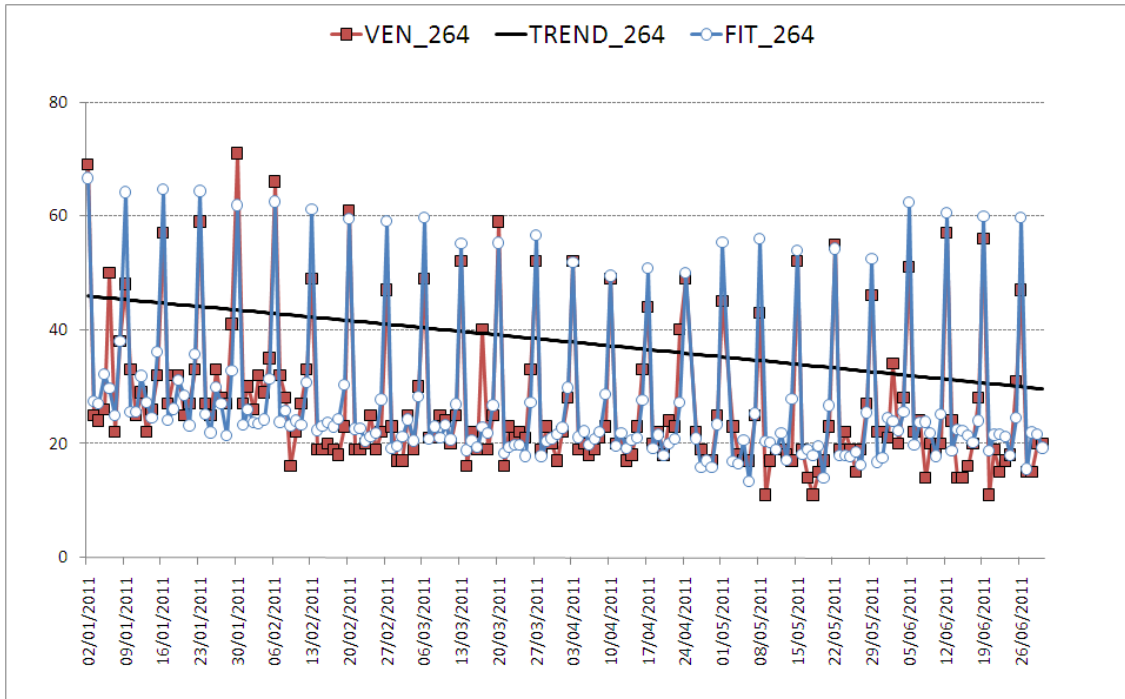


Figura 79. Vendite reali, trend e vendite previste per l'edicola 264

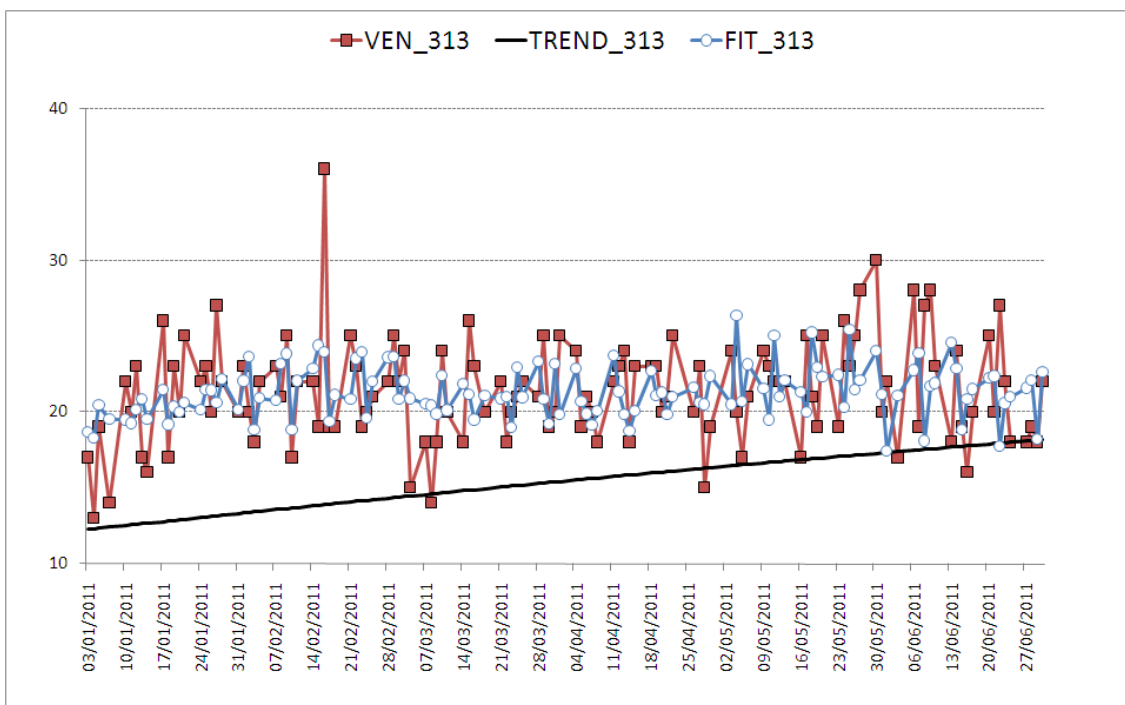


Figura 80. Vendite reali, trend e vendite previste per l'edicola 313

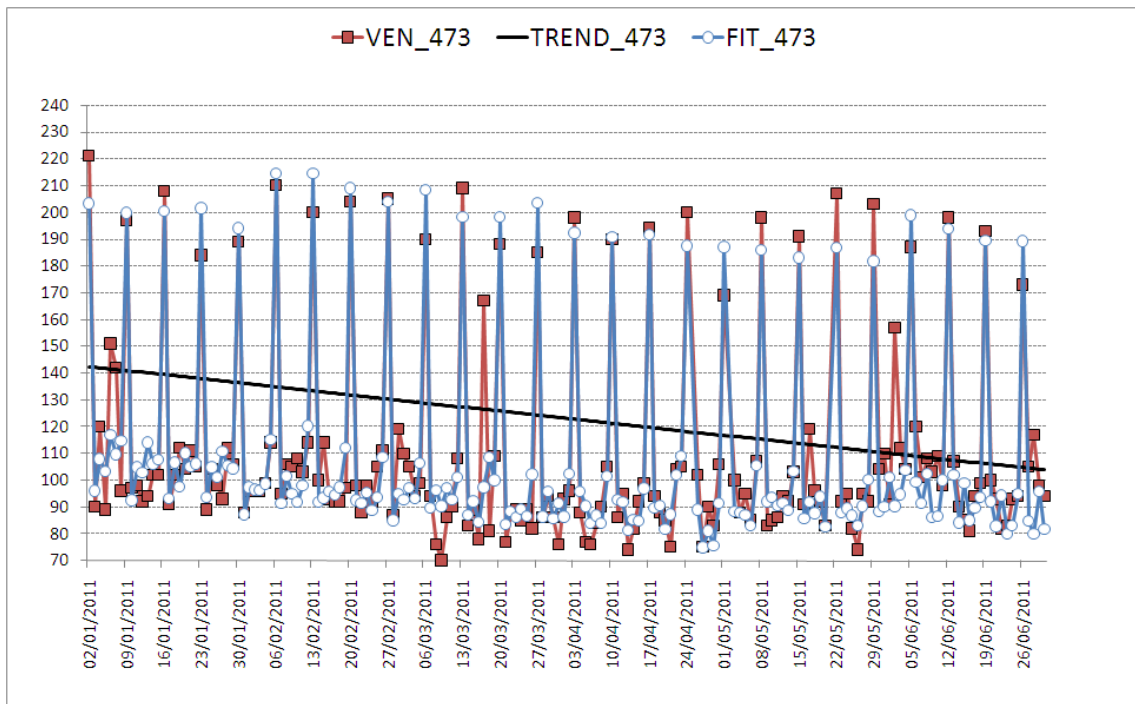


Figura 81. Vendite reali, trend e vendite previste per l'edicola 473

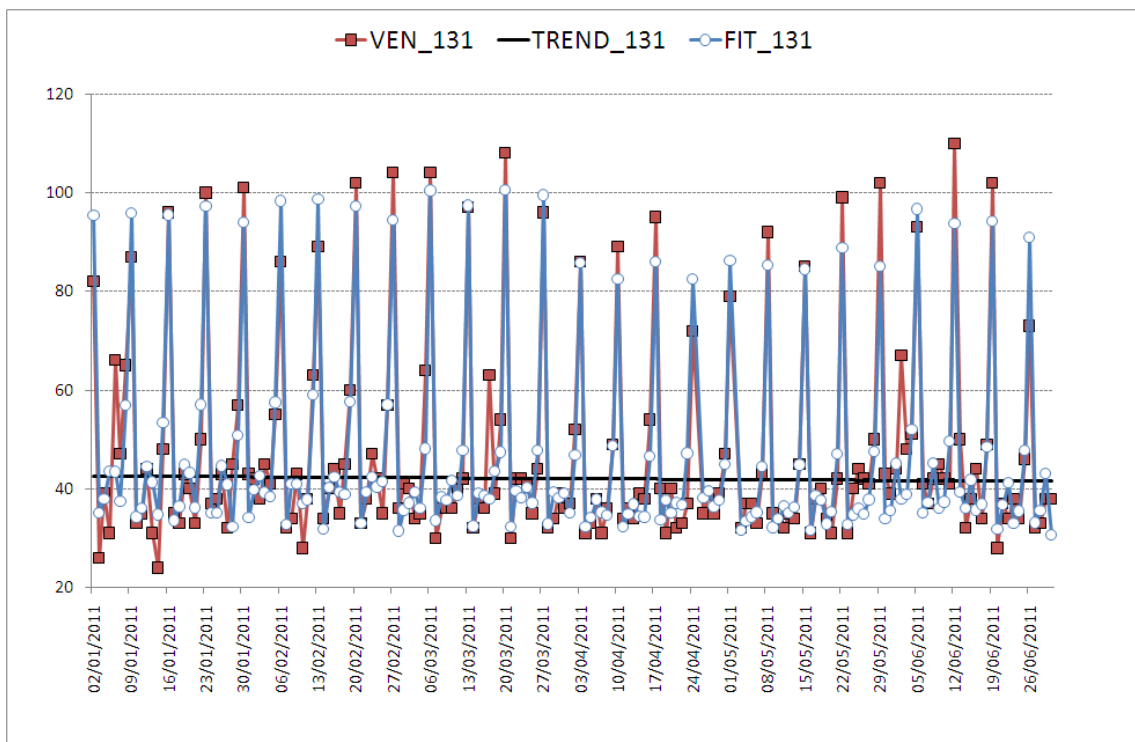


Figura 82. Vendite reali, trend e vendite previste per l'edicola 131

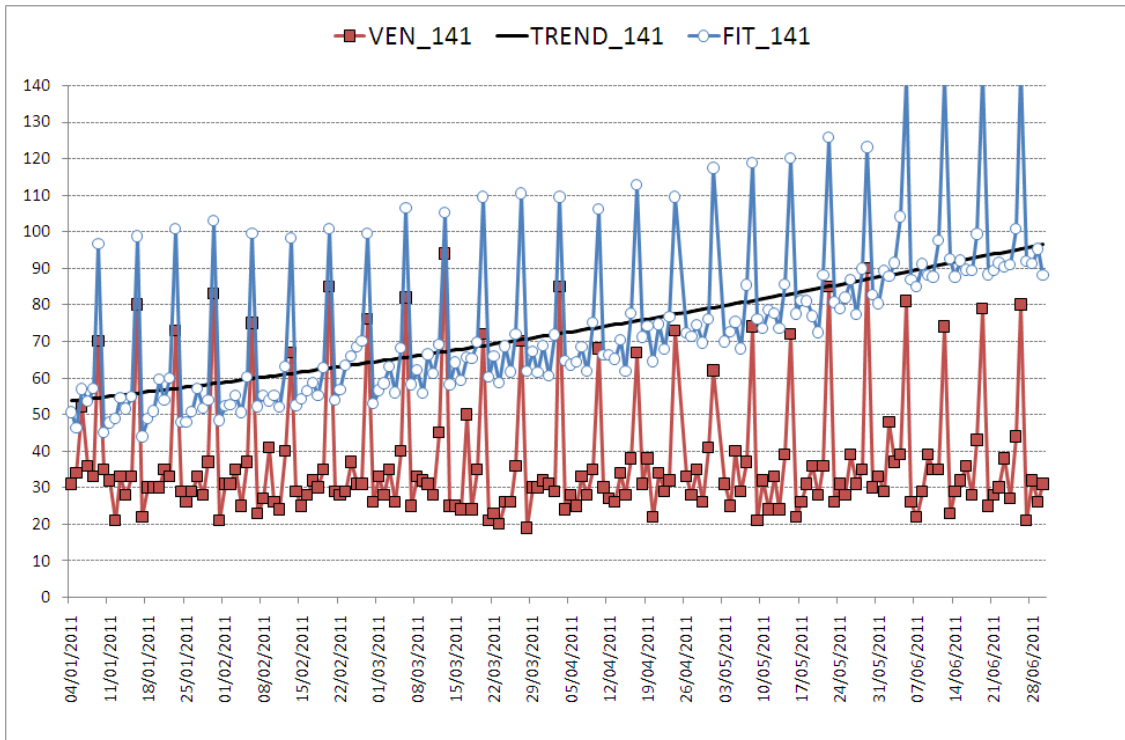


Figura 83. Vendite reali, trend e vendite previste per l'edicola 141

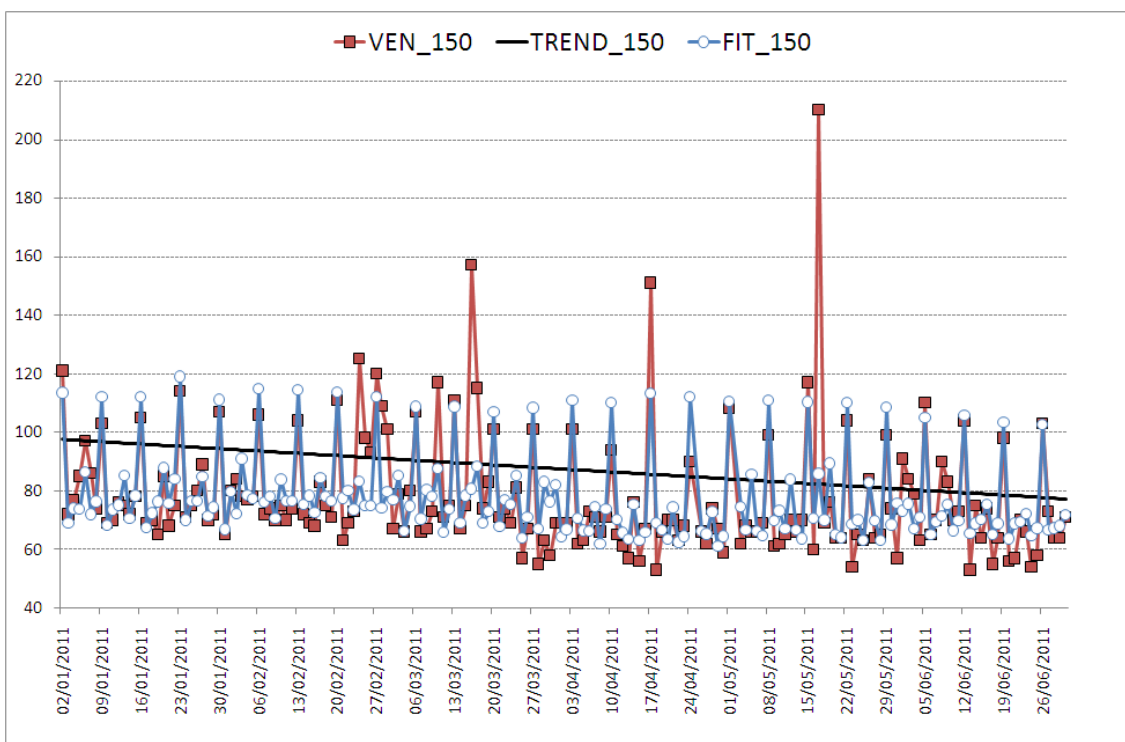


Figura 84. Vendite reali, trend e vendite previste per l'edicola 150

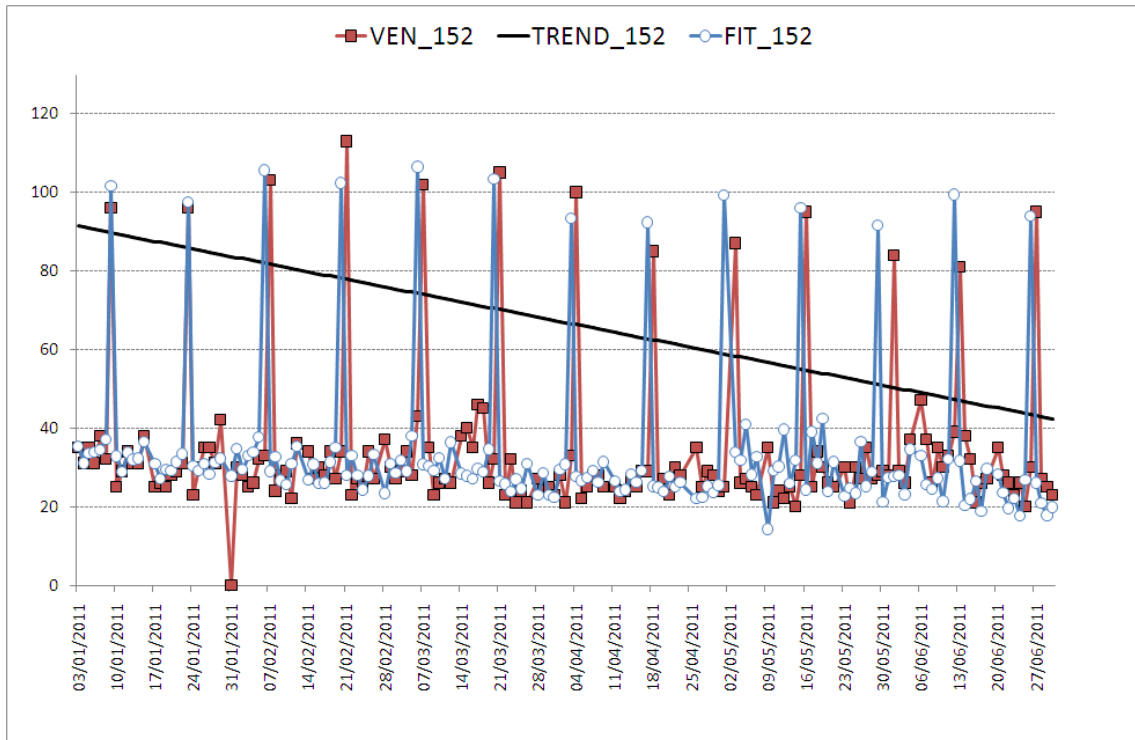


Figura 85. Vendite reali, trend e vendite previste per l'edicola 152

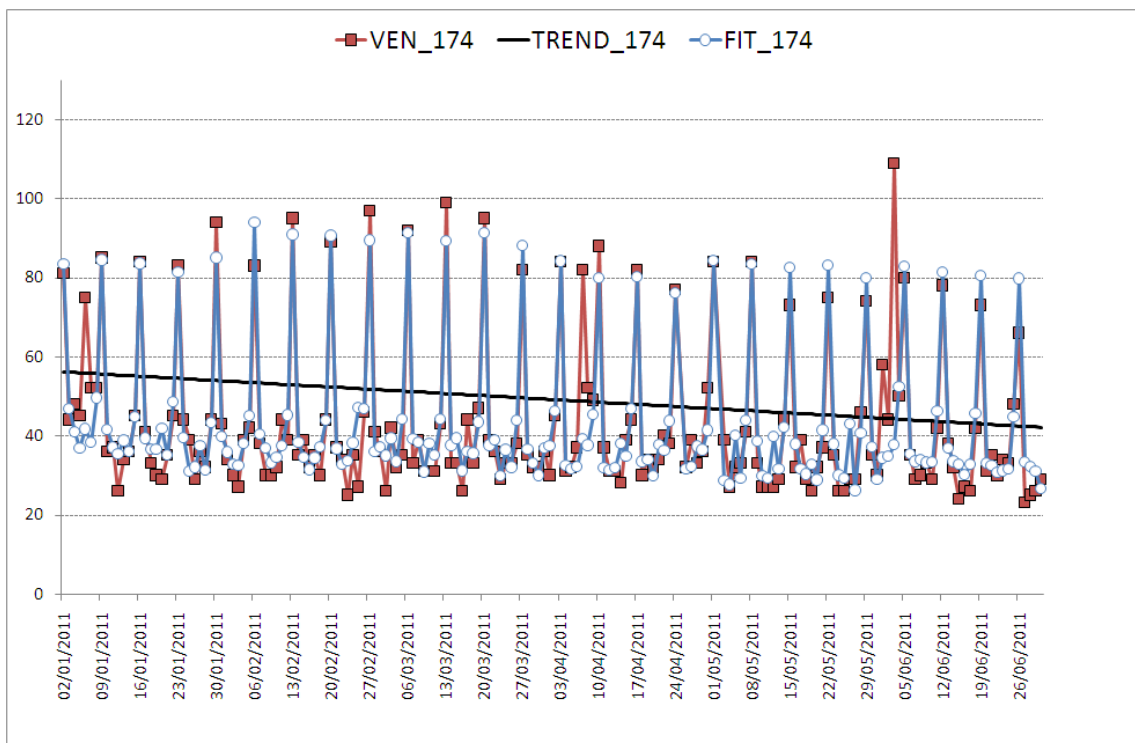


Figura 86. Vendite reali, trend e vendite previste per l'edicola 174

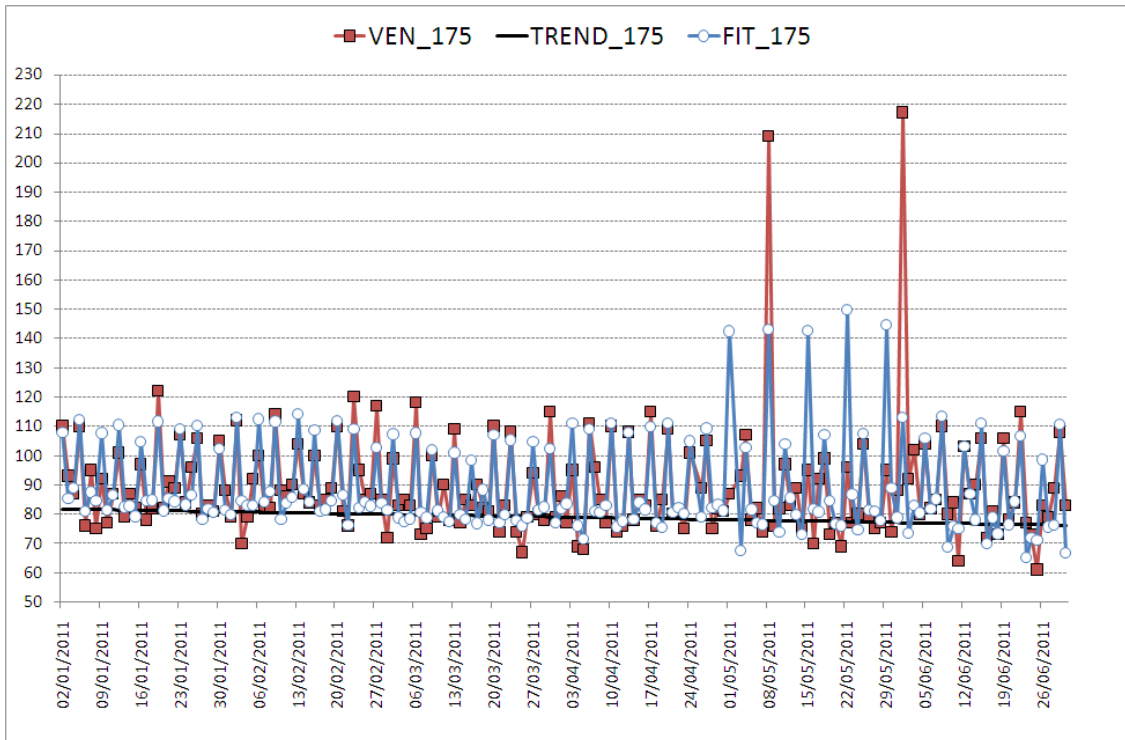


Figura 87. Vendite reali, trend e vendite previste per l'edicola 175

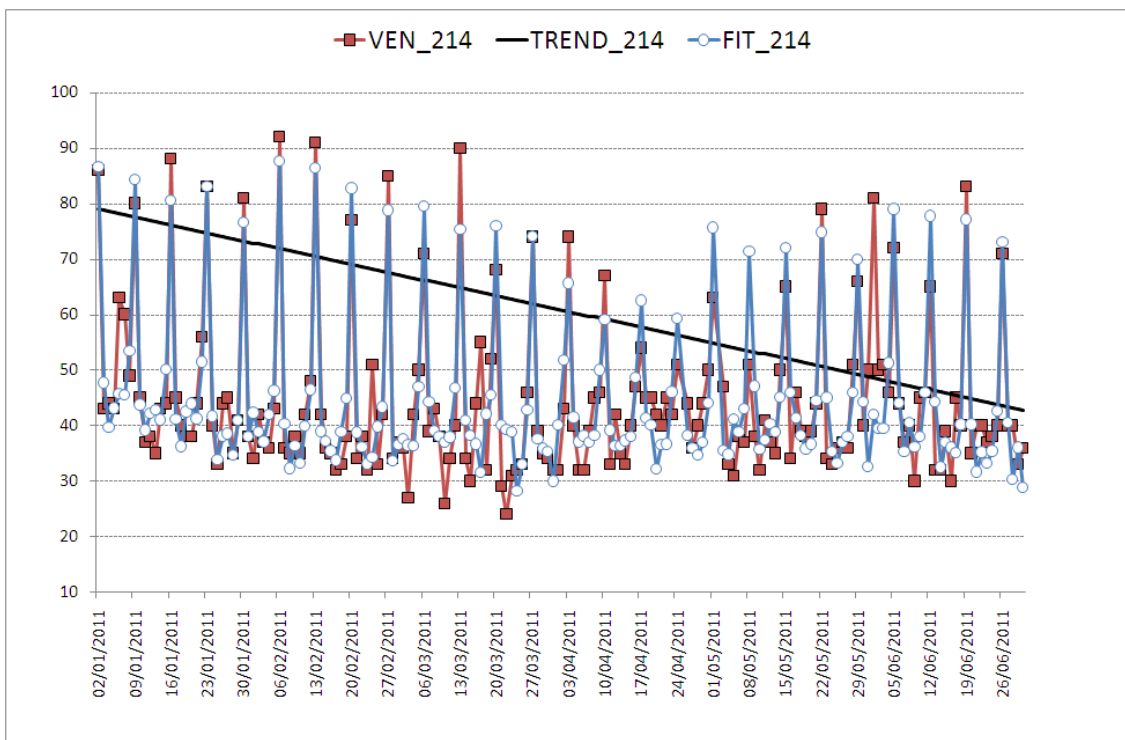


Figura 88. Vendite reali, trend e vendite previste per l'edicola 214

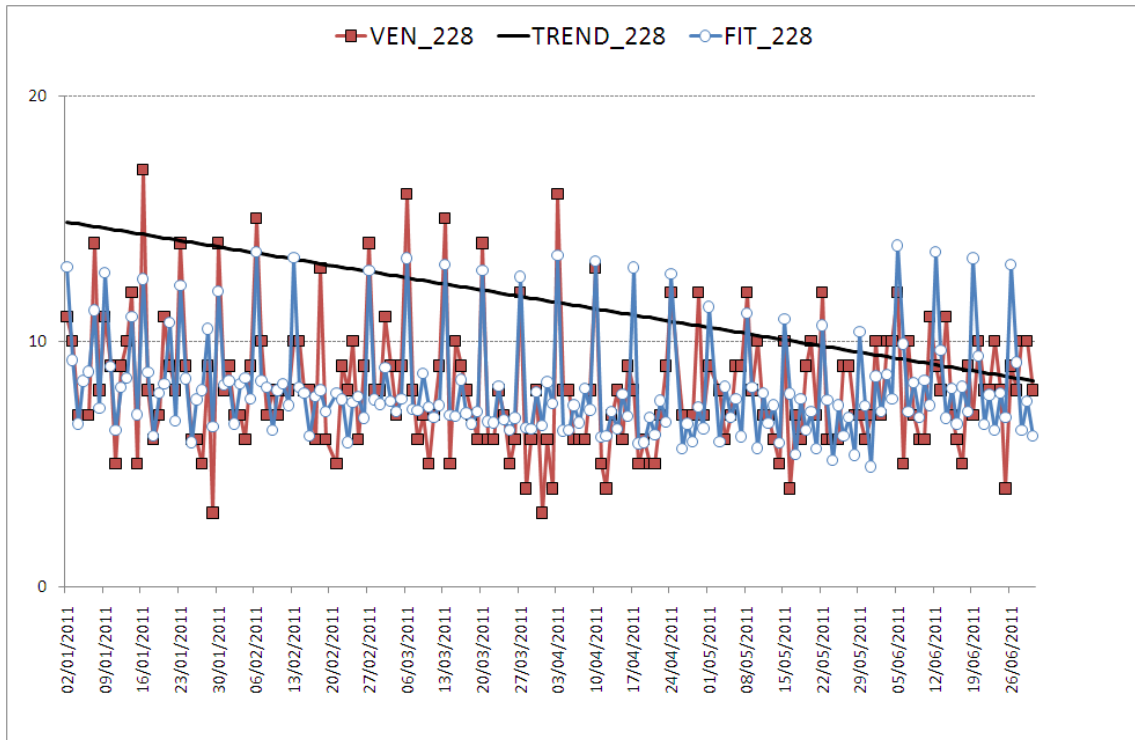


Figura 89. Vendite reali, trend e vendite previste per l'edicola 228

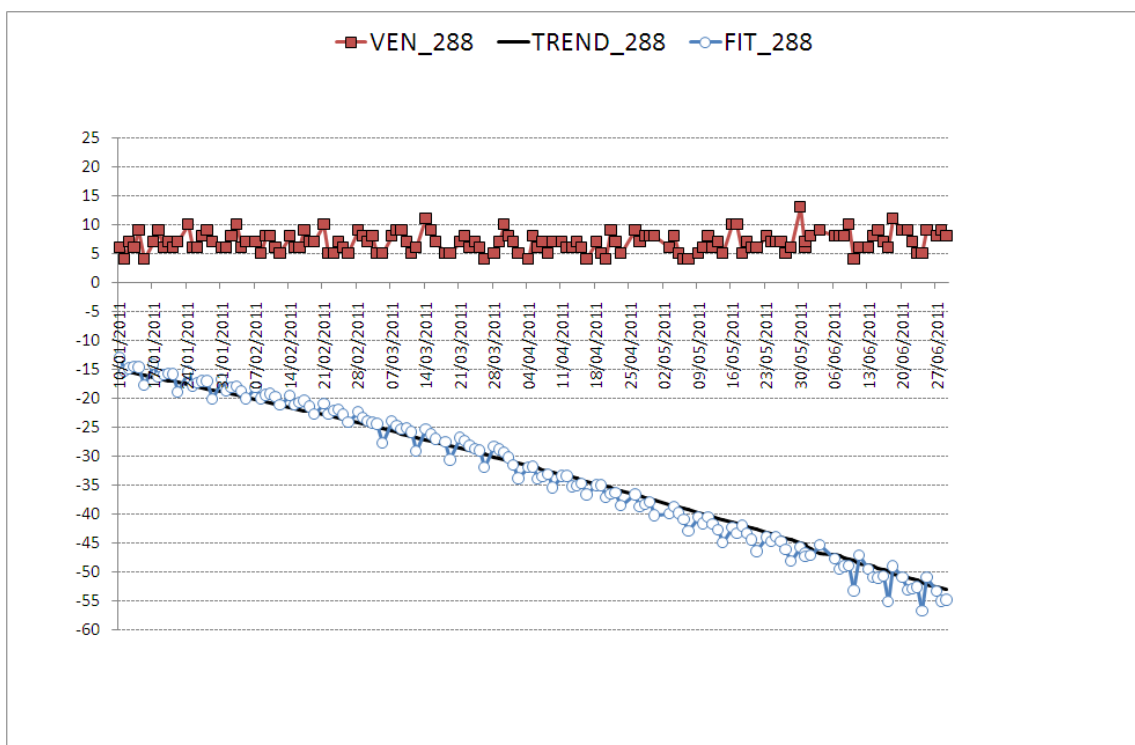


Figura 90. Vendite reali, trend e vendite previste per l'edicola 288

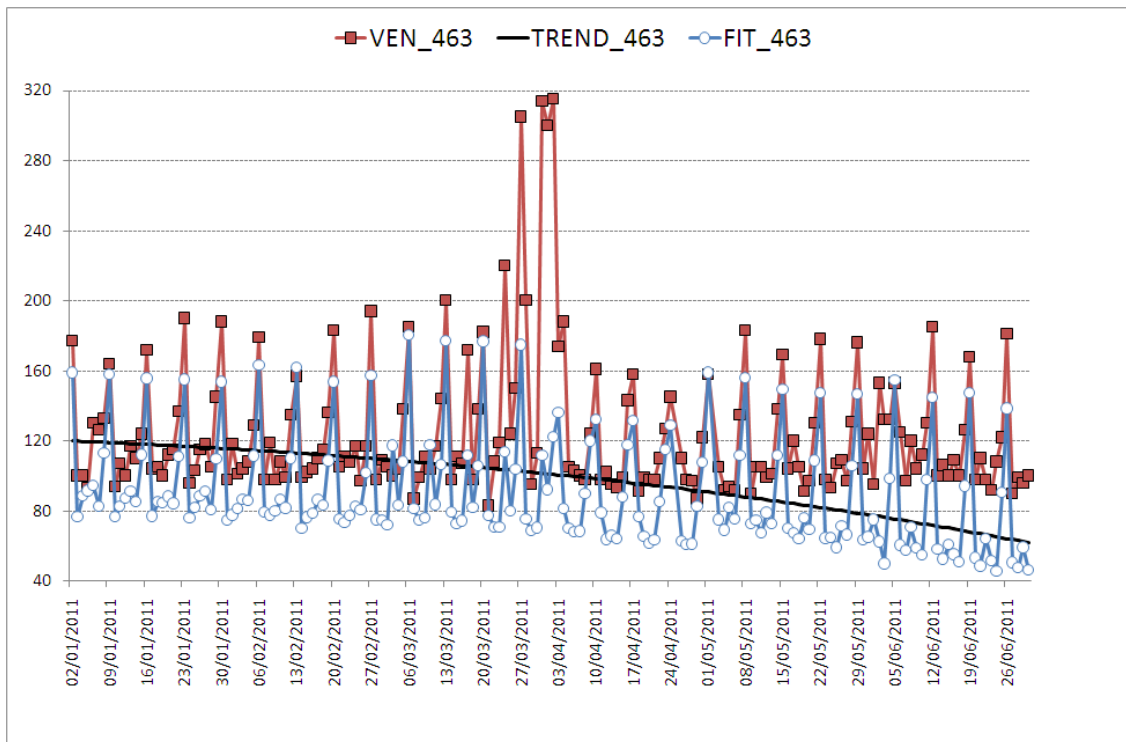


Figura 91. Vendite reali, trend e vendite previste per l'edicola 463

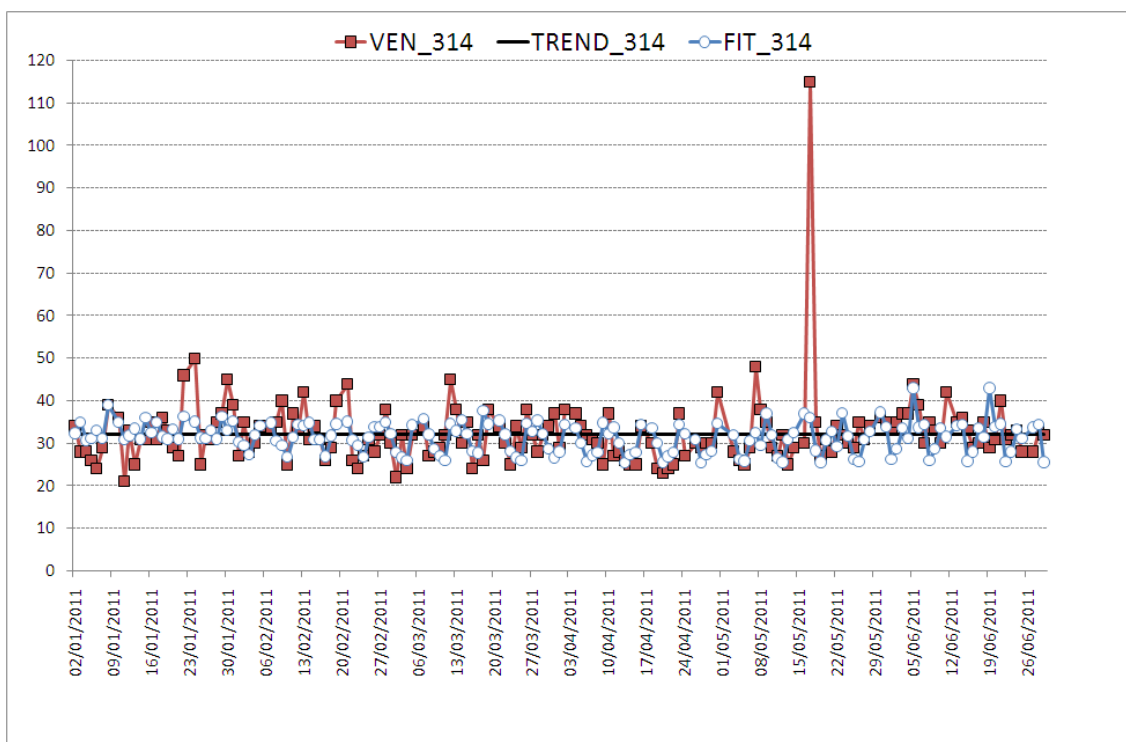


Figura 92. Vendite reali, trend e vendite previste per l'edicola 314

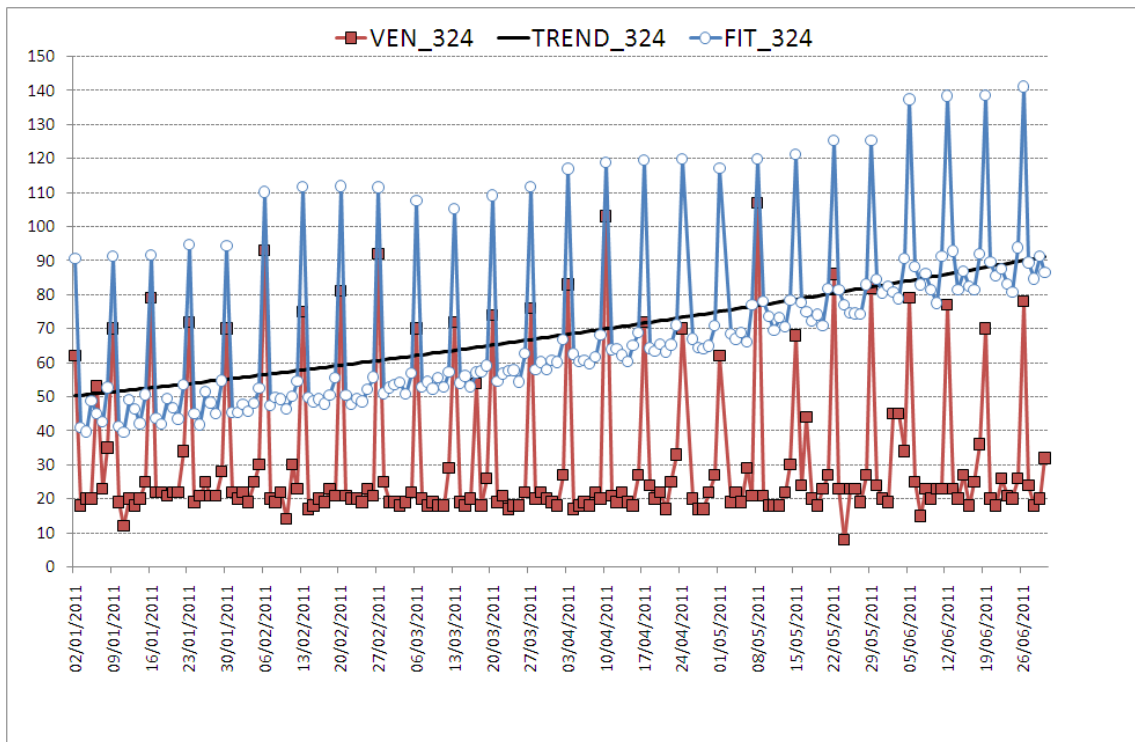


Figura 93. Vendite reali, trend e vendite previste per l'edicola 324

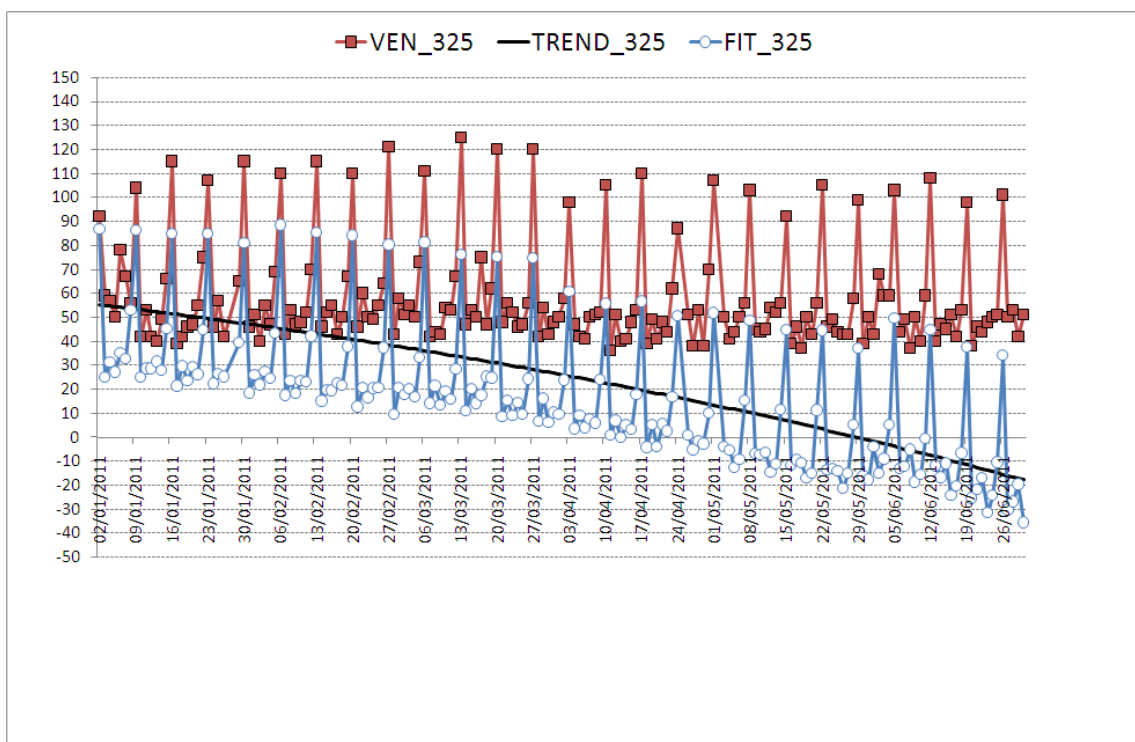


Figura 94. Vendite reali, trend e vendite previste per l'edicola 325

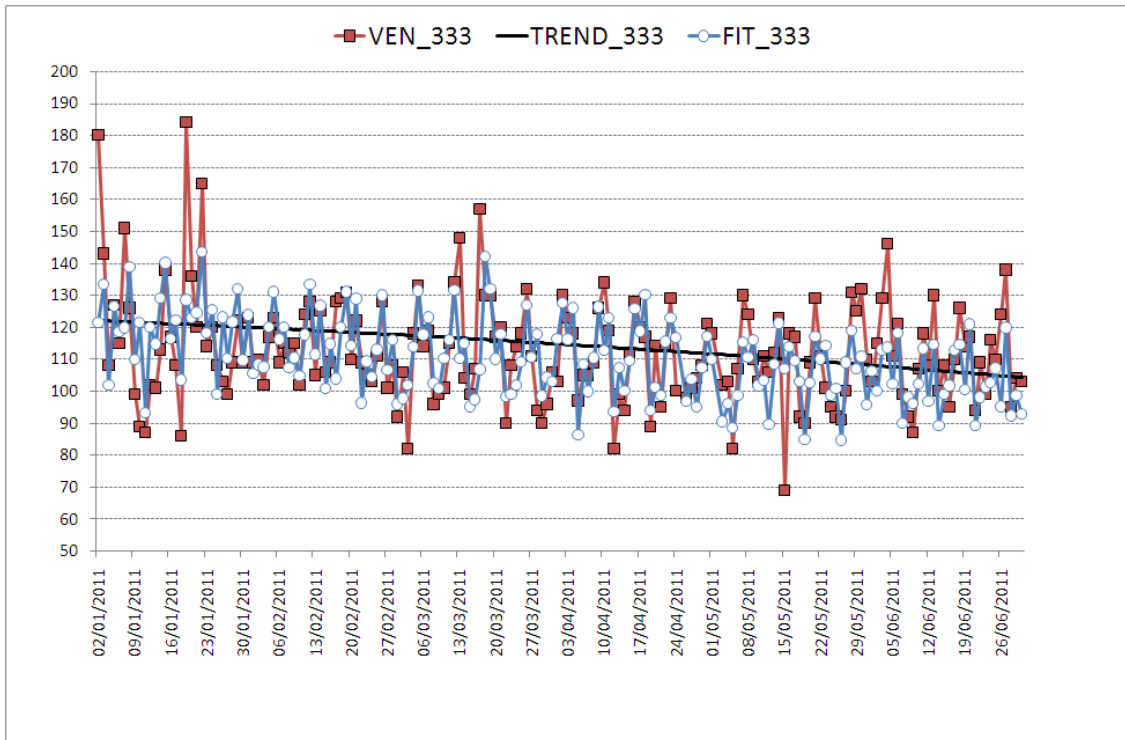


Figura 95. Vendite reali, trend e vendite previste per l'edicola 333

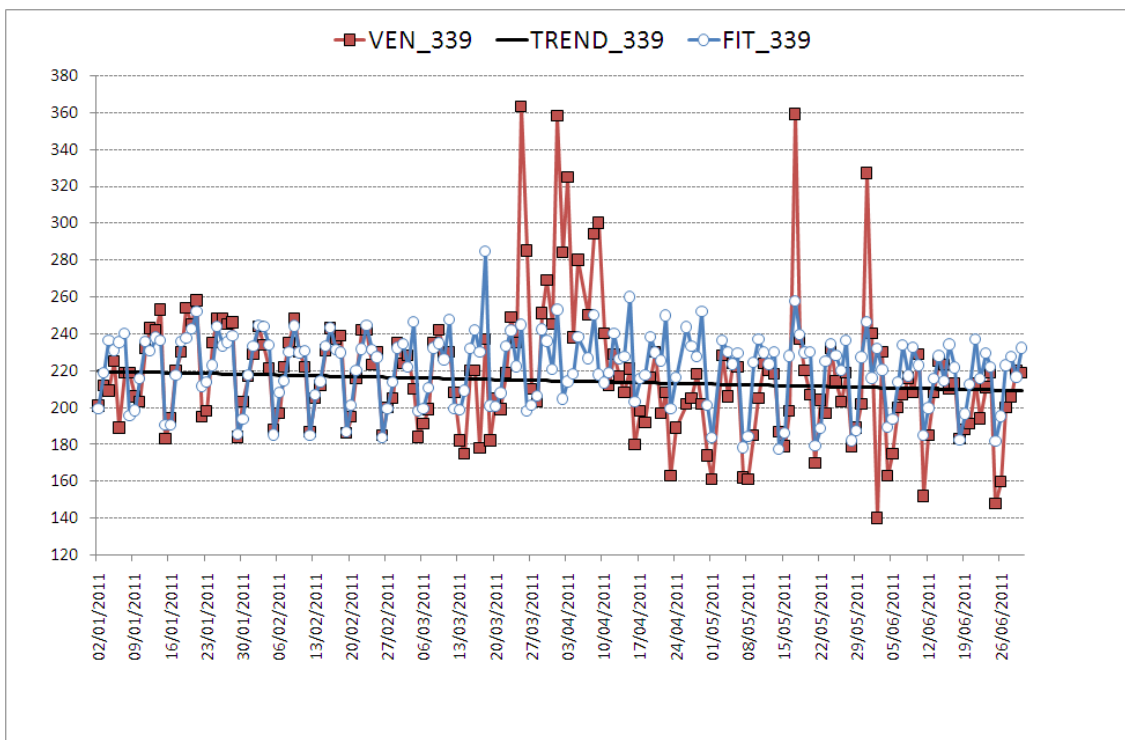


Figura 96. Vendite reali, trend e vendite previste per l'edicola 339

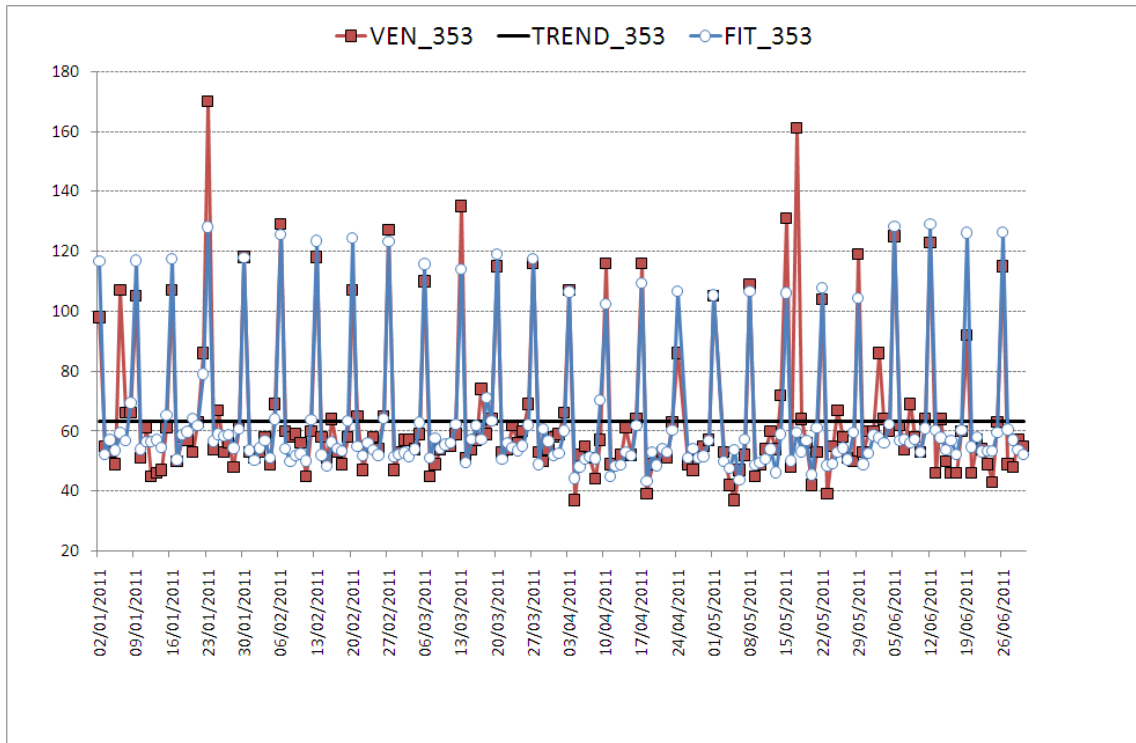


Figura 97. Vendite reali, trend e vendite previste per l'edicola 353

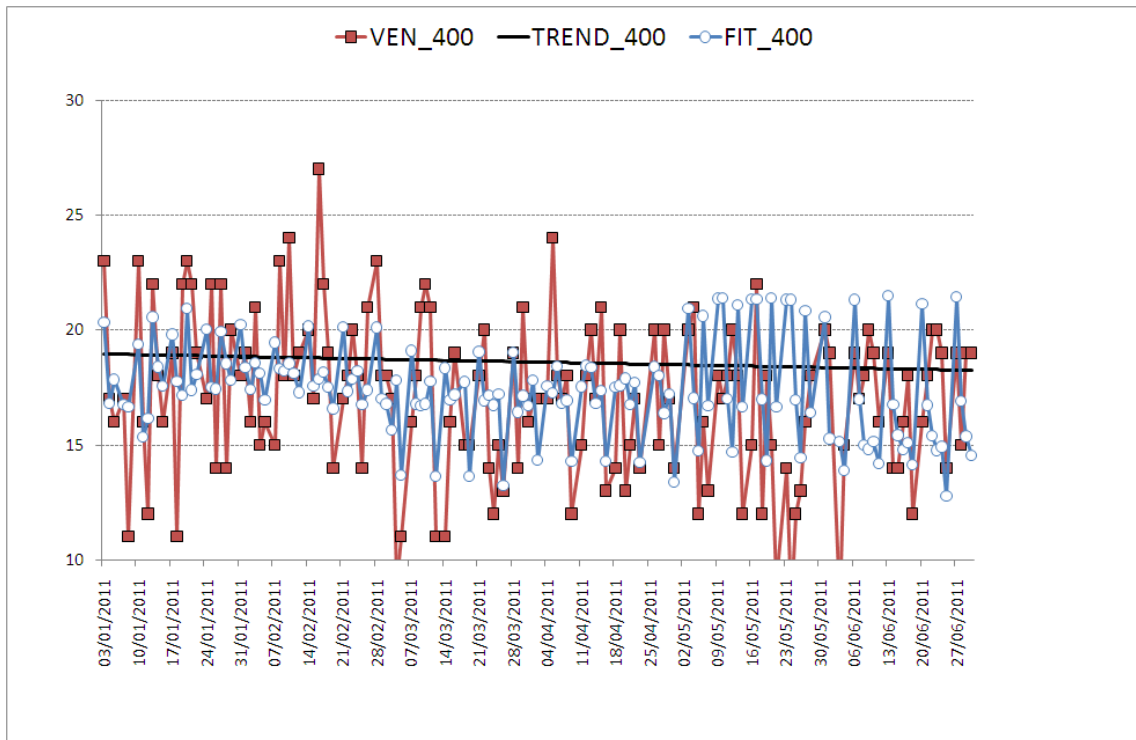


Figura 98. Vendite reali, trend e vendite previste per l'edicola 400

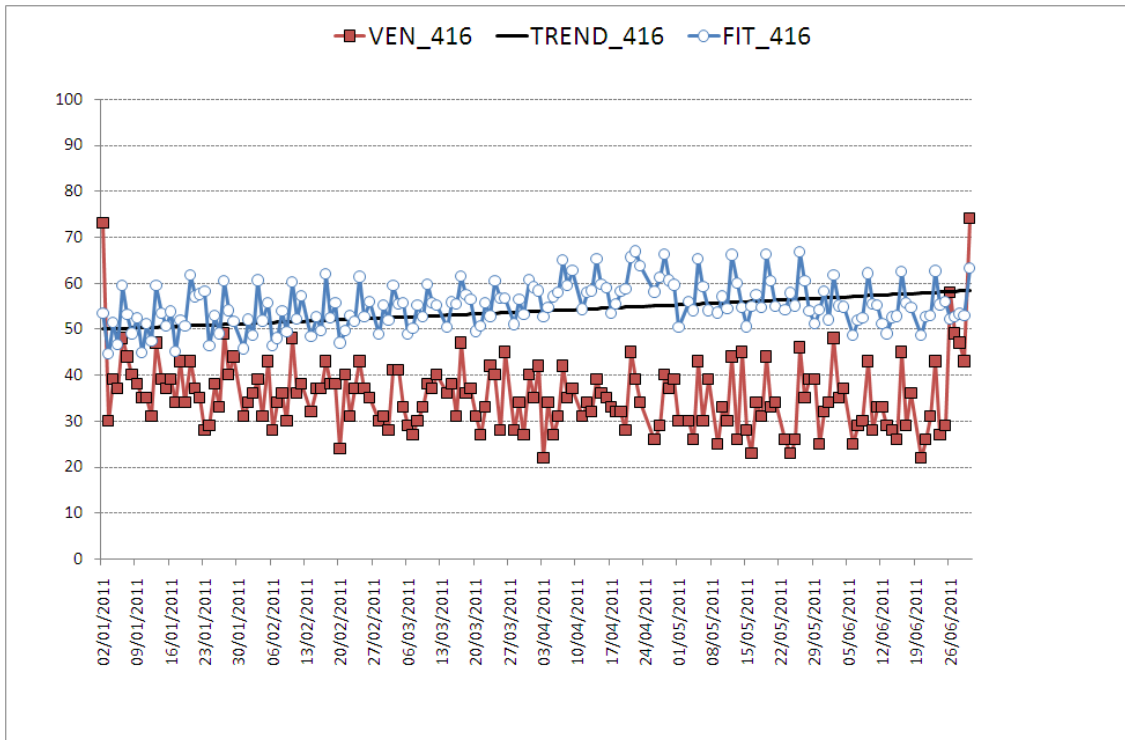


Figura 99. Vendite reali, trend e vendite previste per l'edicola 416

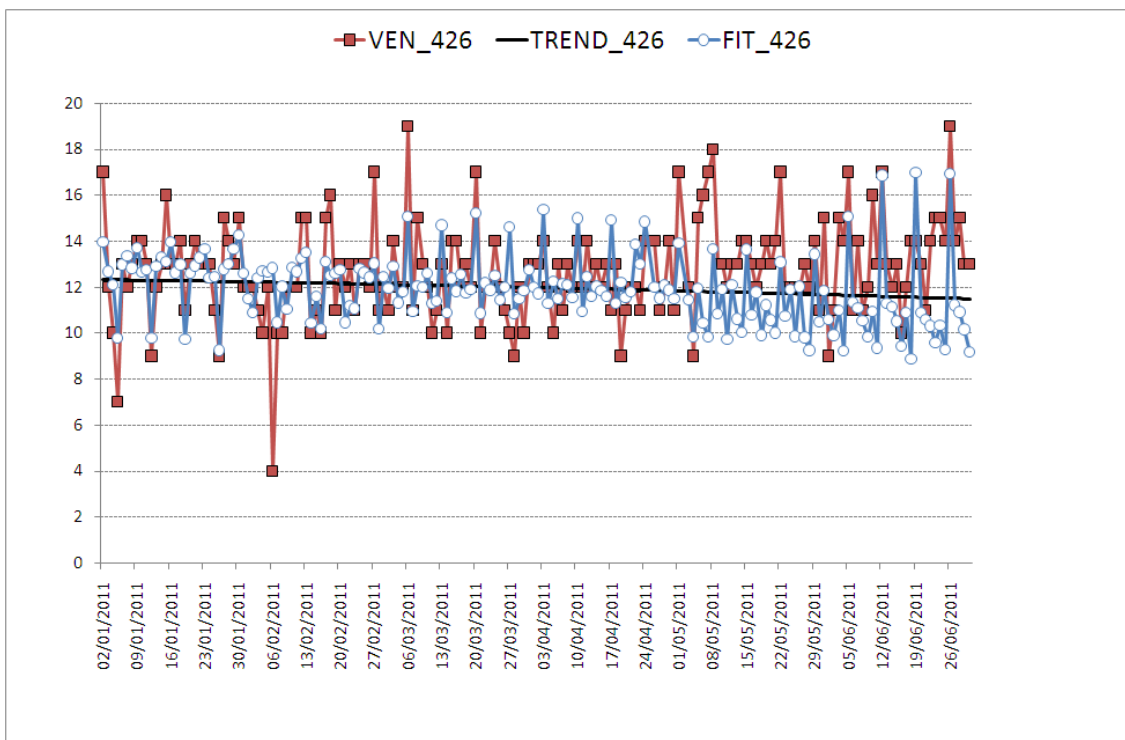


Figura 100. Vendite reali, trend e vendite previste per l'edicola 426

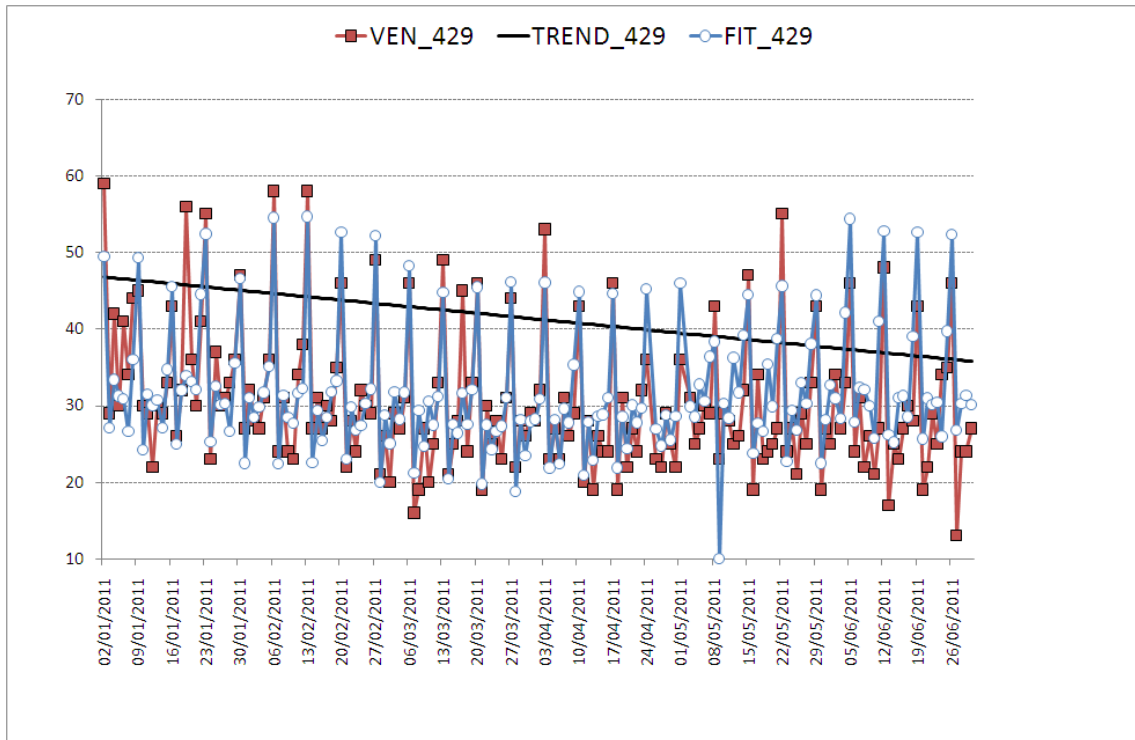


Figura 101. Vendite reali, trend e vendite previste per l'edicola 429

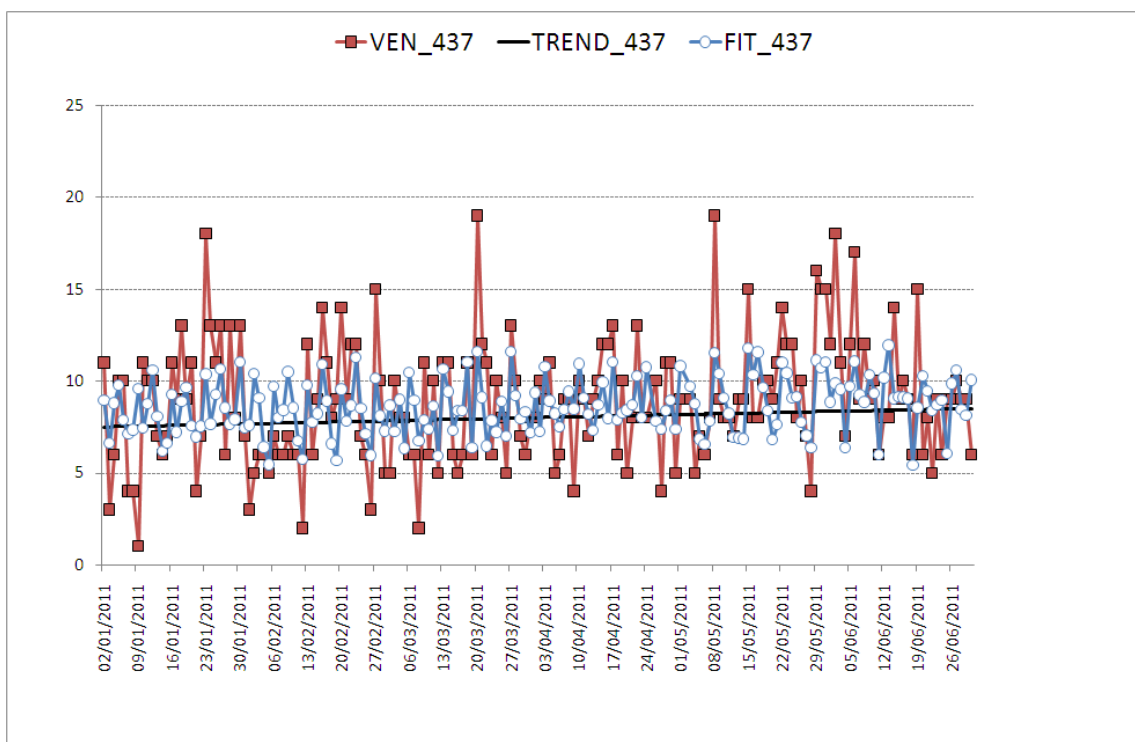


Figura 102. Vendite reali, trend e vendite previste per l'edicola 437

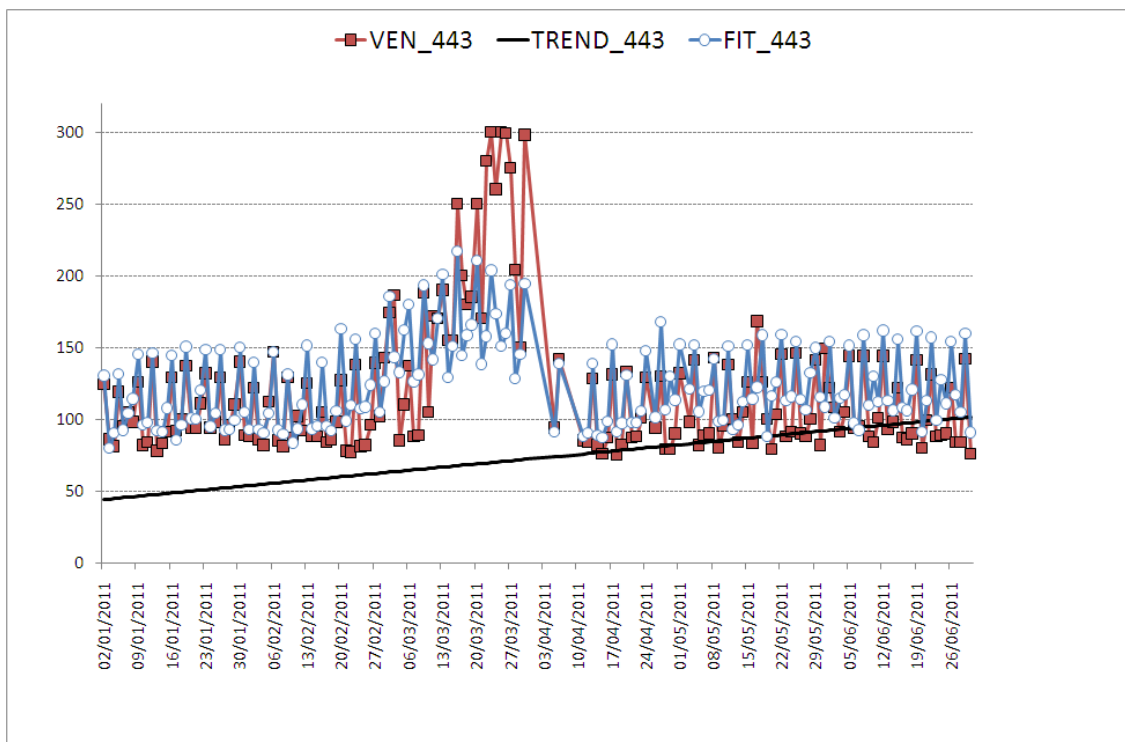


Figura 103. Vendite reali, trend e vendite previste per l'edicola 443

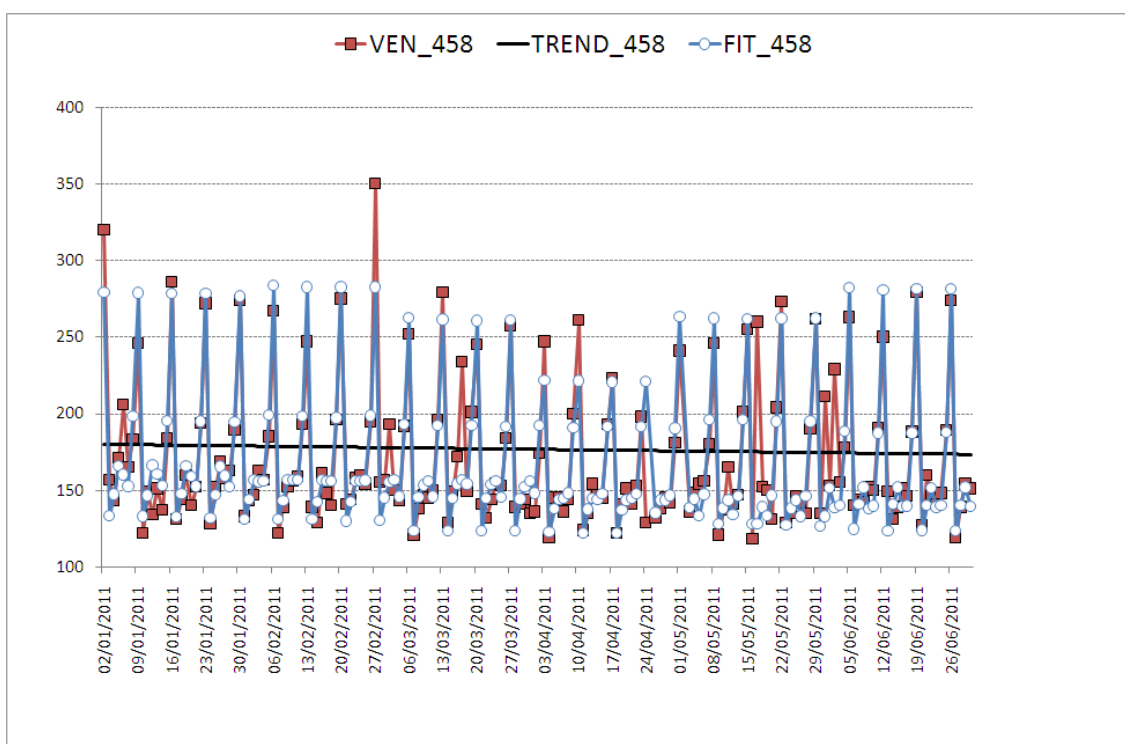


Figura 104. Vendite reali, trend e vendite previste per l'edicola 458

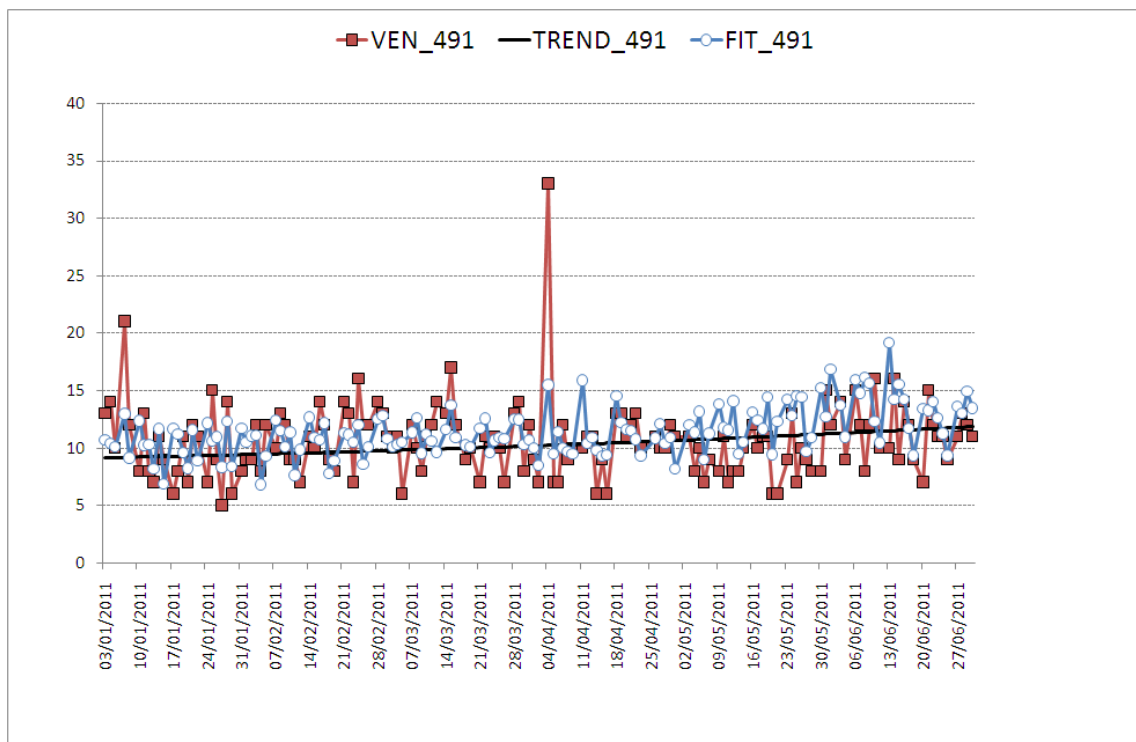


Figura 105. Vendite reali, trend e vendite previste per l'edicola 491

In linea generale, osservando i grafici delle 8 aree si nota che l'andamento delle vendite previste si avvicina a quello delle vendite reali. Tenendo conto poi che i valori di vendite coinvolti sono alti, a maggior ragione si può concludere che il modello costruito è affidabile e accurato.

Per quanto riguarda le edicole, 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 vendite piuttosto basse. Ciò è dovuto al coinvolgimento di altri fattori estranei al nostro modello, che risulta in questo senso poco accurato 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

Dalle analisi statistiche sui 18 mesi, ovvero da gennaio 2010 a giugno 2011, si può concludere che sono state individuate per ogni edicola e per ogni area delle variabili significative e, in particolare, che alcune di queste variabili risultano essere rilevanti per la maggior parte delle edicole/aree studiate. Per variabili significative intendiamo tutte quelle tipologie di notizia che influenzano le vendite del quotidiano studiato, sia in modo positivo che in modo negativo. Infatti, considerando le figure dalla Figura 53 alla Figura 60, si può notare che alcuni coefficienti hanno segno positivo mentre altri hanno segno negativo: tale risultato può essere interpretato, in generale, come un corrispondente aumento delle vendite in presenza di un coefficiente positivo e una diminuzione delle stesse in caso di coefficiente negativo.

Nelle figure considerate si possono, inoltre, osservare dei coefficienti pari a zero: equivale a concludere che la variabile considerata risulta essere non rilevante, ovvero non significativa per la vendita del giornale "Il Mattino di Padova".

Per portare degli esempi, citiamo "cronaca disastri ambientali", la quale è in assoluto la variabile più significativa: se consideriamo, ad esempio, l'edicola 040 notiamo che il relativo coefficiente in Figura 56 è 2.4863 e ciò equivale a affermare che una pagina di "cronaca disastri ambientali" aumenta le vendite di 2.4863 copie. Stesso ragionamento può essere effettuato per l'area Monselicense il cui il coefficiente, riferito sempre alla stessa variabile, è di 55.595 (Figura 54): una pagina di cronaca "disastri ambientali" comporta un incremento delle vendite di 55.595 copie.

Dalla Figura 53 e Figura 54, risulta immediato osservare che, per le 8 aree studiate, le notizie di tipo regionale, politico, sportivo e "cronaca incidenti mortali" sono particolarmente significative in modo positivo: da tale risultato si può concludere che a tali notizie può essere riservato ampio spazio nel giornale.

Per quel che concerne la notizia di tipo economico, notiamo che per la maggior parte delle edicole/aree considerate essa è non rilevante: risulta esserlo in modo positivo solo per l'edicola 313 (coefficiente pari a 0,7134 in Figura 55), mentre risulta essere significativa in termini negativi per tre edicole (Figura 59). In generale si può concludere che risulta non essere conveniente, al fine della vendita del quotidiano, dare spazio alle notizie di tipo economico.

Medesimo ragionamento può essere esteso anche alle notizie "cronaca furti", "cronaca suicidi" e "cronaca violenza sessuale". In particolare si può notare che, in riferimento al

tipo di notizia "cronaca suicidi", tale variabile risulta essere significativa in modo negativo per le edicole 510 e 426 e non rilevante per tutte le altre edicole/aree analizzate.

BIBLIOGRAFIA

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

Montgomery

Levine