

VENDITE DI PRODOTTI BIOLOGICI IN UNA CATENA DI SUPERMERCATI

Donò Filippo, Garbin Sara, Zamengo Silvia



INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

DOMANDA DI RICERCA

Analizzare le vendite di **prodotti biologici** al fine di ottimizzare le strategie di marketing



- Analizzare la domanda di prodotti biologici nell'arco di un anno
Sono presenti trend e stagionalità?
- Studiare la propensione all'acquisto dei prodotti biologici in funzione delle altre variabili disponibili
Vale la pena proporre una varietà di offerta di prodotti biologici?
- Valutare le associazioni tra i prodotti biologici e gli altri
Come disporre i prodotti negli scaffali?
Proporre promozioni mirate in base alle abitudini dei consumatori

INDICE

1. Domanda di ricerca
- 2. Caratteristiche del dataset**
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

CARATTERISTICHE DEL DATASET

- 100087 ACQUISTI in una catena di supermercati
- 28 VARIABILI

CLIENTE_ID

NEGOZIO_ID

QUANTITÀ

IMPORTO

BIO

Altre 16 categorie di prodotti per esempio: DOP, FRESCHI, CASA, PERSONA, MULTIMEDIA...

Informazioni sul momento d'acquisto (GIORNO DEL MESE, GIORNO DELLA SETTIMANA, SETTIMANA, MESE, ANNO, TRIMESTRE)

N° SCONTRINO

FONTE: Esame di data mining del 6 febbraio 2017

INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

SISTEMAZIONE DATASET

- Variabili eliminate:

- **N° SCONTRINO**

- **ANNO (2012)**

- Modalità eliminate:

- **NEGOZIO_ID**

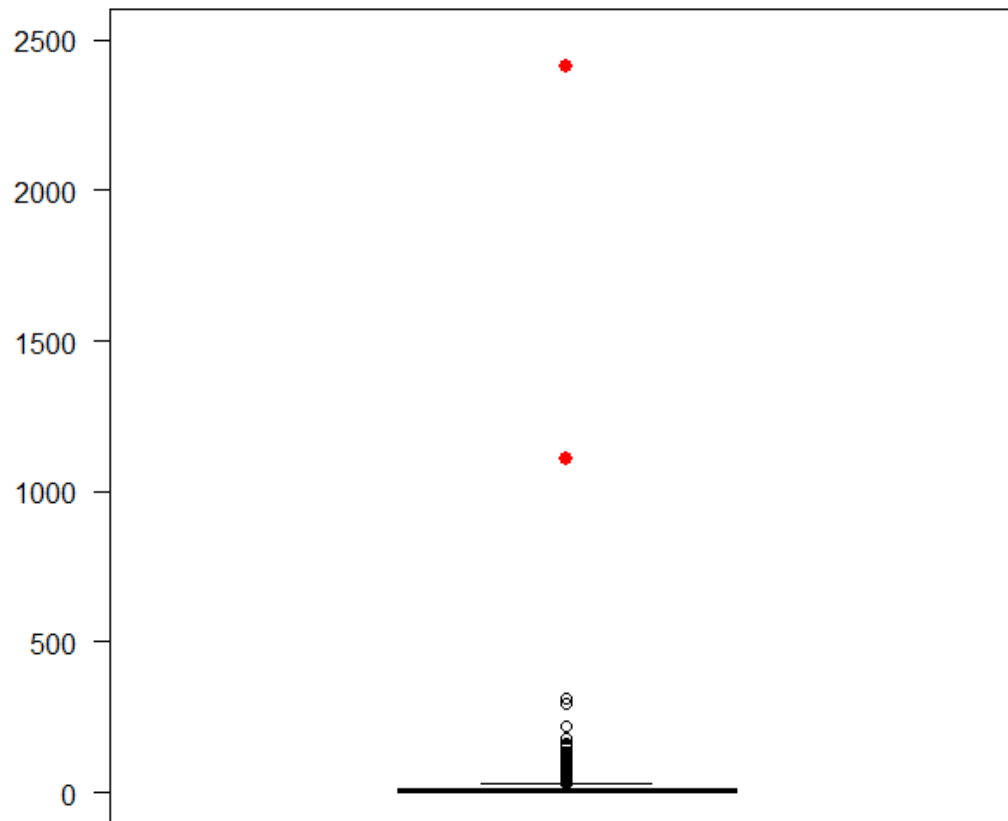
Avevamo 23 negozi, ma due li abbiamo eliminati poiché con poche osservazioni e nessuno acquistava prodotti biologici

- Variabile **CLIENTE_ID** (n° 1249)

ANALISI ESPLORATIVE

- VARIABILE **QUANTITÀ**

Numero di prodotti acquistati in ogni transazione

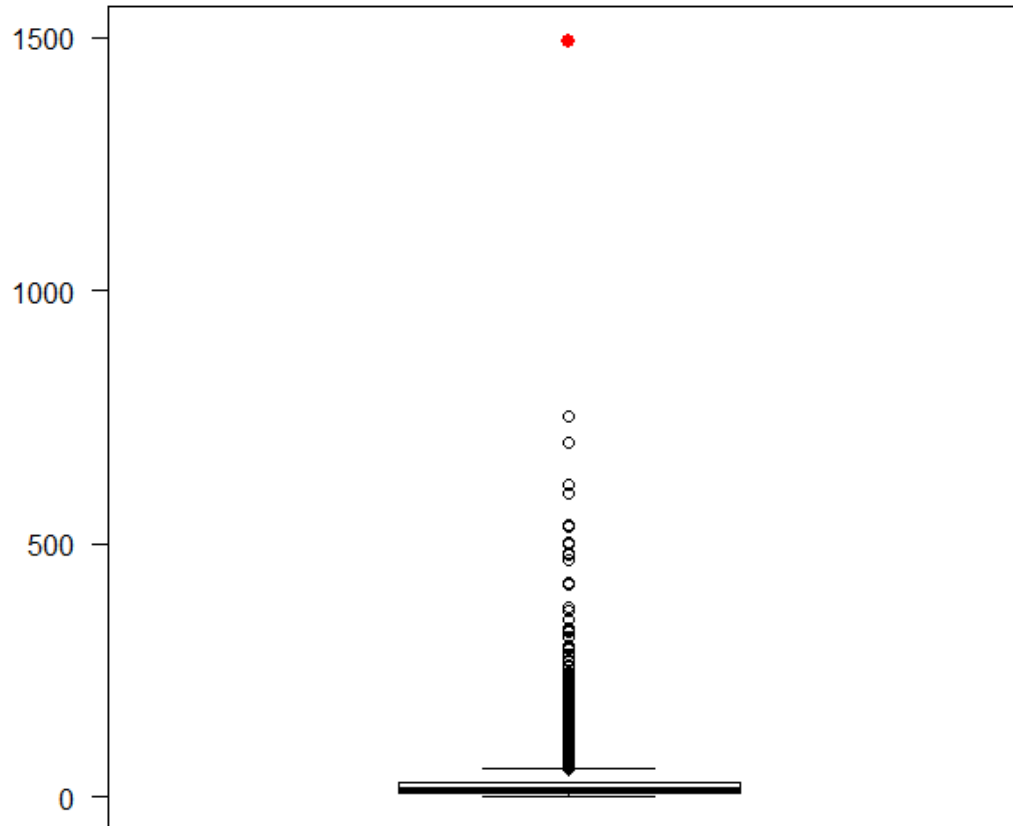


ANALISI ESPLORATIVE

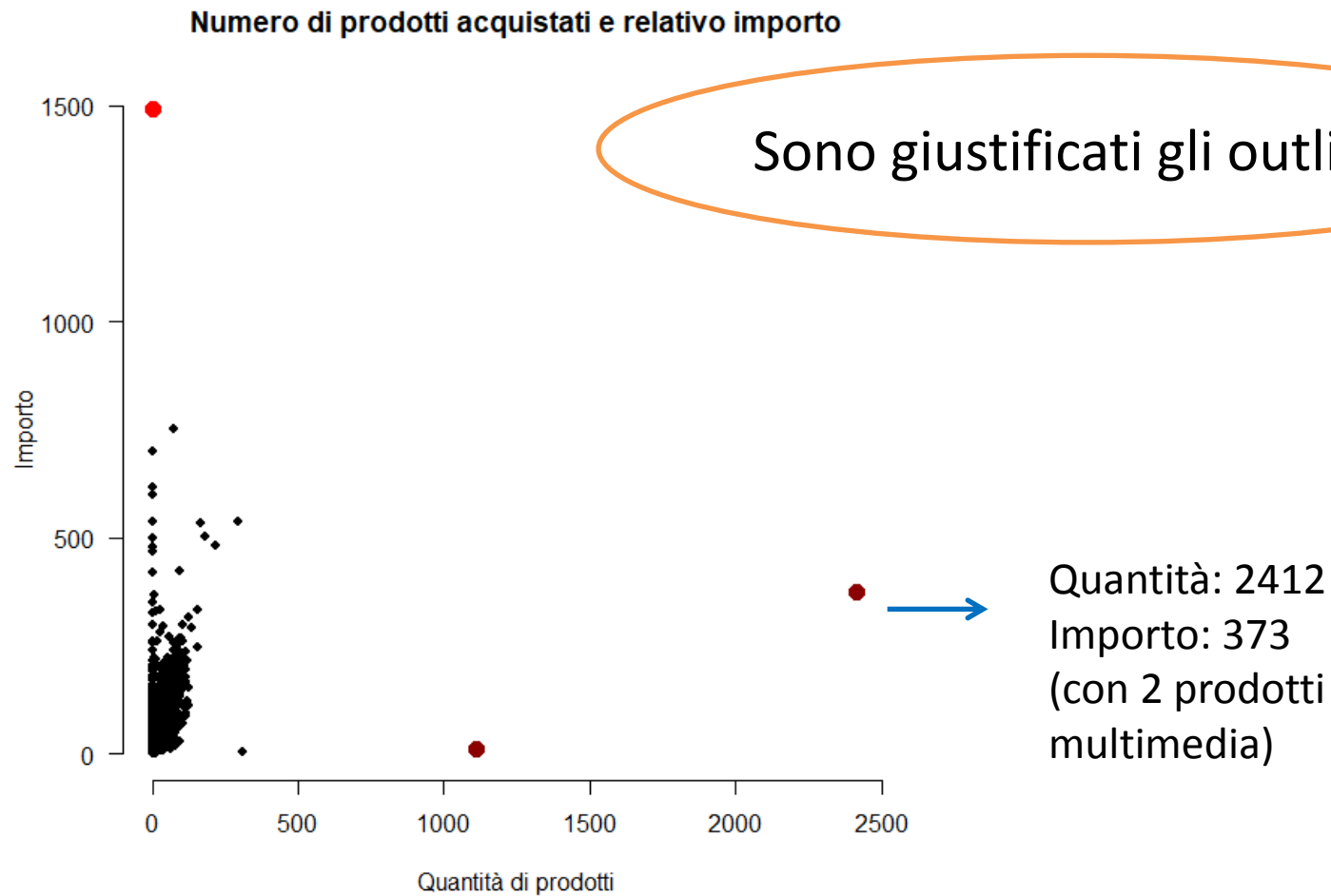
- VARIABILE **IMPORTO**



Importo speso per ogni acquisto



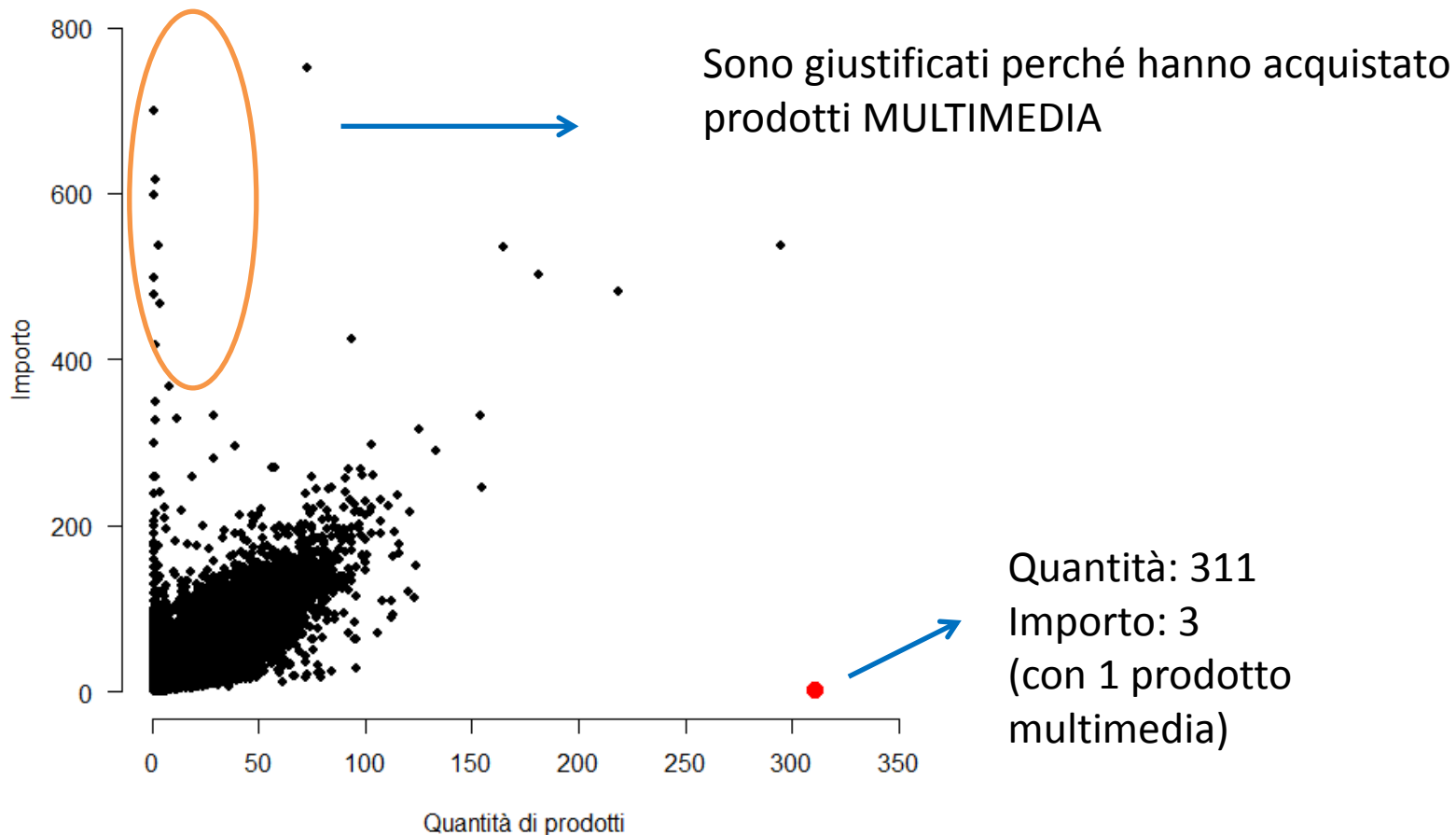
- **RELAZIONE** tra la variabile **QUANTITÀ** e **IMPORTO**



ANALISI ESPLORATIVE

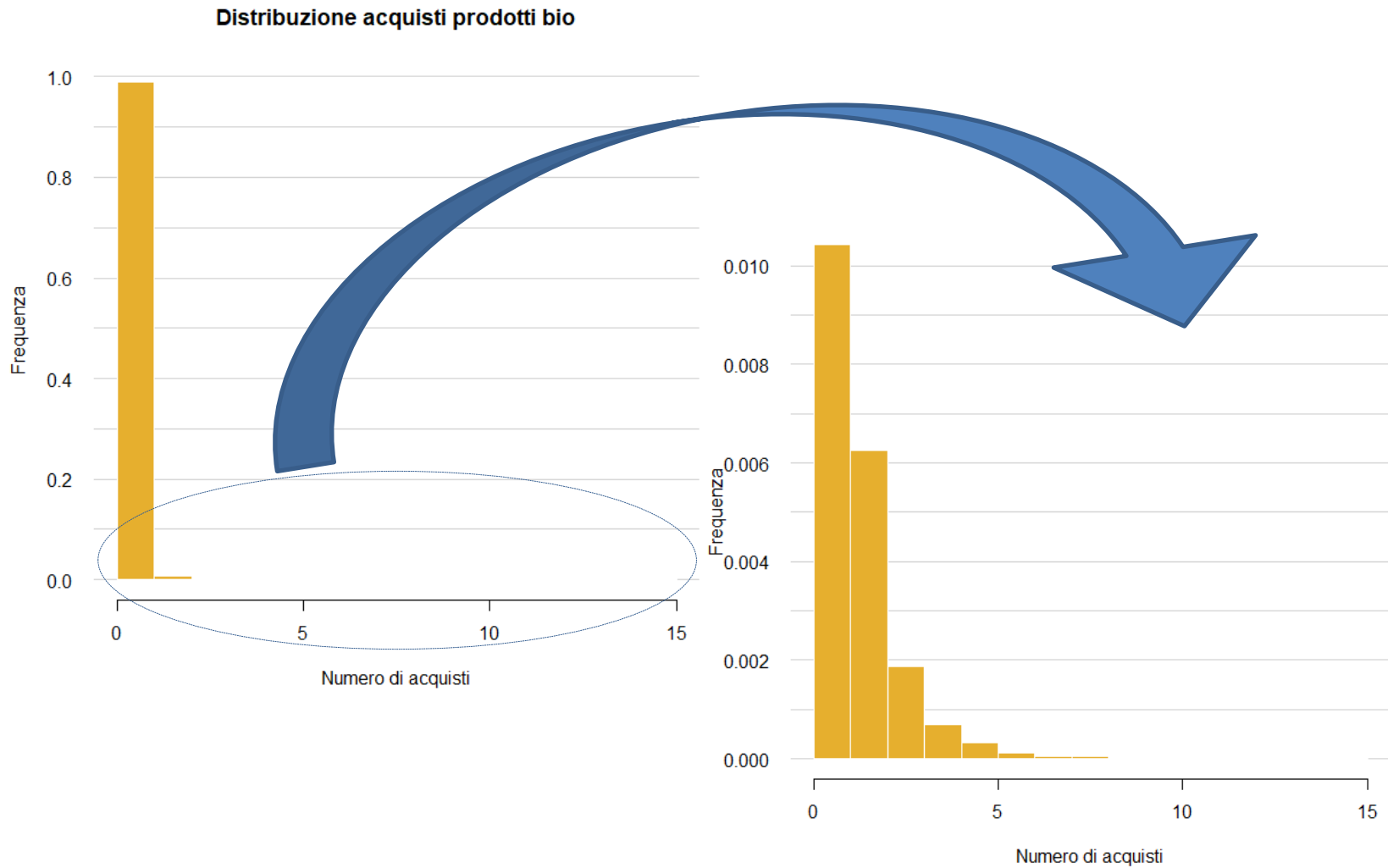
-> Abbiamo dunque rimosso queste osservazioni
Emerge però un altro outlier che abbiamo eliminato

Numero di prodotti acquistati e relativo importo



ANALISI ESPLORATIVE

- **VARIABILE BIO**

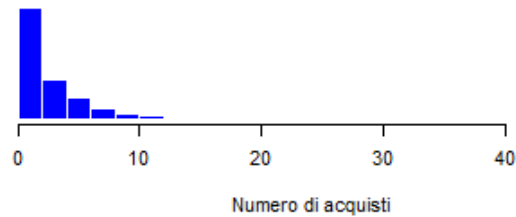


ANALISI ESPLORATIVE

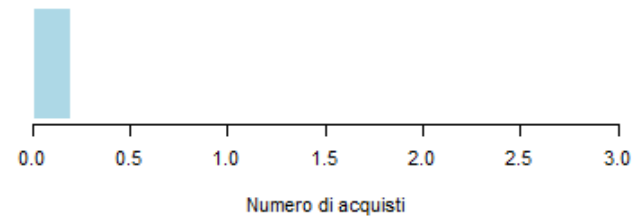
- VARIABILI relative al numero di acquisti per CATEGORIE DI PRODOTTO

■ variabili numeriche
■ variabili rese dicotomiche

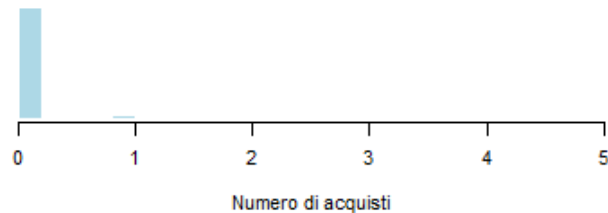
BRAND



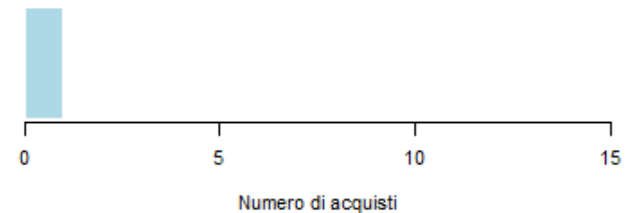
SOLIDALE



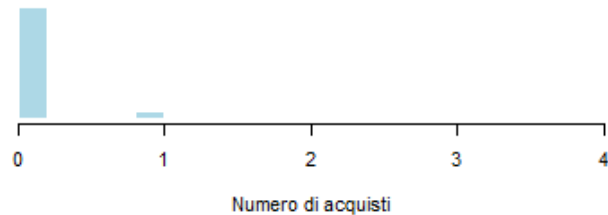
TIPICO



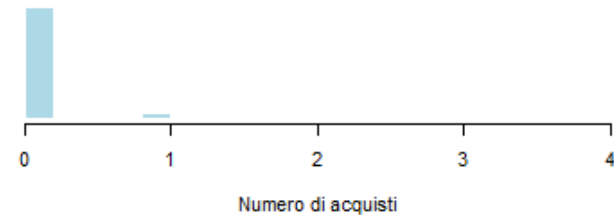
CELIACI



DOP



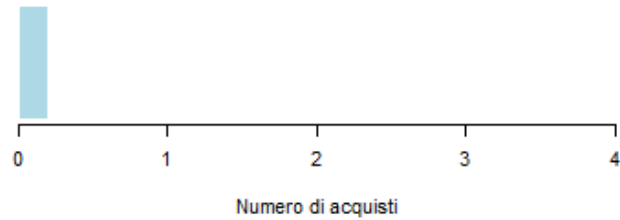
IGP



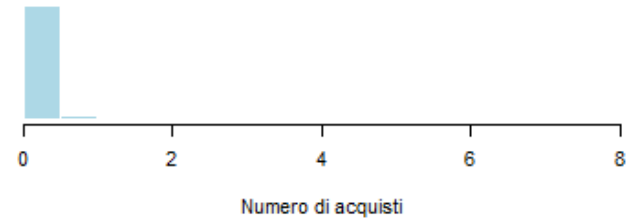
ANALISI ESPLORATIVE

■ variabili numeriche
■ variabili rese dicotomiche

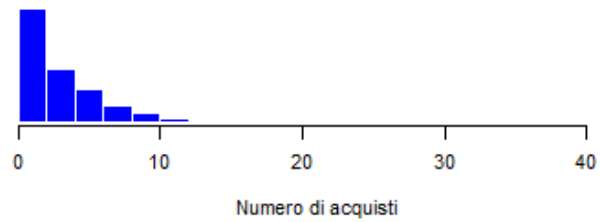
DOCG



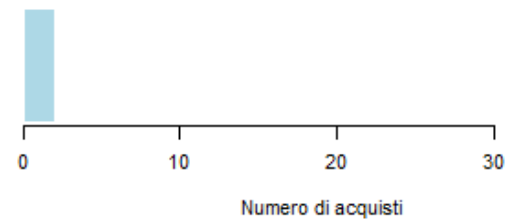
CASA



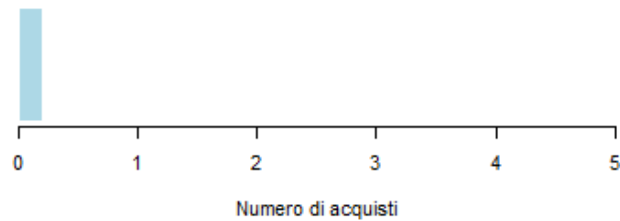
FRESCHISSIMI



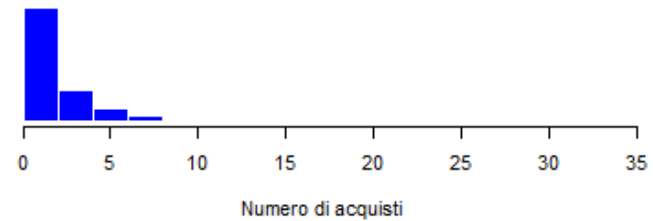
STAGIONALI



SALUTE



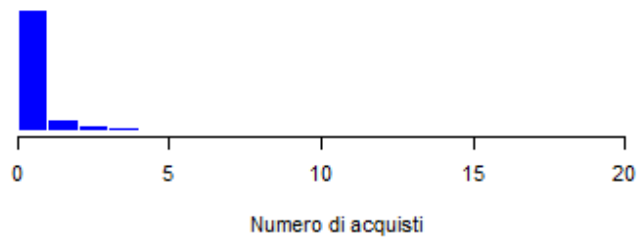
FRESCHI



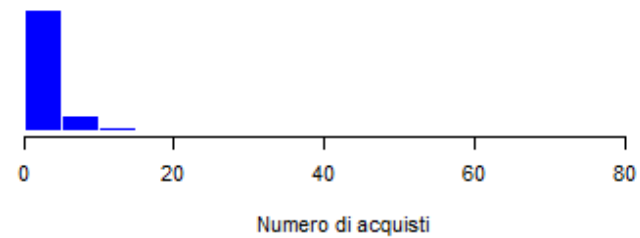
ANALISI ESPLORATIVE

■ variabili numeriche
■ variabili rese dicotomiche

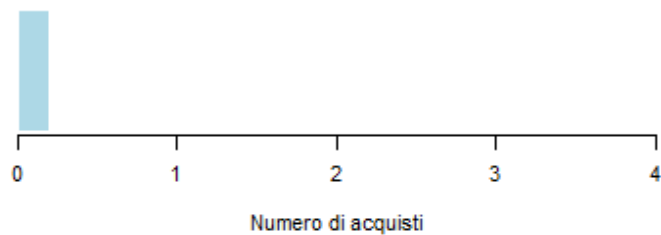
CHIMICA



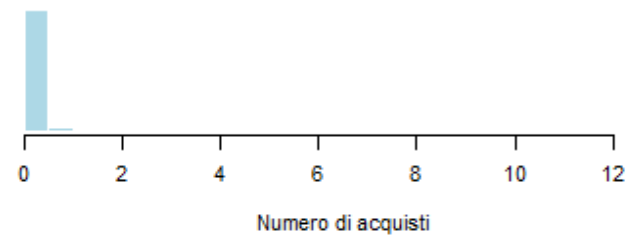
GROCERY_ALIMENTARI



MULTIMEDIA



PERSONA



- **VARIABILI TEMPORALI**

DISPONIBILI: GIORNO DEL MESE, GIORNO DELLA SETTIMANA, SETTIMANA, MESE, TRIMESTRE

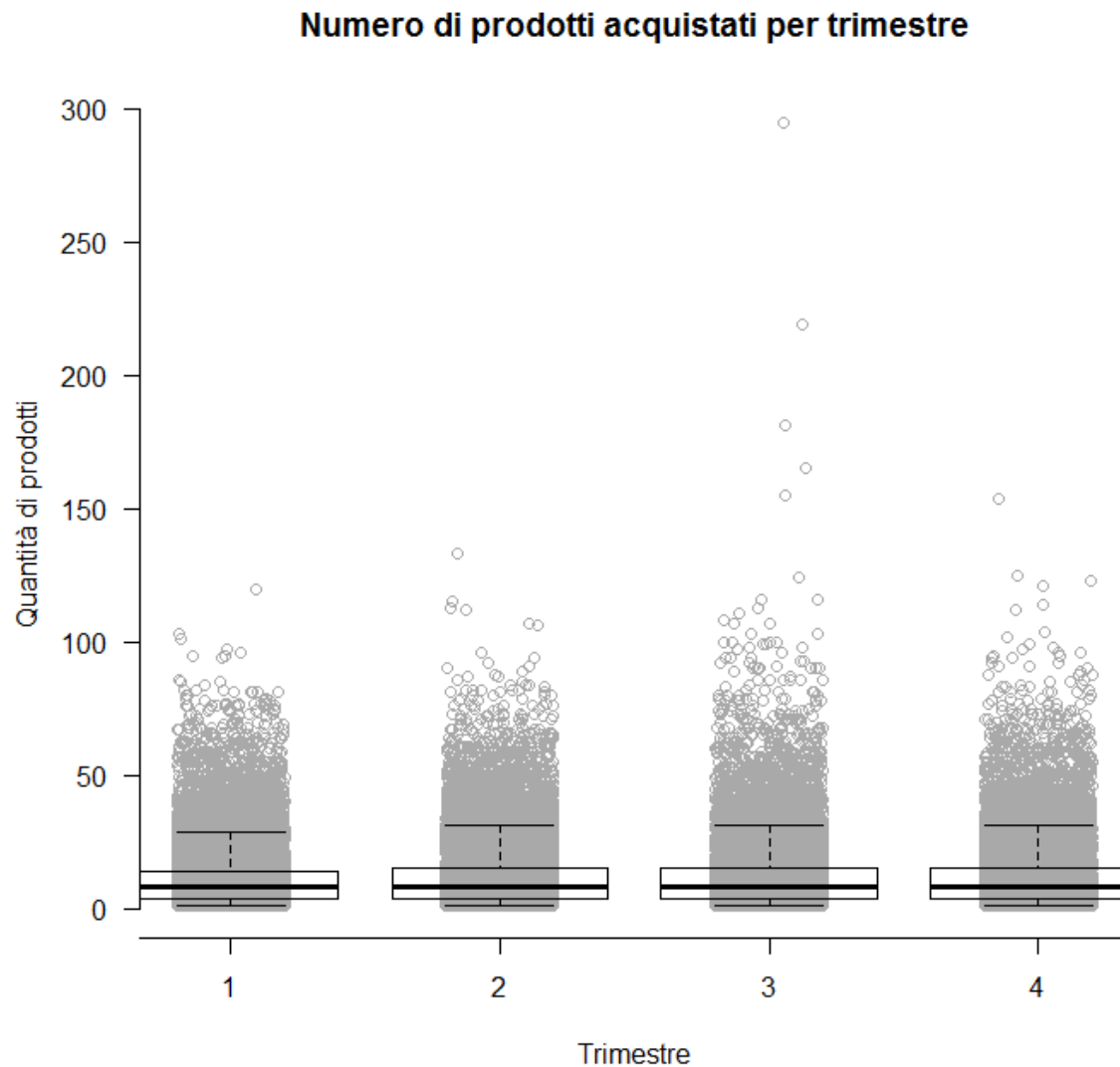
CALCOLATA : GIORNO DELL' ANNO



Con particolare focus per
l'analisi temporale delle vendite
di prodotti BIOLOGICI

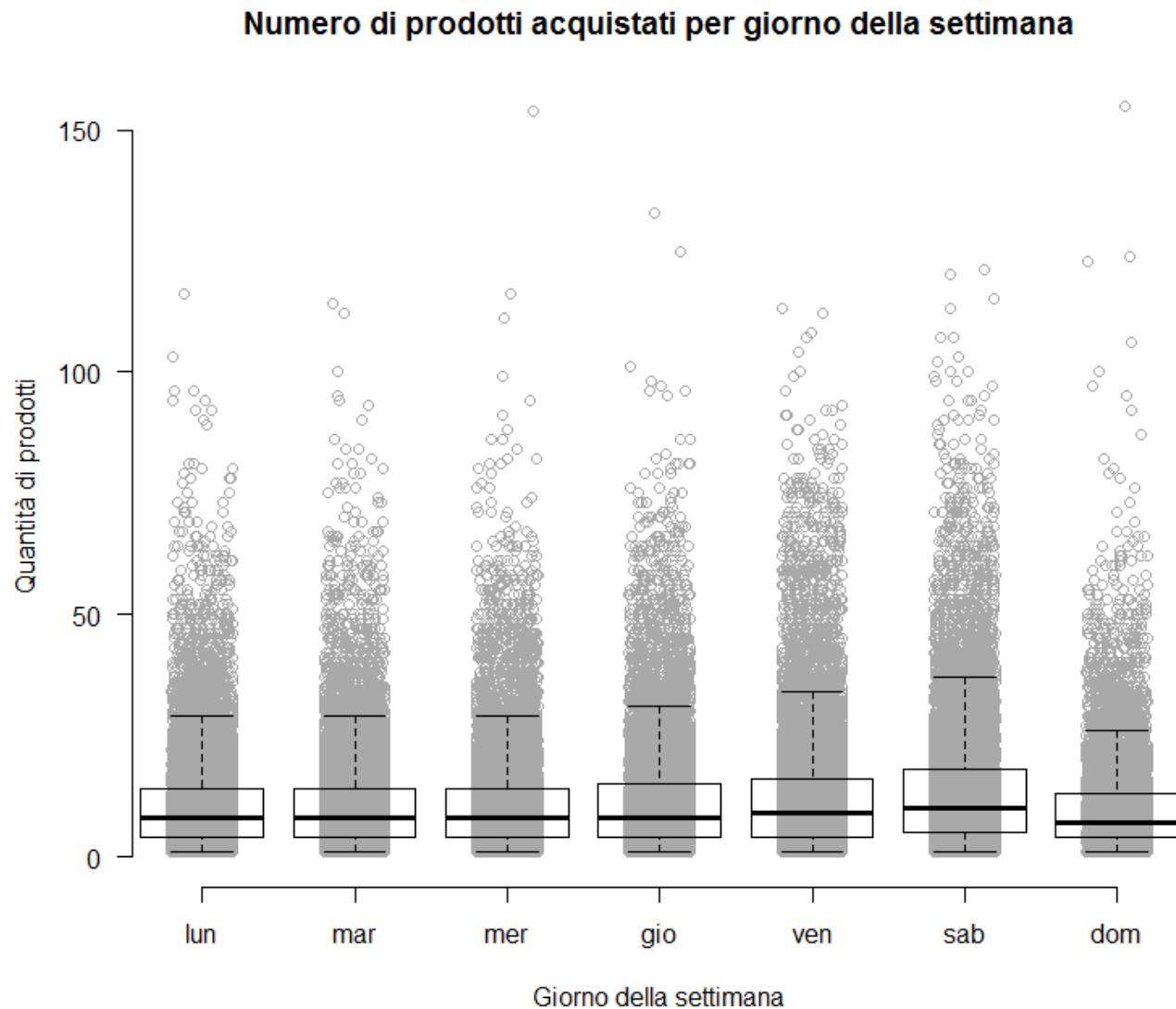
ANALISI ESPLORATIVE

- Andamento della **QUANTITÀ DI PRODOTTI** nei TRIMESTRI



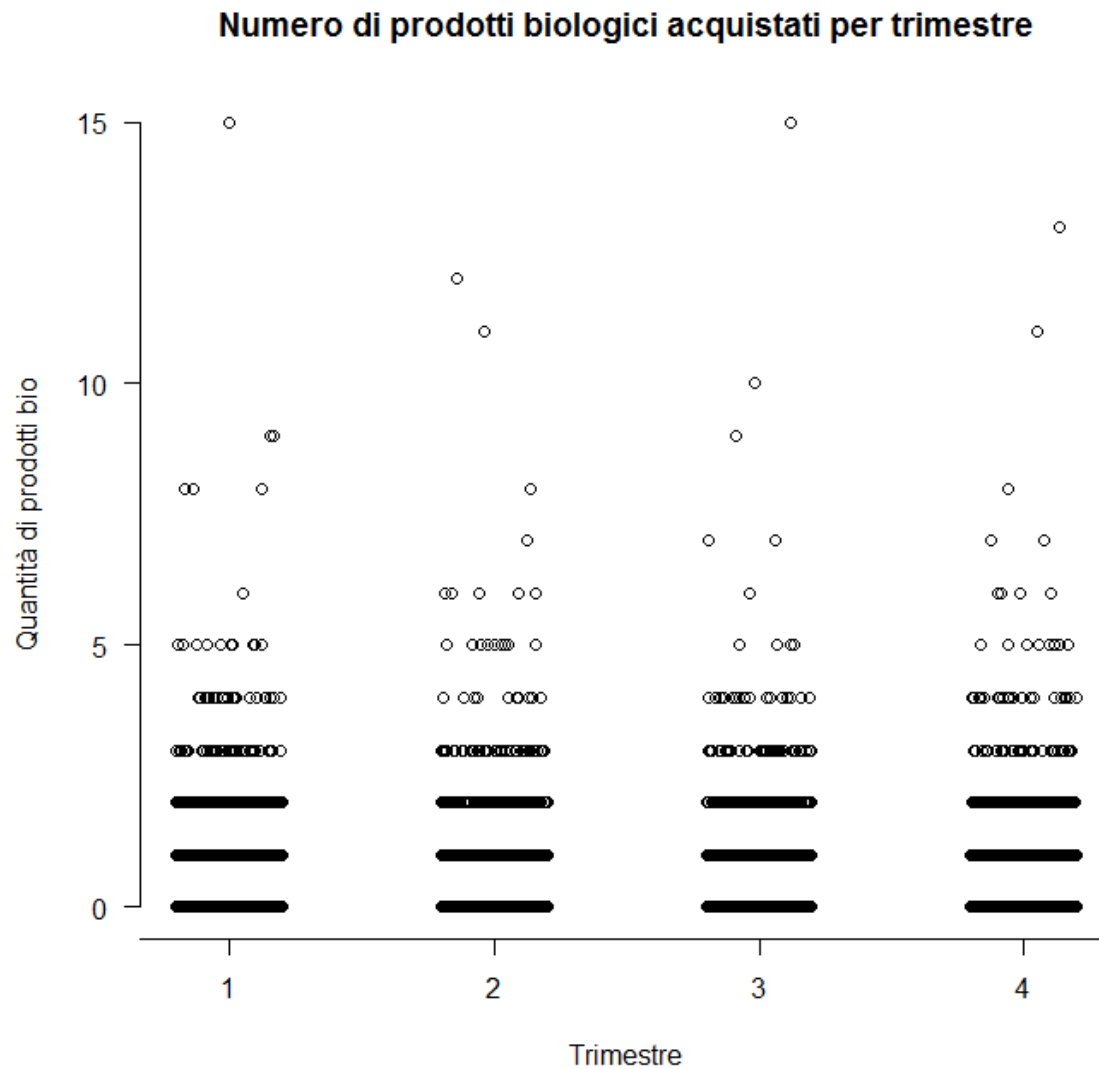
ANALISI ESPLORATIVE

- Andamento della **QUANTITÀ DI PRODOTTI** nella SETTIMANA



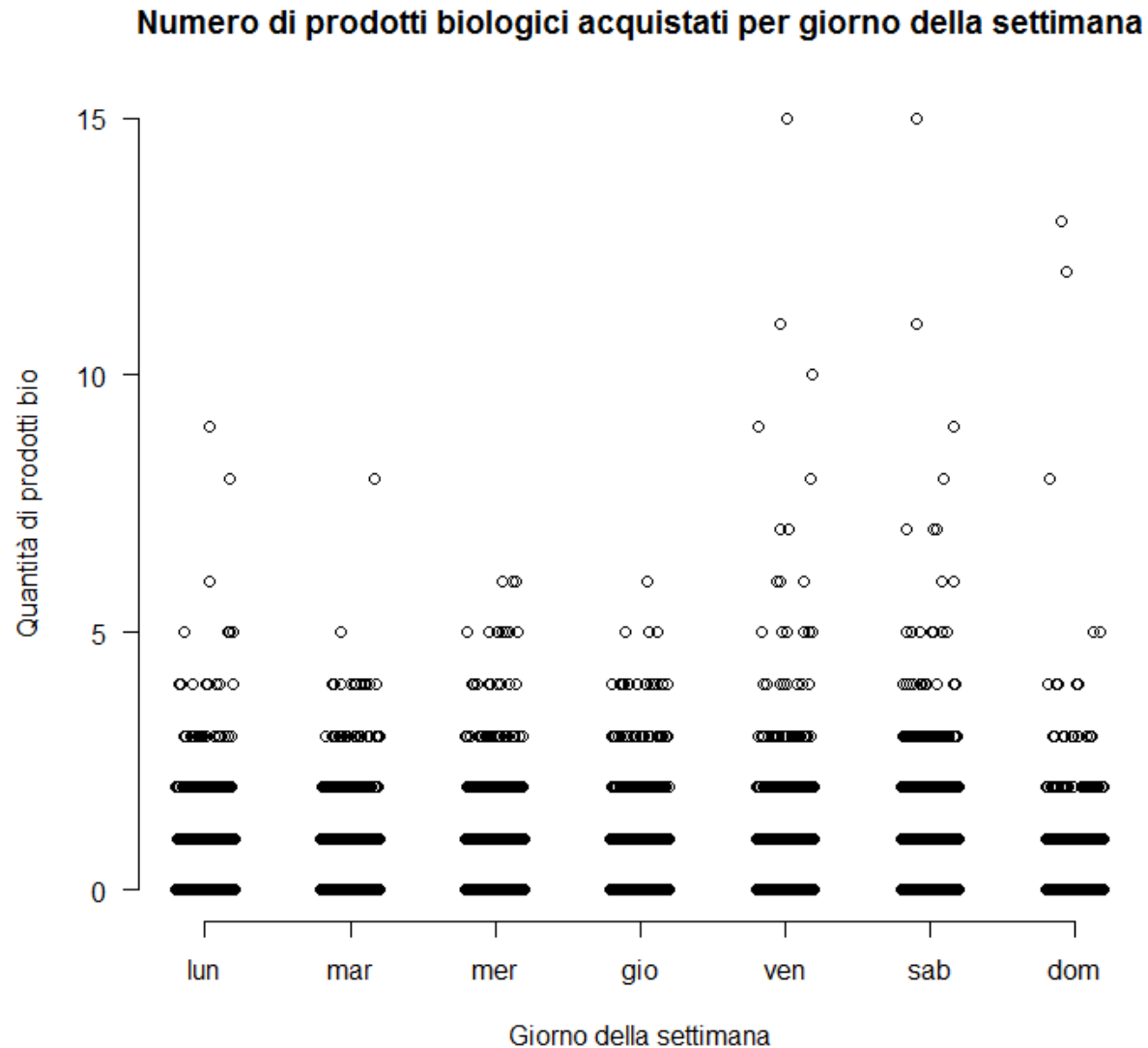
ANALISI ESPLORATIVE

- Andamento della **QUANTITÀ DI PRODOTTI BIOLOGICI** nei TRIMESTRI



ANALISI ESPLORATIVE

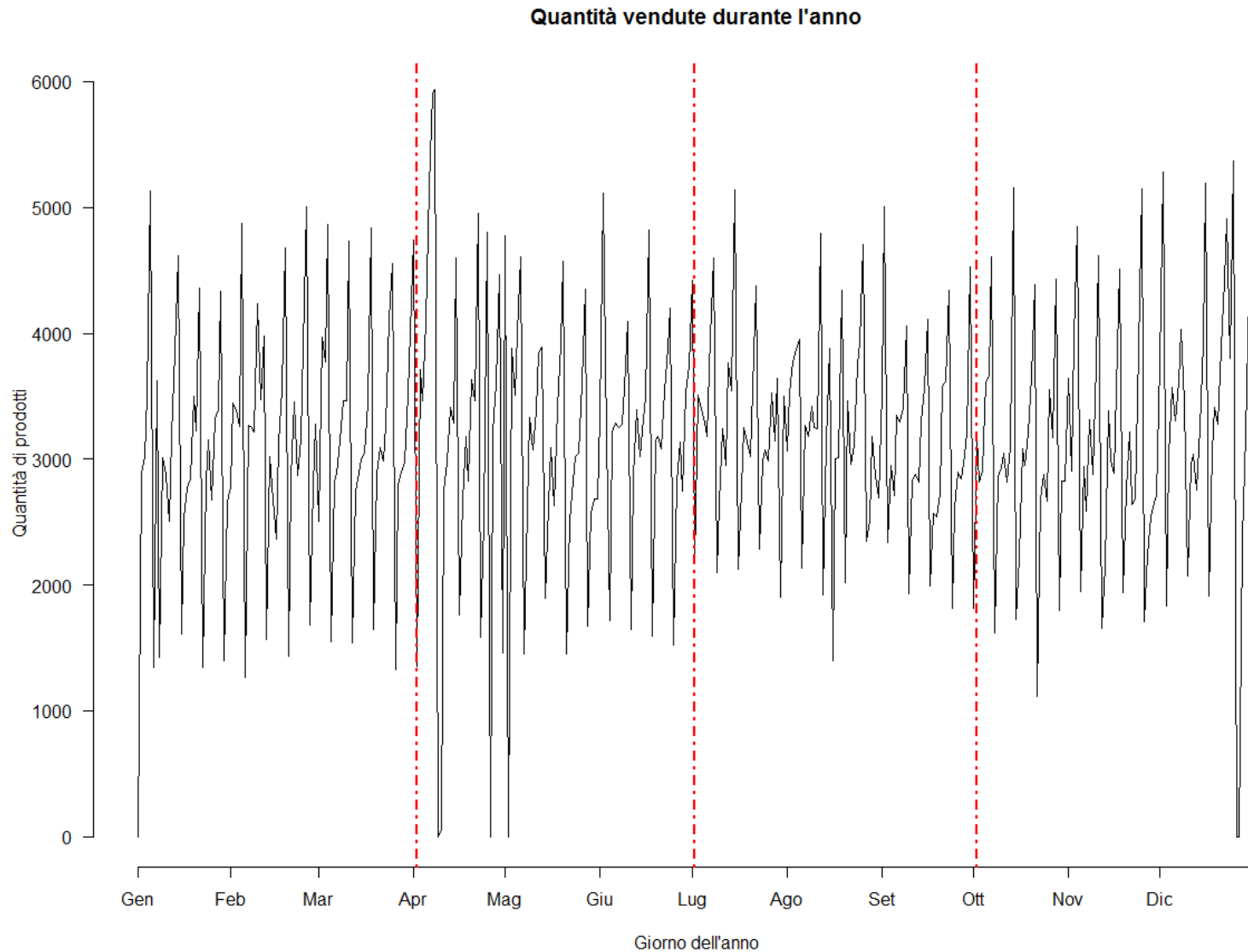
- Andamento della **QUANTITÀ DI PRODOTTI BIOLOGICI** nella SETTIMANA



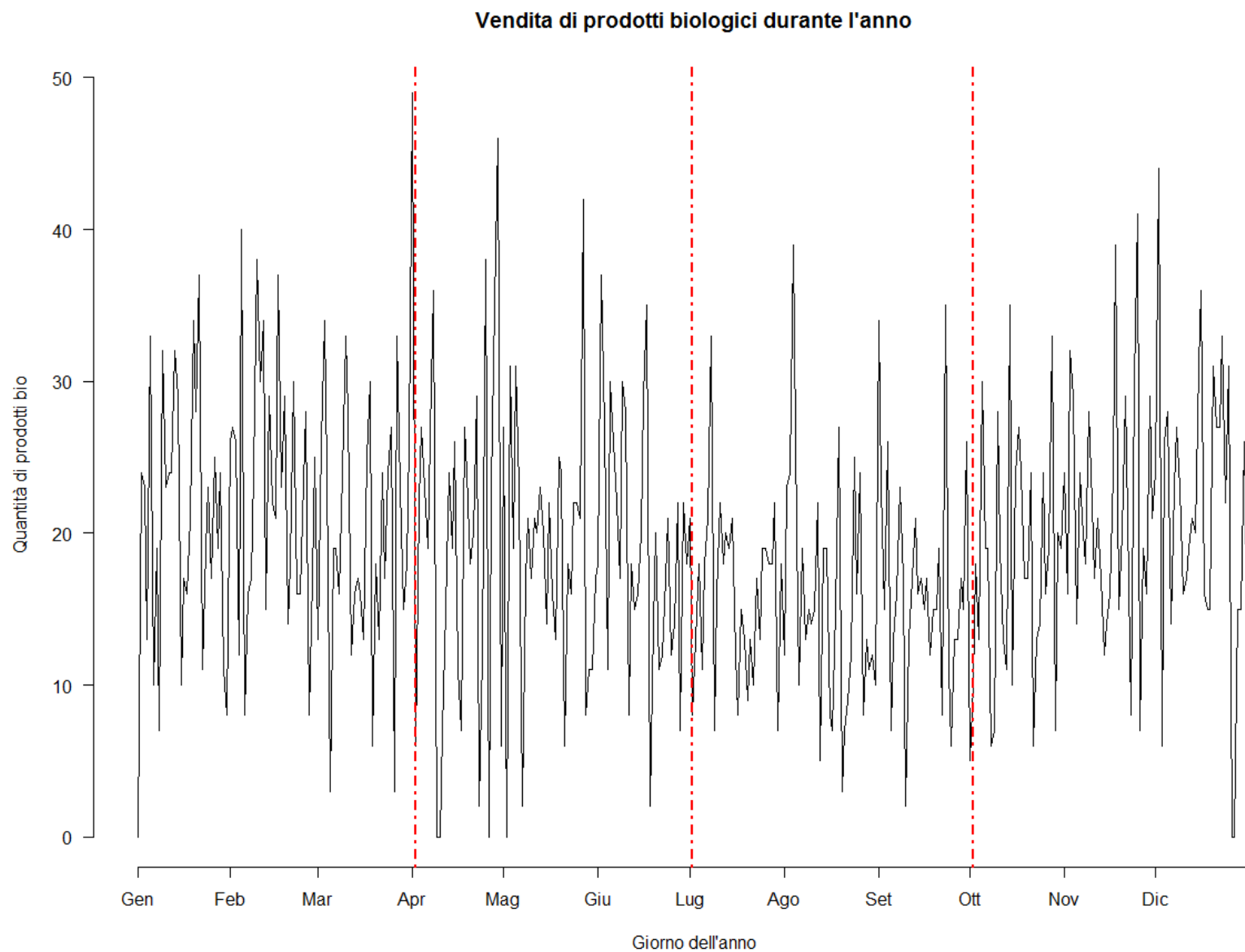
INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

SERIE STORICHE

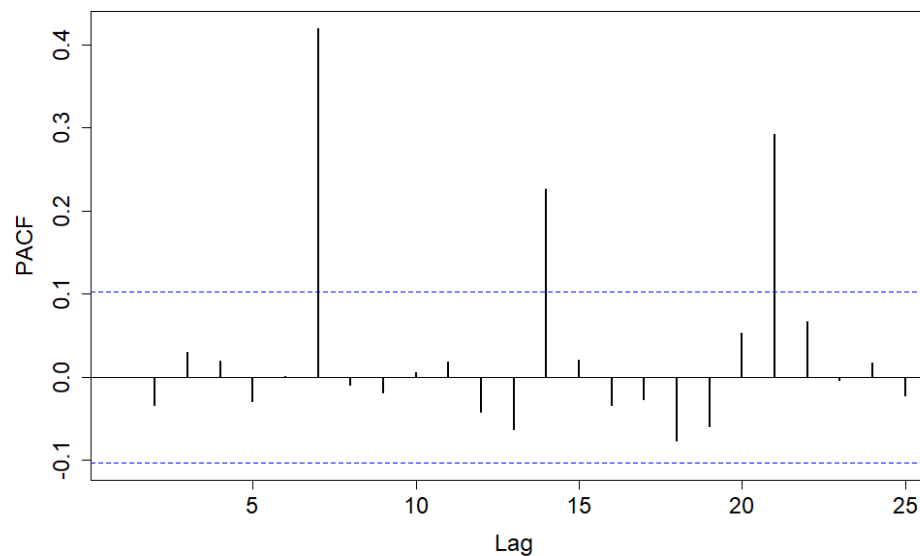
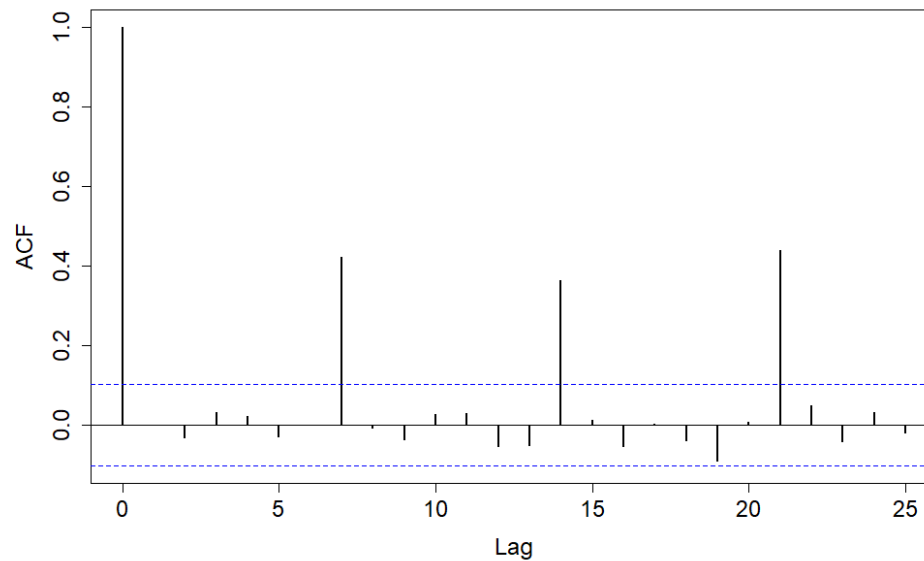


SERIE STORICHE



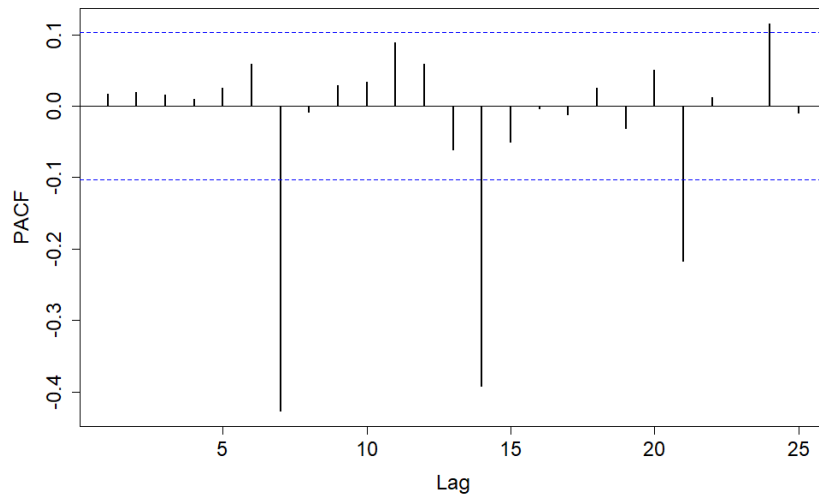
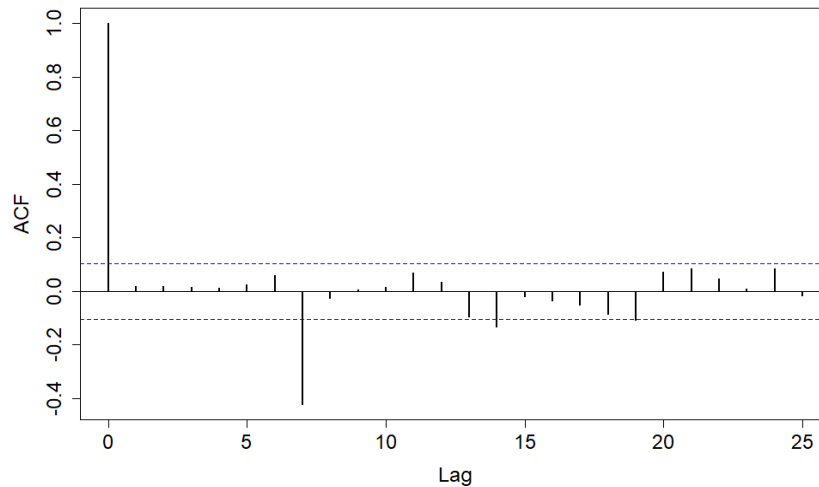
SERIE STORICHE

- **ACF** e **PACF** della serie della vendita di prodotti biologici



SERIE STORICHE

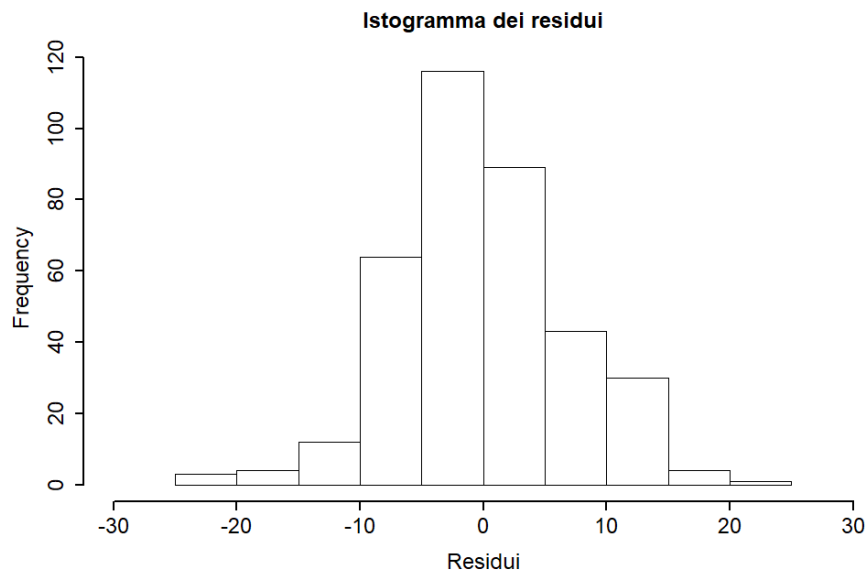
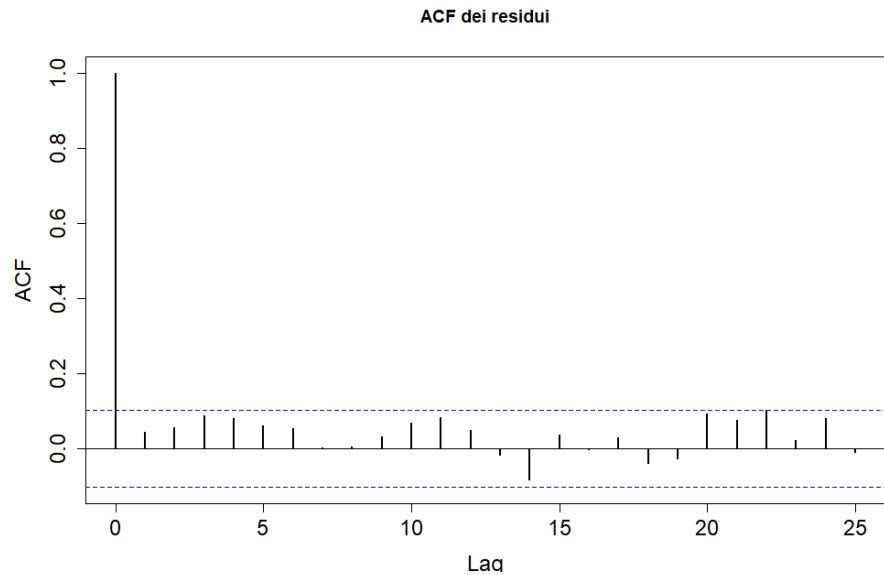
- **ACF** e **PACF** della serie della vendita di prodotti biologici differenziata



$\text{SARIMA}(0,0,0)(0,1,1)_7$

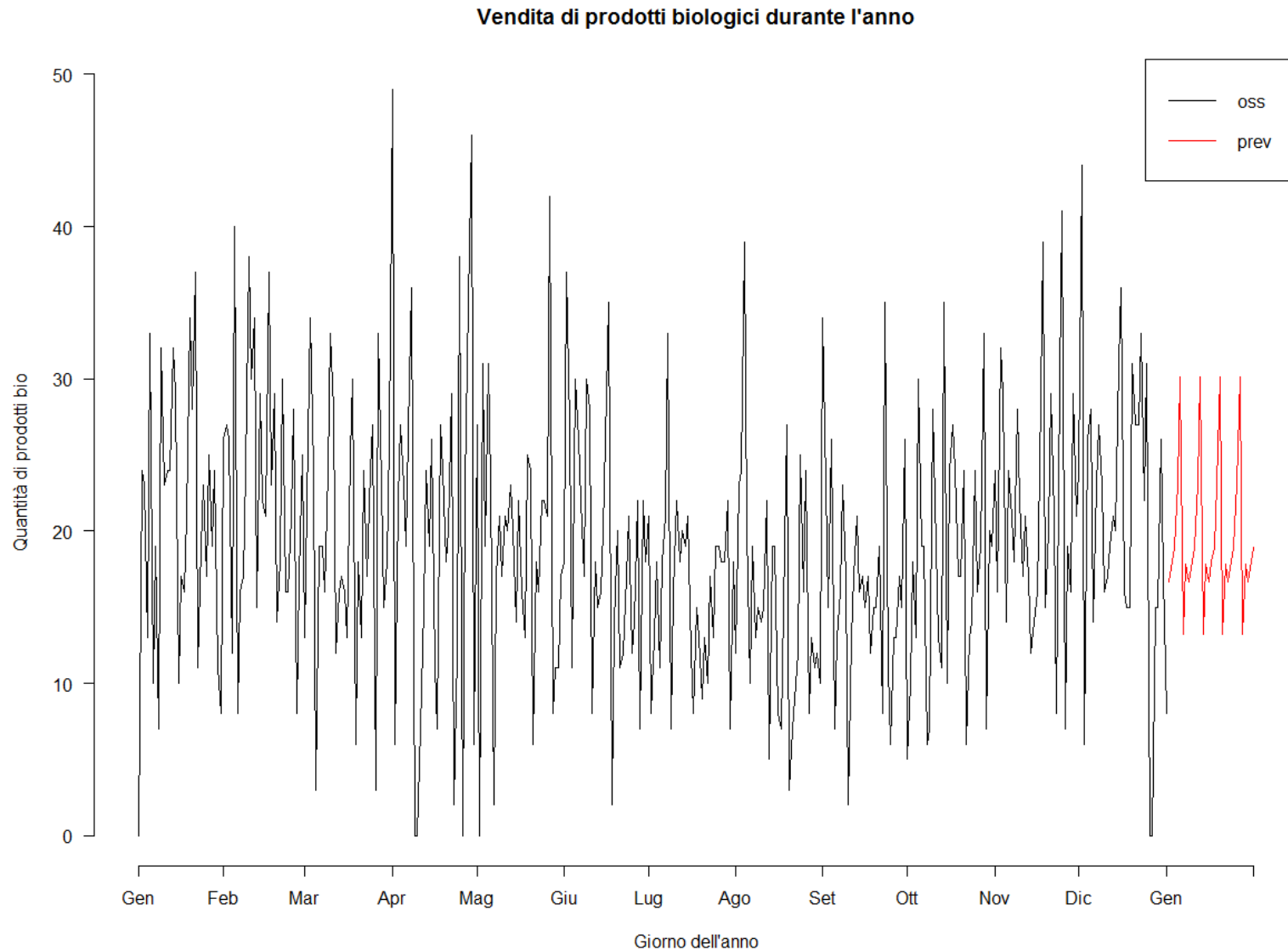
SERIE STORICHE

- ANALISI DEI RESIDUI** - Modello SARIMA(0,0,0)(0,1,1)₇



SERIE STORICHE

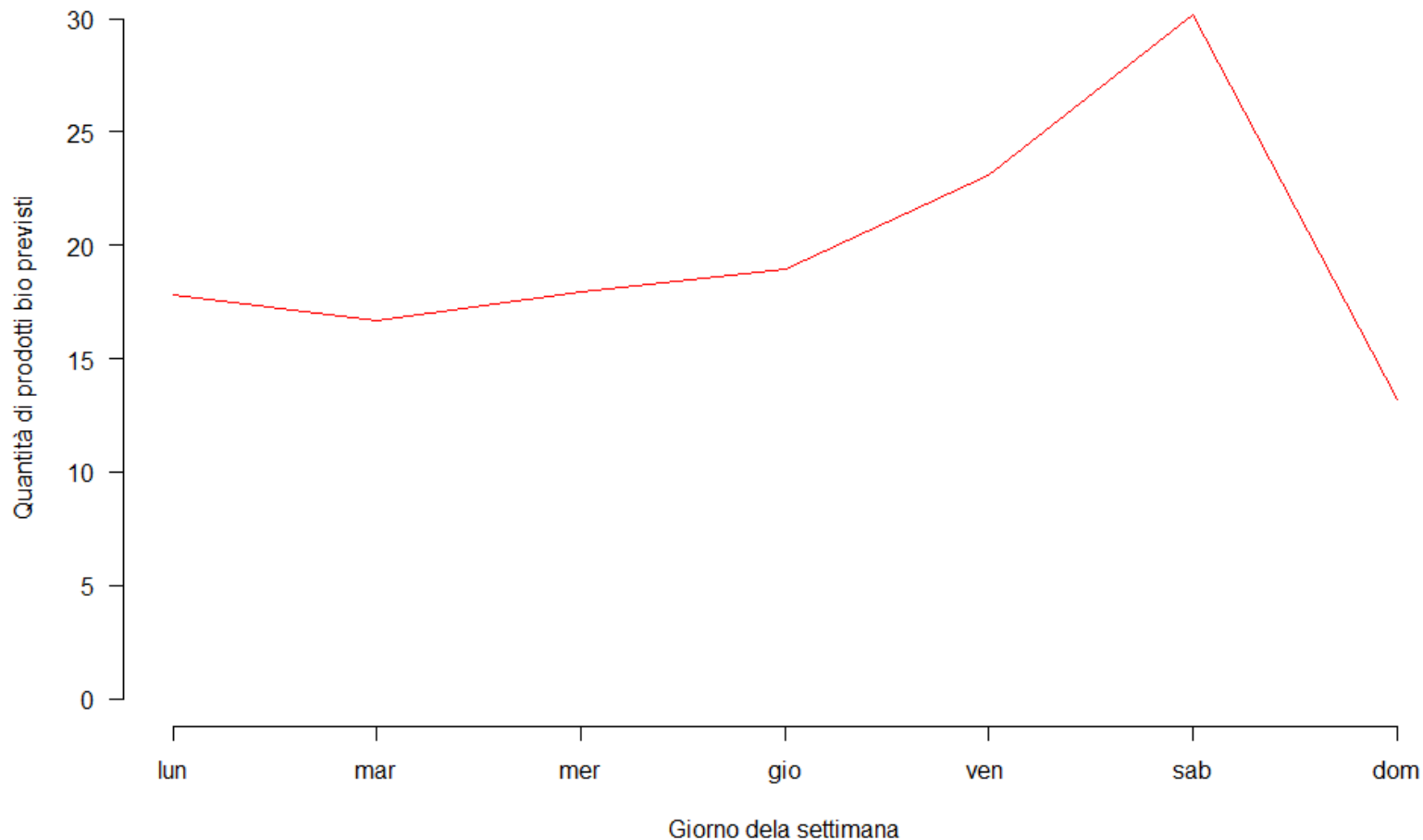
- VALORI OSSERVATI e PREVISTI** - Modello SARIMA(0,0,0)(0,1,1)₇



SERIE STORICHE

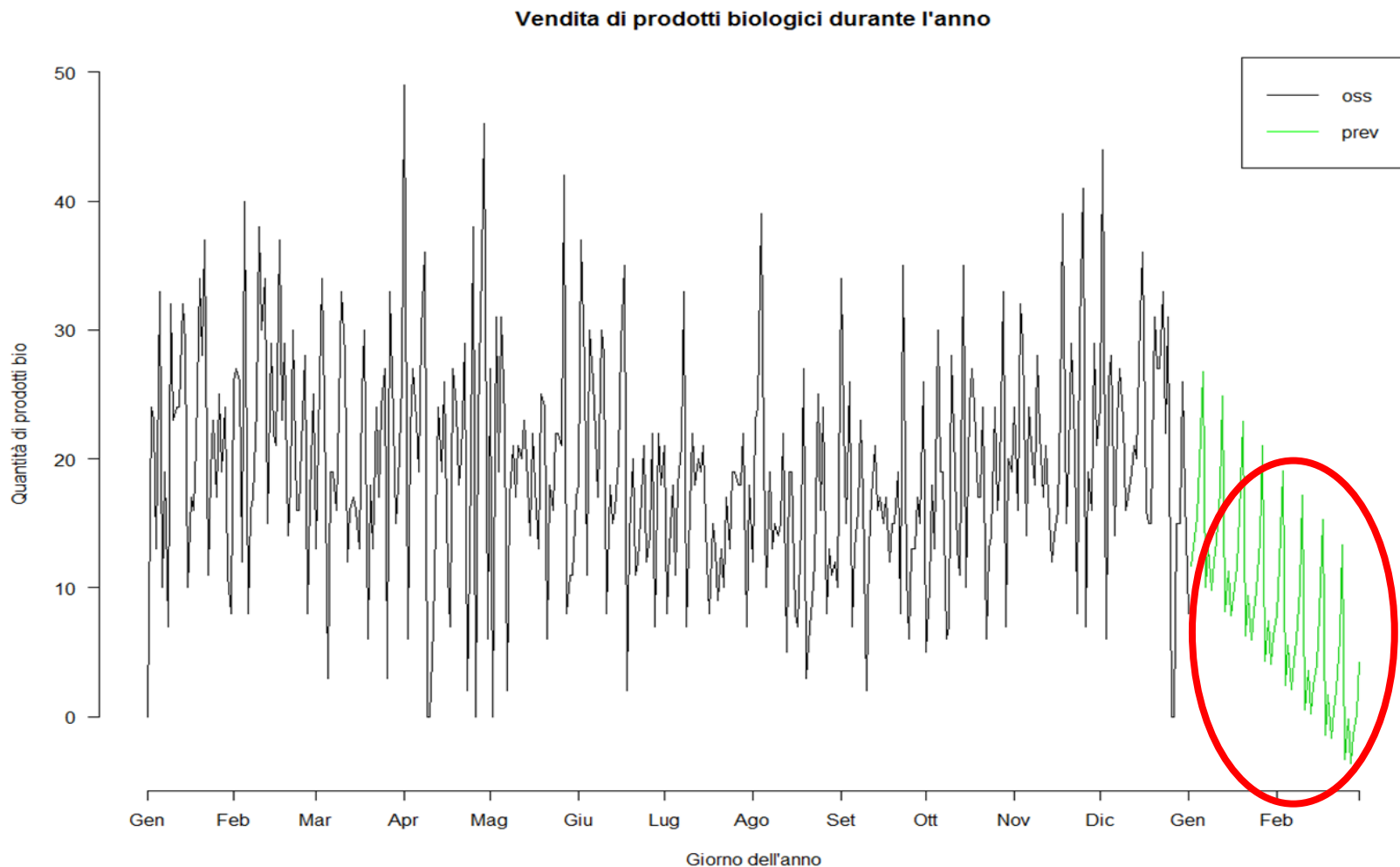
- ZOOM PREVISIONE SETTIMANALE**

Previsione settimanale vendita di prodotti biologici



MODELLO DI HOLT-WINTERS: tiene conto della stagionalità

- **VALORI OSSERVATI e PREVISTI**



INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
- 5. Modelli di classificazione**
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

MODELLI DI CLASSIFICAZIONE

- Con questi modelli vogliamo prevedere, per ogni transazione commerciale, **se viene acquistato almeno un prodotto biologico**



- Rendiamo dicotomica la variabile BIO:

$$y = \begin{cases} 1, & BIO > 0 \\ 0, & BIO = 0 \end{cases}$$

MODELLI DI CLASSIFICAZIONE

- Prima di stimare i modelli....
- Divisione del dataset:

75 % per il TRAINING SET
25 % per il TEST SET

- Ma analizzando il dataset di STIMA:

NON ACQUISTA BIO	ACQUISTA BIO
94,7%	5,3%

- **FOCUS DELLA RICERCA:** L'acquisto di prodotti biologici



DATASET DI STIMA BILANCIATO

MODELLI DI CLASSIFICAZIONE

- Come facciamo a trovare il **MIGLIOR MODELLO**?
- Il nostro obiettivo è quello di trovare un modello, tra quelli possibili, che sia al tempo stesso interpretabile e abbia buone performance rispetto a:

1. TASSO DI ERRATA CLASSIFICAZIONE

2. PERCENTUALE DI FALSI POSITIVI

3. PERCENTUALE DI FALSI NEGATIVI

MODELLI DI CLASSIFICAZIONE

MODELLO	TASSO ERRATA CLASS. %	% FALSI POSITIVI	% FALSI NEGATIVI
Regressione lineare (stepwise)	22 ,21%	21,14 %	39,55 %
Regressione logistica (stepwise)	15,44 %	13 ,14%	52,82 %
Modello lineare gerarchico (stepwise)	22,23 %	21 ,20%	39,00 %
Modello logistico gerarchico (stepwise)	15 ,36%	13,05 %	52,35%
Mars	31,68%	32,04 %	25,86%
Albero di classificazione	22 ,04%	20,86%	41 ,13%
Bagging	33,32 %	33,79 %	25,72%
Boosting	31 ,26%	31,65 %	24,97 %
Random forest	32,33 %	32,78%	24,90 %
Modello Naive di Bayes	30,55%	30,62 %	29,37 %

MODELLI DI CLASSIFICAZIONE

MODELLO	TASSO ERRATA CLASS. %	% FALSI POSITIVI	% FALSI NEGATIVI
Regressione lineare (stepwise)	22 ,21%	21,14 %	39,55 %
Regressione logistica (stepwise)	15,44 %	13 ,14%	52,82 %
Modello lineare gerarchico (stepwise)	22,23 %	21 ,20%	39,00 %
Modello logistico gerarchico (stepwise)	15 ,36%	13,05 %	52,35%
Mars	31,68%	32,04 %	25,86%
Albero di classificazione	22 ,04%	20,86%	41 ,13%
Bagging	33,32 %	33,79 %	25,72%
Boosting	31 ,26%	31,65 %	24,97 %
Random forest	32,33 %	32,78%	24,90 %
Modello Naive di Bayes	30,55%	30,62 %	29,37 %

MODELLI DI CLASSIFICAZIONE

Il modello di classificazione migliore in termini di tasso di errata classificazione è il **MODELLO LOGISTICO GERARCHICO**

Alcune considerazioni:

- Considera comportamenti diversi a livello di negozi
- Facilmente interpretabile

Criticità:

- % di FALSI NEGATIVI ALTA



ASSORTIMENTO PRODOTTI BIO < DOMANDA



MODELLI DI CLASSIFICAZIONE

Output del modello logistico gerarchico con selezione delle variabili basata sull'AIC

Random effects:

Groups	Name	Variance	Std.Dev.
NEGOZIO_ID	(Intercept)	0.03889	0.1972

Number of obs: 7944, groups: NEGOZIO_ID, 21

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.384070	0.089276	-15.503	< 2e-16	***
QUANTITA	-0.011982	0.005731	-2.091	0.03656	*
IMPORTO	0.005341	0.002240	2.384	0.01713	*
BRAND	0.191421	0.013884	13.787	< 2e-16	***
SOLIDALETRUE	2.516239	0.305391	8.239	< 2e-16	***
TIPICOTRUE	0.451219	0.147396	3.061	0.00220	**
DOCGTRUE	0.380877	0.192011	1.984	0.04730	*
FRESCHISSIMI	0.148543	0.011541	12.871	< 2e-16	***
STAGIONALITRUE	-0.267287	0.102444	-2.609	0.00908	**
FRESCHI	-0.045292	0.014840	-3.052	0.00227	**
CHIMICA	-0.069427	0.021063	-3.296	0.00098	***
GROCERY_ALIMENTARI	0.039011	0.014302	2.728	0.00638	**
PERSONATRUE	-0.297572	0.161133	-1.847	0.06478	.
TRIMESTRE_N2	-0.180790	0.072352	-2.499	0.01246	*
TRIMESTRE_N3	-0.206938	0.073671	-2.809	0.00497	**
TRIMESTRE_N4	-0.055319	0.071884	-0.770	0.44157	

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

MODELLI DI CLASSIFICAZIONE

Output del modello logistico gerarchico con selezione delle variabili basata sull'AIC

- EFFETTI FISSI

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.384070	0.089276	-15.503	< 2e-16	***
QUANTITA	-0.011982	0.005731	-2.091	0.03656	*
IMPORTO	0.005341	0.002240	2.384	0.01713	*
BRAND	0.191421	0.013884	13.787	< 2e-16	***
SOLIDALETRUE	2.516239	0.305391	8.239	< 2e-16	***
TIPICOTRUE	0.451219	0.147396	3.061	0.00220	**
DOCGTRUE	0.380877	0.192011	1.984	0.04730	*
FRESCHISSIMI	0.148543	0.011541	12.871	< 2e-16	***
STAGIONALITRUE	-0.267287	0.102444	-2.609	0.00908	**
FRESCHI	-0.045292	0.014840	-3.052	0.00227	**
CHIMICA	-0.069427	0.021063	-3.296	0.00098	***
GROCERY_ALIMENTARI	0.039011	0.014302	2.728	0.00638	**
PERSONATRUE	-0.297572	0.161133	-1.847	0.06478	.
TRIMESTRE_N2	-0.180790	0.072352	-2.499	0.01246	*
TRIMESTRE_N3	-0.206938	0.073671	-2.809	0.00497	**
TRIMESTRE_N4	-0.055319	0.071884	-0.770	0.44157	

MODELLI DI CLASSIFICAZIONE

Output del modello logistico gerarchico con selezione delle variabili basata sull'AIC

- EFFETTI CASUALI

```
$`NEGOZIO_ID`  
  (Intercept)  
1  0.0905031870  
2  0.0226252563  
3  0.0127626860  
4  0.1411705695  
5  0.2045855194  
6  0.0420977655  
7  0.0004146452  
8 -0.0990538375  
9 -0.1098573598  
10 0.1843501957  
11 -0.3563067475  
12 -0.1368257055  
13 0.2502257245  
14 -0.0104151195  
15 -0.0059021606  
16 -0.0816369233  
17 -0.0158494842  
18 -0.1145682542  
19 -0.0542372687  
20 0.0227988971  
21 0.0131184146
```

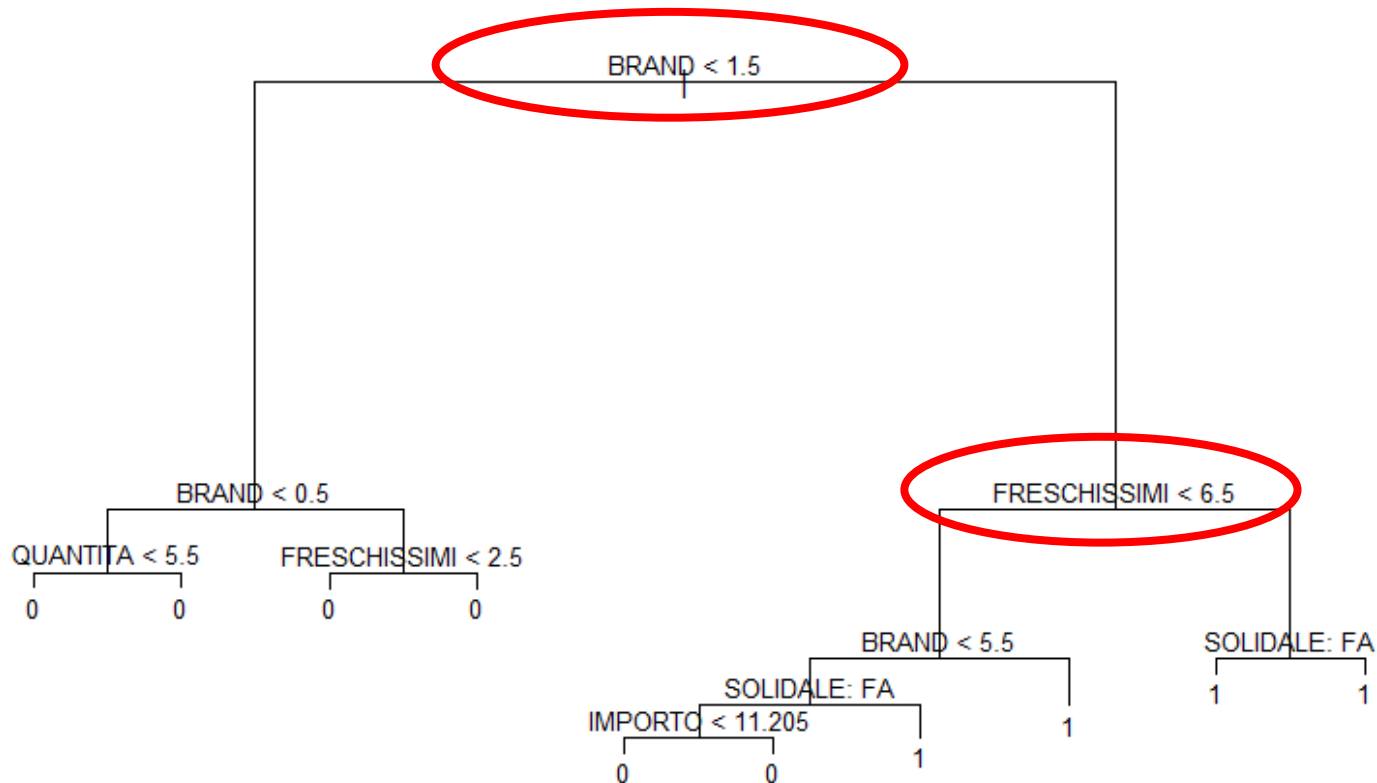
Negozio con minor propensione
media d'acquisto

Negozio con maggior propensione
media d'acquisto

MODELLI DI CLASSIFICAZIONE

Altri tipi di modelli danno indicazioni molto simili

Albero di classificazione

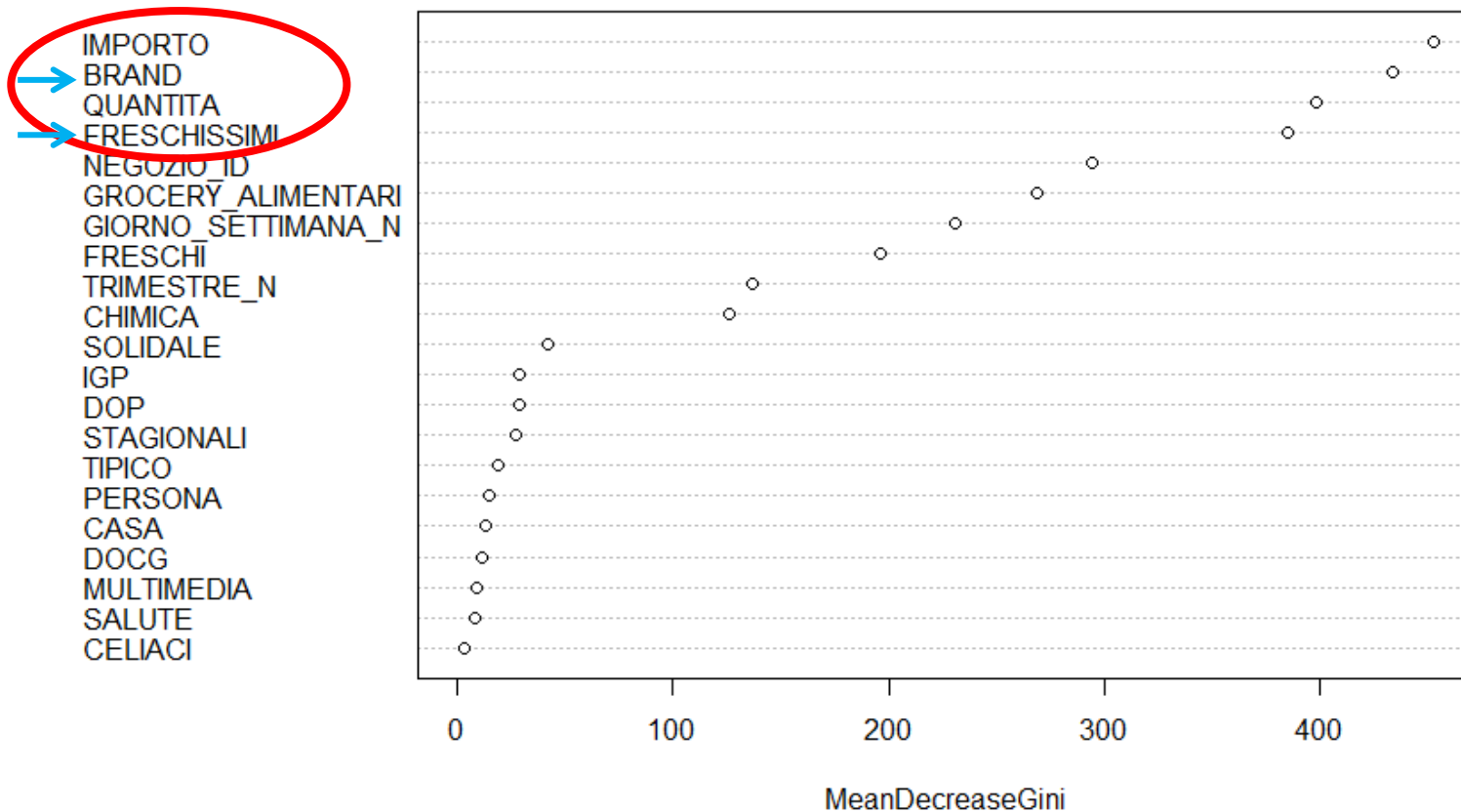


MODELLI DI CLASSIFICAZIONE

Altri tipi di modelli danno indicazioni molto simili

Random Forest

Indice di importanza delle variabili



INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

MODELLI DI PREVISIONE

Con questi modelli vogliamo prevedere la **quantità di prodotti biologici** comprati in ogni acquisto



Escludiamo dal dataset di stima tutte le transazioni che non hanno avuto neanche un acquisto di prodotti biologici



TRAINING SET

ACQUISTI CON
ALMENO UN
PRODOTTO BIO

MODELLI DI PREVISIONE

Il nostro obiettivo è quello di trovare un modello, tra quelli possibili, che sia al tempo stesso interpretabile e abbia il minore EQM calcolato sul test set:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

MODELLI DI PREVISIONE

Il nostro obiettivo è quello di trovare un modello, tra quelli possibili, che sia al tempo stesso interpretabile e abbia il minore EQM calcolato sul test set:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$


Come calcoliamo la previsione?

Per minimizzare l'EQM dovremmo prevedere:

- 0 quando non ci sono stati acquisti BIO
- la quantità quando c'è stato almeno un acquisto BIO

MODELLI DI PREVISIONE

Il nostro obiettivo è quello di trovare un modello, tra quelli possibili, che sia al tempo stesso interpretabile e abbia il minore EQM calcolato sul test set:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$


Come calcoliamo la previsione?

$$\hat{y}_i = \text{acquista} * \hat{y}_{i,prev}$$

dove $\text{acquista} = \begin{cases} 1, & \hat{\mathbb{P}}(BIO) > 0.5 \\ 0, & \text{altrimenti} \end{cases}$

$\hat{y}_{i,prev}$ = previsione data dal modello quantitativo

MODELLI DI PREVISIONE

- SCELTA DEL MODELLO PER $\hat{y}_{i,prev}$

MODELLO	EQM
Regressione lineare (stepwise)	0.31230
Regressione lineare (lasso)	0.23108
Regressione Poisson (stepwise)	0.13255
Modello gerarchico lineare (stepwise)	0.31435
Modello gerarchico Poisson (stepwise)	0.13175
Albero di regressione	0.20978
Rete neurale (12 nodi)	0.35260
Mars	0.25994
Projection pursuit	0.26190
Random forest	0.26394

MODELLI DI PREVISIONE

- SCELTA DEL MODELLO PER $\hat{y}_{i,prev}$

MODELLO	EQM
Regressione lineare (stepwise)	0.31230
Regressione lineare (lasso)	0.23108
Regressione Poisson (stepwise)	0.13255
Modello gerarchico lineare (stepwise)	0.31435
Modello gerarchico Poisson (stepwise)	0.13175
Albero di regressione	0.20978
Rete neurale (12 nodi)	0.35260
Mars	0.25994
Projection pursuit	0.26190
Random forest	0.26394

MODELLI DI PREVISIONE

- **MODELLO FINALE**

Il metodo migliore per prevedere la quantità di prodotti biologici in ogni transazione è in 2 passi:

1. **Modello gerarchico con legame logistico**
2. **Modello gerarchico di Poisson**



In entrambi i passi, si utilizza il modello più "naturale" per la tipologia di dati trattati

MODELLI DI PREVISIONE

Output del modello gerarchico di Poisson con selezione delle variabili basata sull'AIC

Random effects:

Groups	Name	Variance	Std.Dev.
NEGOZIO_ID	(Intercept)	0.004808	0.06934

Number of obs: 3972, groups: NEGOZIO_ID, 21

Fixed effects:

	Estimate	Std. Error	z	value	Pr(> z)	
(Intercept)	0.0125440	0.0358686	0.350	0.72655		
QUANTITA	0.0036683	0.0024094	1.522	0.12789		
IMPORTO	0.0019303	0.0008819	2.189	0.02860	*	
BRAND	0.0587453	0.0043274	13.575	< 2e-16	***	
SOLIDALETRUE	0.3312894	0.0522280	6.343	2.25e-10	***	
DOCGTRUE	0.1843473	0.0798327	2.309	0.02093	*	
FRESCHISSIMI	-0.0117646	0.0047983	-2.452	0.01421	*	
SALUTETTRUE	0.1673285	0.0991376	1.688	0.09144	.	
FRESCHI	-0.0389704	0.0064362	-6.055	1.40e-09	***	
CHIMICA	-0.0357896	0.0086442	-4.140	3.47e-05	***	
GROCERY_ALIMENTARI	-0.0219134	0.0056672	-3.867	0.00011	***	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

MODELLI DI PREVISIONE

Output del modello gerarchico di Poisson con selezione delle variabili basata sull'AIC

- EFFETTI FISSI

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.0125440	0.0358686	0.350	0.72655	
QUANTITA	0.0036683	0.0024094	1.522	0.12789	
IMPORTO	0.0019303	0.0008819	2.189	0.02860	*
BRAND	0.0587453	0.0043274	13.575	< 2e-16	***
SOLIDALETRUE	0.3312894	0.0522280	6.343	2.25e-10	***
DOCGTRUE	0.1843473	0.0798327	2.309	0.02093	*
FRESCHISSIMI	-0.0117646	0.0047983	-2.452	0.01421	*
SALUTETTRUE	0.1673285	0.0991376	1.688	0.09144	.
FRESCHI	-0.0389704	0.0064362	-6.055	1.40e-09	***
CHIMICA	-0.0357896	0.0086442	-4.140	3.47e-05	***
GROCERY_ALIMENTARI	-0.0219134	0.0056672	-3.867	0.00011	***

Per la maggior parte delle variabili i segni dei coefficienti sono **concordi** con quelli del modello di classificazione

MODELLI DI PREVISIONE

Output del modello gerarchico di Poisson con selezione delle variabili basata sull'AIC

- EFFETTI FISSI

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.0125440	0.0358686	0.350	0.72655	
QUANTITA	0.0036683	0.0024094	1.522	0.12789	
IMPORTO	0.0019303	0.0008819	2.189	0.02860	*
BRAND	0.0587453	0.0043274	13.575	< 2e-16	***
SOLIDALETRUE	0.3312894	0.0522280	6.343	2.25e-10	***
DOCTRUE	0.1843473	0.0798327	2.309	0.02093	*
FRESCHISSIMI	-0.0117646	0.0047983	-2.452	0.01421	*
SALUTETURE	0.1673285	0.0991376	1.688	0.09144	.
FRESCHI	-0.0389704	0.0064362	-6.055	1.40e-09	***
CHIMICA	-0.0357896	0.0086442	-4.140	3.47e-05	***
GROCERY_ALIMENTARI	-0.0219134	0.0056672	-3.867	0.00011	***

Per queste variabili i segni dei coefficienti sono **discordi** con quelli del modello di classificazione

MODELLI DI PREVISIONE

Output del modello gerarchico di Poisson con selezione delle variabili basata sull'AIC

- EFFETTI FISSI

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0125440	0.0358686	0.350	0.72655
QUANTITA	0.0036683	0.0024094	1.522	0.12789
IMPORTO	0.0019303	0.0008819	2.189	0.02860 *
BRAND	0.0587453	0.0043274	13.575	< 2e-16 ***
SOLIDALETRUE	0.3312894	0.0522280		
DOCGTRUE	0.1843473	0.0798327		
FRESCHISSIMI	-0.0117646	0.0047983		
SALUTETTRUE	0.1673285	0.0991376	1.688	0.09144 .
FRESCHI	-0.0389704	0.0064362	-6.055	1.40e-09 ***
CHIMICA	-0.0357896	0.0086442	-4.140	3.47e-05 ***
GROCERY_ALIMENTARI	-0.0219134	0.0056672	-3.867	0.00011 ***

Non era presente nel
modello di classificazione

Rispetto al modello di
classificazione, qui **non** sono
presenti le variabili: **tipico**,
stagionali, **persona** e **trimestre**

MODELLI DI PREVISIONE

Output del modello gerarchico di Poisson con selezione delle variabili basata sull'AIC

- EFFETTI CASUALI

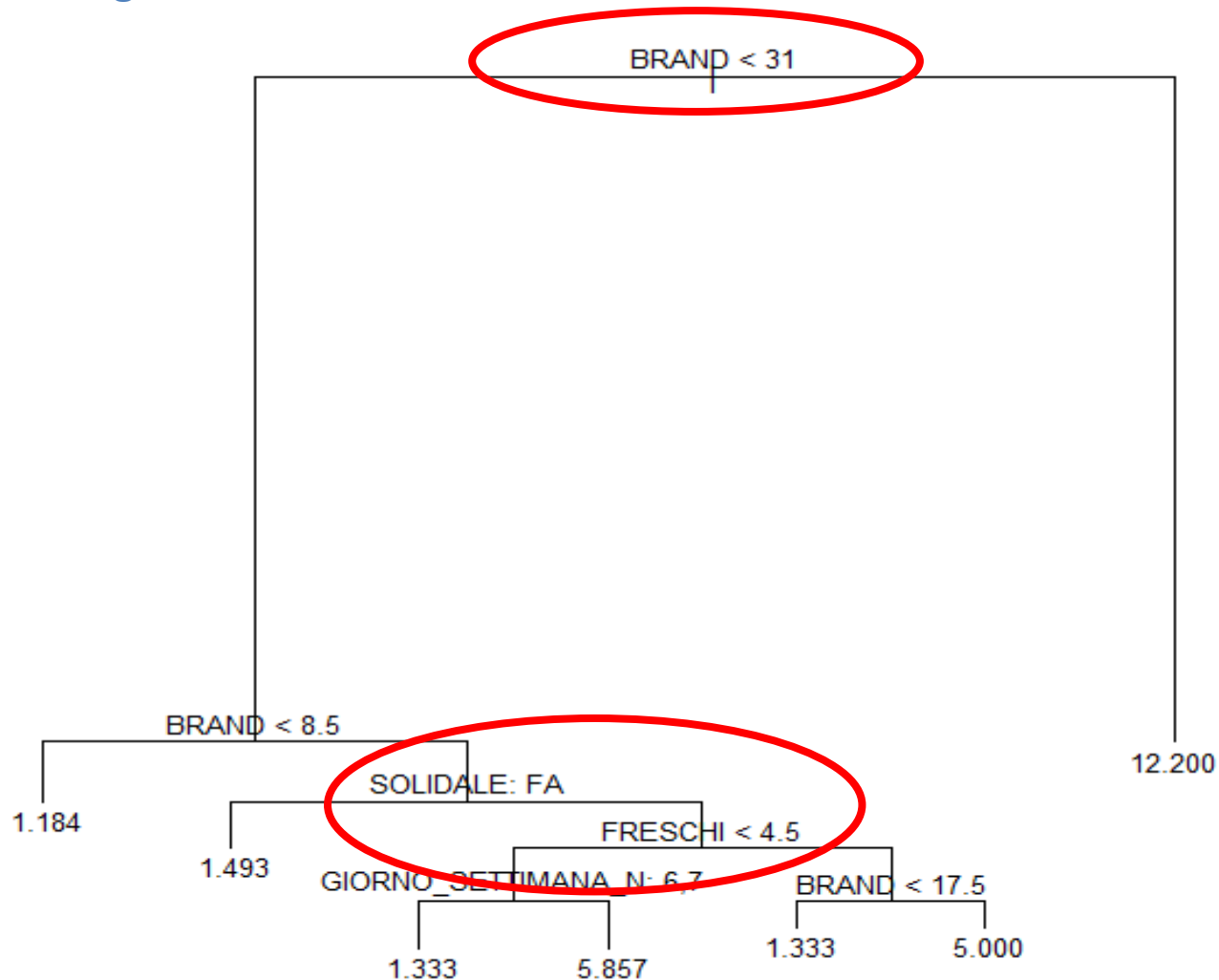
```
$NEGOZIO_ID  
  (Intercept)  
1    0.097354801  
2    0.013798634  
3   -0.050544198  
4    0.020435272  
5    0.088647889  
6   -0.006736863  
7    0.010068361  
8   -0.008121724  
9   -0.003793452  
10   0.017038874  
11  -0.042259253  
12  -0.025763039  
13   0.011160513  
14  -0.006739289  
15  -0.005416380  
16  -0.070366307  
17  -0.026728144  
18  -0.007482538  
19  -0.001578692  
20  -0.007167382  
21   0.004192919
```

Ad eccezione dei negozi evidenziati, i segni degli effetti casuali sono concordi a quelli del modello per la classificazione

MODELLI DI PREVISIONE

Altri tipi di modelli danno indicazioni molto simili

Albero di regressione

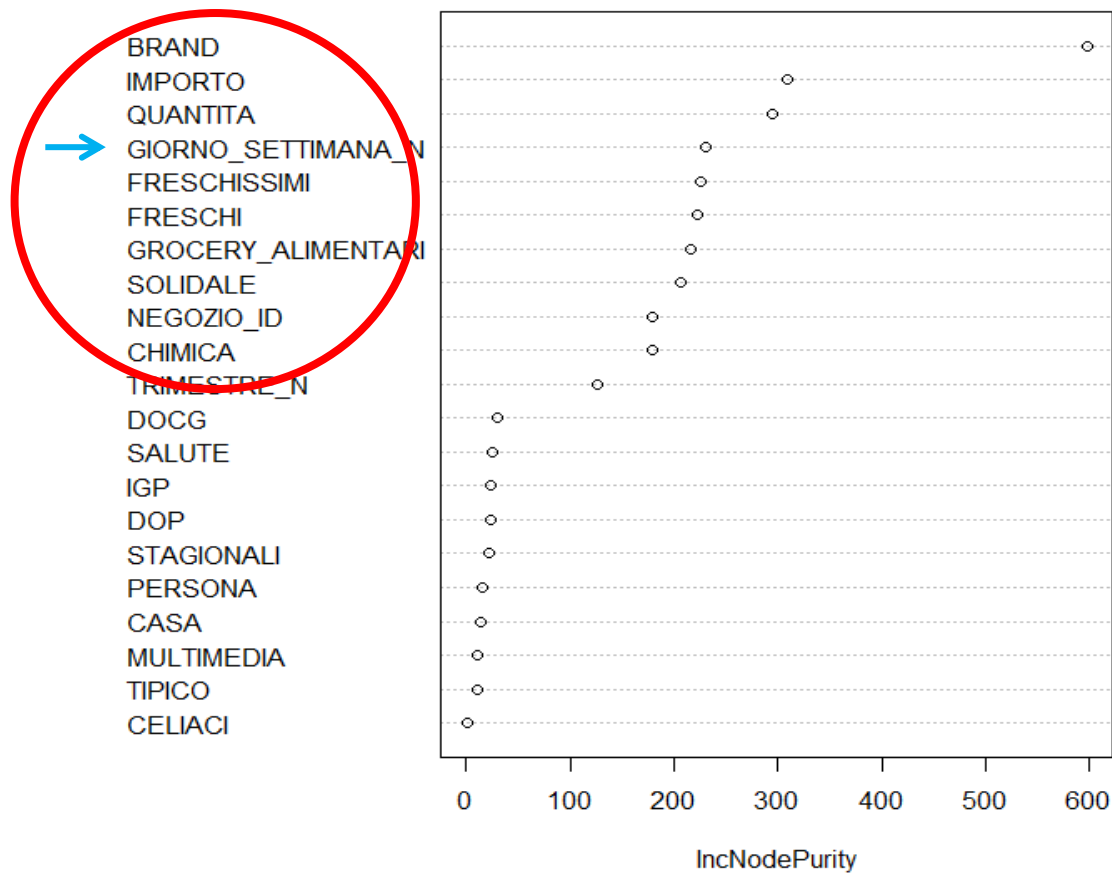


MODELLI DI PREVISIONE

Altri tipi di modelli danno indicazioni molto simili

Random Forest

Indice di importanza delle variabili



INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

REGOLE ASSOCIATIVE (MBA)

FOCUS: PRODOTTI BIOLOGICI

- Come disporre i prodotti biologici negli scaffali?
- Proporre promozioni mirate in base alle abitudini dei consumatori

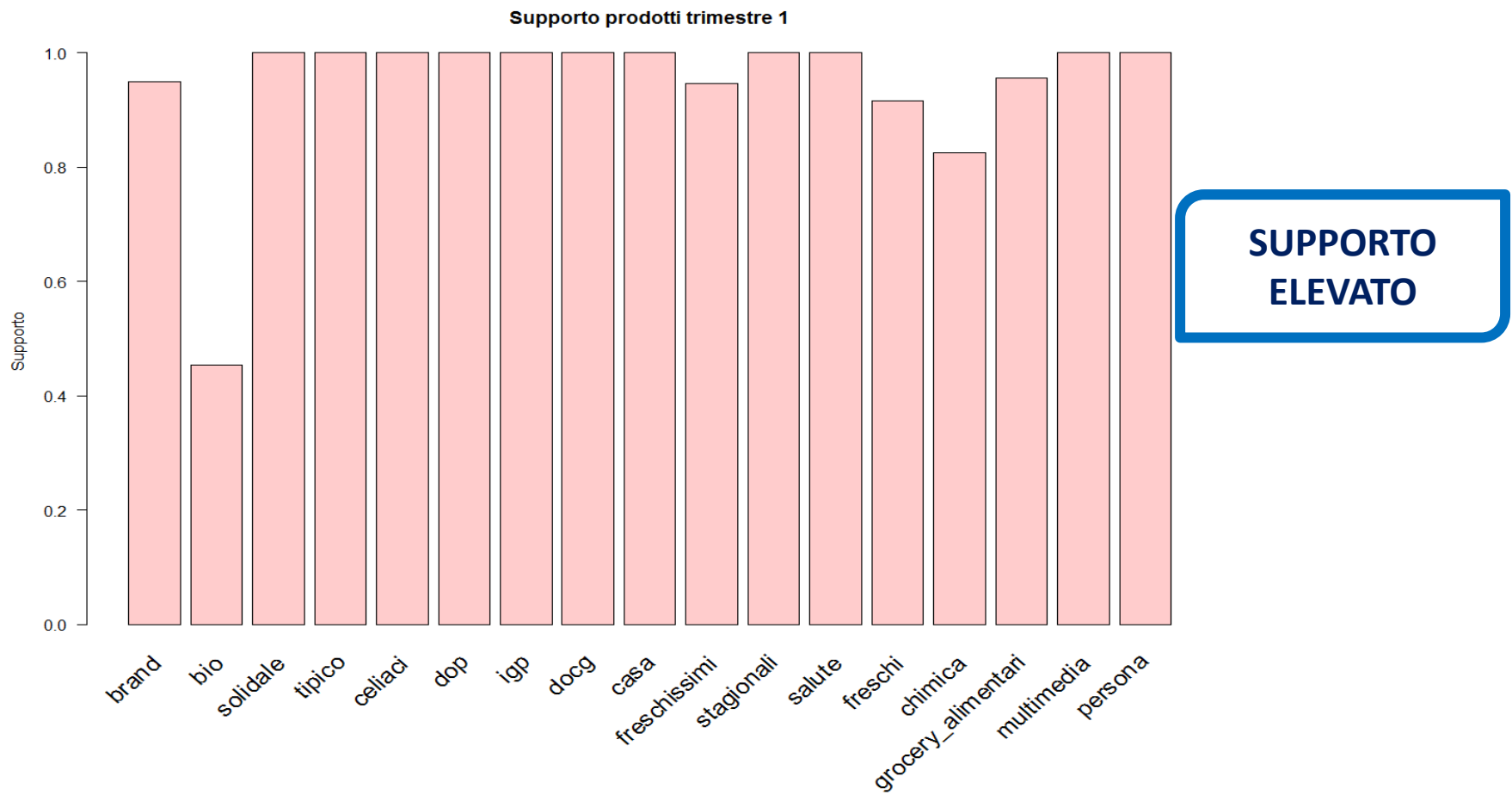


- Creiamo una lista dove per ogni cliente compaiono tutti i prodotti da lui comprati nel periodo temporale considerato

REGOLE ASSOCIATIVE (MBA)

PROBLEMA: Considerando un arco temporale ampio (anno, trimestre, mese) tutti i consumatori acquistano tutte le categorie di prodotti

Esempio con trimestre 1



REGOLE ASSOCIATIVE (MBA)

ASSUNZIONE: Scegliamo tre settimane significative durante l'anno e valutiamo in queste, l'acquisto di prodotti biologici a seconda degli acquisti di altri tipi di prodotti

1. **Settimana 3:** indicazione sulla situazione a inizio anno
2. **Settimana 26:** indicazione sulla situazione a metà anno (Giugno)
3. **Settimana 51:** indicazione sulla situazione a fine anno

REGOLE ASSOCIATIVE (MBA)

- Conseguente: **BIO**
- Lift >1



TANTE REGOLE

Ordinando per confidenza decrescente, si nota che l'antecedente della prima regola ricorre in tutte le successive

Esempio per la settimana 3:

lhs	rhs	support	confidence	lift	count
{grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{salute,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{stagionali,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{casa,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{docg,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{igp,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{dop,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{celiaci,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{tipico,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{solidale,grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111
{grocery_alimentari,persona}	=> {bio}	0.1604046	0.1770335	1.103668	111
{salute,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{stagionali,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{casa,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{docg,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{igp,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{dop,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{celiaci,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111
{tipico,grocery_alimentari,multimedia}	=> {bio}	0.1604046	0.1770335	1.103668	111

REGOLE ASSOCIATIVE (MBA)

Teniamo solo la prima regola

1. All'inizio dell'anno:

lhs	rhs	support	confidence	lift	count
{grocery_alimentari}	=> {bio}	0.1604046	0.1770335	1.103668	111

Antecedente: prodotti alimentari

2. A metà dell'anno:

lhs	rhs	support	confidence	lift	count
{brand,freschissimi,freschi,grocery_alimentari}	=> {bio}	0.09552239	0.121673	1.235165	64

Antecedente: prodotti della catena, prodotti alimentari, di ortofrutta, carne e pesce e altri prodotti freschi

3. A fine anno:

lhs	rhs	support	confidence	lift	count
{freschi}	=> {bio}	0.140647	0.1644737	1.135347	100

Antecedente: prodotti freschi

INDICE

1. Domanda di ricerca
2. Caratteristiche del dataset
3. Sistemazione dataset e analisi esplorative
4. Serie storiche
5. Modelli di classificazione
6. Modelli di previsione
7. Regole associative (MBA)
8. Conclusioni

CONCLUSIONI

- Analizzare la domanda di prodotti biologici nell'arco di un anno

Stagionalità settimanale → consigliato un maggior rifornimento di prodotti per venerdì e soprattutto sabato

- Valutare le relazioni tra i prodotti biologici e le altre variabili

Modelli e regole associative → i prodotti legati a quelli biologici variano durante l'anno, quindi le promozioni e la disposizione negli scaffali vanno riviste frequentemente

La propensione all'acquisto dei prodotti biologici è in funzione di altri prodotti, del tempo e del negozio

Grazie per l'attenzione

