

# La **R**ete di **I**nformazione **C**ontabile **A**gricola

**Nota metodologica**

---

Ufficio di statistica CREA  
C. Cardillo, G. Gabrieli, M. Vassallo

## Sommario

1.1 La normativa comunitaria.....	3
1.2 La Produzione Standard.....	4
1.3 La classificazione delle aziende agricole.....	5
2 Il campo di osservazione.....	7
3 Il Piano di selezione.....	9
3.1 Piano di campionamento stratificato .....	9
3.2 Variabili strategiche e numerosità del campione.....	10
3.4 Liste per l'estrazione delle aziende .....	11
4 Le stime .....	12
4.1 Pesi del campione .....	12
4.2 Calcolo degli errori.....	12
4.3 Calibrazione dei pesi .....	14
5 Gli outlier .....	15
Bibliografia .....	21
Appendice.....	23
Glossario .....	23
Utilizzo dei Pesi.....	25

# 1 Il quadro di riferimento

## 1.1 La normativa comunitaria

Di seguito si riportano i principali regolamenti di riferimento per la RICA, con alcune brevi considerazioni, per un maggiore approfondimento si rimanda ai regolamenti stessi.

**Reg.CE n. 79/65:** istituisce la Rete d'Informazione Contabile Agricola (RICA) per la raccolta dei dati contabili relativi ai redditi ed alle operazioni finanziarie delle aziende agricole dell'Unione Europea. Insieme al **Reg.CE n°1859/82**, esso definisce, insieme al precedente regolamento, il campo di osservazione dell'indagine RICA, rappresentato dalle aziende agricole aventi una dimensione economica maggiore o uguale ad una determinata soglia minima, periodicamente riesaminata sulla base dei risultati forniti dall'indagine sulla struttura delle aziende agricole. Le aziende da rilevare sono scelte in base ad un piano di selezione, che deve assicurare la rappresentatività delle aziende a livello globale ed è redatto dall'organo di collegamento di ogni Stato membro (in Italia è il CREA), sulla base dei dati statistici più recenti. La procedura di stratificazione utilizzata si basa sulla tipologia comunitaria, tenendo in considerazione, dove necessario, anche criteri nazionali aggiuntivi. Il piano deve indicare, inoltre, le procedure per determinare il numero minimo di aziende presenti in ogni strato; quelle per la selezione e la trasmissione delle aziende e dei risultati dell'indagine alla DG-Agri, quelle per un possibile aggiornamento del piano ed il periodo di validità dello stesso.

**Reg. CE n. 1242/2008:** modifica la Decisione della Commissione 85/377/CEE e stabilisce una nuova tipologia comunitaria per le aziende agricole da applicare, a partire dal 2010, alla Rete d'Informazione Contabile Agricola (RICA) ed alle indagini sulla struttura delle aziende agricole. Essa rappresenta una classificazione delle aziende agricole basata su un criterio comune di natura economica e realizzata dagli Stati membri al fine di poter confrontare i risultati economici delle aziende agricole, analizzare la loro situazione e l'impatto che le diverse politiche possono avere su di esse. La nuova tipologia tiene in considerazione tre aspetti delle aziende agricole: la localizzazione geografica, l'Orientamento Tecnico Economico (OTE) e la classe di Produzione Standard (o Standard Output) (PS). Inoltre, il regolamento prevede un elemento di classificazione aggiuntivo rappresentato dalle Altre Attività Lucrative (AAL) direttamente collegate all'azienda.

**Reg. CE n. 1217/2009:** abroga definitivamente il Reg. n. 79/65 e definisce quale campo di osservazione dell'indagine tutte le aziende di dimensione economica superiore o uguale ad un minimo espresso in euro, pari ad uno dei limiti inferiori delle classi di dimensione economica definite dalla tipologia comunitaria. Entrato in vigore nel gennaio 2010, esso stabilisce le regole per la raccolta di dati contabili da consegnare alla Commissione Europea, essi formano la cosiddetta "scheda aziendale" e sono identici per tutti i paesi dell'UE, indipendentemente dalle loro pratiche contabili e dal modo in cui organizzano il processo di raccolta. I dati sono soggetti a rigide norme di riservatezza e possono essere utilizzati solo per soddisfare le esigenze della Politica Agricola Comune. La Commissione è assistita dal Comitato per la Rete di Informazione Contabile Agricola, composto da rappresentanti di tutti i paesi dell'UE.

**Reg. (UE) n. 1318/2013:** applicato dal 1 ° gennaio 2014, modifica il regolamento (CE) n. 1217/2009 e offre alla Commissione il potere di adottare atti delegati e di esecuzione che possono riferirsi alla definizione delle unità territoriali di riferimento per la RICA, alle regole per la determinazione della dimensione economica delle aziende che partecipano al campione e per il

piano di selezione delle aziende stesse, alla principale classificazione tipologica delle aziende agricole dell'UE, ai principali gruppi di dati contabili da raccogliere e alle regole generali per la raccolta dei dati. In virtù di questo potere, la Commissione, attraverso il **regolamento delegato (UE) n. 1198/2014**, ha stabilito la corrispondenza tra gli orientamenti tecnici economici o l'elenco dei principali dati contabili da raccogliere.

**Reg. (UE) 2015/220 della Commissione:** stabilisce norme per l'applicazione del regolamento (CE) n. 1217/2009 del Consiglio, che includono la descrizione dettagliata dei dati contabili da raccogliere in ciascuna scheda dell'azienda agricola e le soglie economiche applicate in ciascun paese. Su tale regolamento si basa anche la definizione della tipologia di azienda agricola, la corrispondenza tra le rubriche dell'indagine sulla struttura delle aziende agricole (SPA) e della RICA e i codici di raggruppamento di determinate caratteristiche, il calcolo della dimensione economica delle aziende e delle classi di dimensione economica, la definizione e il calcolo dei coefficienti di PS, la definizione delle AAL direttamente correlate all'azienda e la stima del loro peso sul totale della produzione aziendale.

## 1.2 La Produzione Standard

Fino all'entrata in vigore del Reg. CE 1242/2008, il criterio economico utilizzato per la classificazione tipologica delle aziende agricole europee era il Reddito Lordo Standard (RLS), calcolato a livello regionale, esso rappresentava il valore della produzione lorda di una attività agricola, al netto di alcuni costi specifici direttamente imputabili alla stessa, e corrispondeva al reddito medio teorico ottenibile da quella attività in una determinata regione.

La graduale eliminazione dei sussidi legati alla produzione dall'ambito normativo della PAC ha però portato ad una revisione dei coefficienti considerati per la classificazione delle aziende agricole, al fine di evitare che per alcune produzioni si avessero risultati negativi per i valori dei RLS. Si è quindi proposta la sostituzione degli RLS con le Produzioni Standard (PS) o Standard output, che prendono in considerazione solo i risultati relativi alla produzione lorda, ottenuta moltiplicando le rese per i prezzi e non si considerano invece i costi variabili ed i sussidi ricevuti dalle aziende. L'utilizzo della PS al posto del Reddito Lordo Standard (RLS) ha rappresentato dunque la reale innovazione rispetto alla normativa precedente, infatti l'OTE e la classe economica dell'azienda devono essere determinate sulla base di un criterio economico che rimane sempre positivo. Per questo la Commissione ha considerato appropriato l'uso della PS calcolata per prodotto e basata su valori medi di un periodo di riferimento di 5 anni e regolarmente aggiornati tenendo in considerazione gli andamenti economici. Per il calcolo della PS per l'anno N, il periodo di riferimento comprende cinque anni consecutivi dall'anno N-5 all'anno N-1 e l'anno di riferimento è N-3. Ad esempio, per la PS 2007 i valori medi utilizzati fanno riferimento al periodo 2005 – 2009, per la PS 2010 i valori medi utilizzati sono quelli relativi agli anni 2008 – 2012.

Le informazioni utilizzate per il calcolo delle PS sono derivate dalle statistiche congiunturali ISTAT, per le informazioni su superfici e produzioni, dalle statistiche sulla macellazione ISTAT, per quanto concerne le produzioni animali, dalla contabilità nazionale ISTAT, per produzioni e valori. Inoltre, si usa la banca dati RICA del CREA, per una verifica delle informazioni strutturali ricavate dalle altre fonti, per le informazioni relative alle attività non presenti nelle altre fonti e per quanto riguarda gli aspetti economici.

Per ogni azienda si definisce la PS totale aziendale, moltiplicando prima gli ettari delle colture ed i capi di bestiame presenti nell'azienda per i rispettivi coefficienti di PS unitari, relativi alla regione di appartenenza e, successivamente, sommando tutti i valori di PS ottenuti per ogni singola attività produttiva agricola dell'azienda. Questo calcolo consente quindi di definire la dimensione economica delle aziende misurata direttamente in euro, e da questa è possibile costruire delle classi di dimensione economica, che costituiscono il criterio di base per la classificazione della tipologia comunitaria, insieme all'Orientamento tecnico-economico (OTE). Quest'ultimo definisce l'indirizzo produttivo dell'azienda ed il suo grado di specializzazione, sulla base dell'incidenza percentuale, in termini di PS, delle diverse attività produttive sul totale del reddito aziendale.

Lo schema seguente riporta, per ogni annualità del campione RICA, il criterio economico utilizzato ed il rispettivo periodo di riferimento.

**Figura 1: Sintesi dei coefficienti e anno contabile del campione RICA**

REDDITO LORDO STANDARD	CAMPIONE	PRODUZIONE STANDARD	CAMPIONE
2007	2008	2010	2010
	2009		2011
			2013
		2013	
		2014	
		2015	
		2016	2016
			2017
			2018

### 1.3 La classificazione delle aziende agricole

L'attività di classificazione della tipologia comunitaria si basa, come già detto, su quanto stabilito nella normativa comunitaria di riferimento e nelle linee guida fornite dalla Commissione Europea. Per la definizione del campione RICA, il Regolamento impone l'utilizzo delle informazioni contenute nell'ultima indagine disponibile sulla struttura delle aziende agricole, in particolare, le informazioni riguardanti gli ettari coltivati per le singole colture, ed i capi allevati per ogni specie animale.

La procedura di classificazione è stata realizzata dal CREA tramite lo sviluppo di un'applicazione web, denominata "ClassCE"<sup>1</sup>, nella quale è stato implementato l'algoritmo di classificazione delle aziende agricole, seguendo le indicazioni contenute nel regolamento CE n. 1242/2008 e successivi aggiornamenti. La procedura utilizza i valori delle PS calcolati per ciascuna regione e i dati strutturali delle aziende, superfici e capi di allevamento, determinando, per ciascuna azienda rilevata, sia il suo Orientamento Tecnico-Economico (OTE) che la dimensione economica, attraverso la PS aziendale, pari alla somma del valore di PS delle singole attività agricole in essa realizzate.

Di seguito sono riportati alcuni esempi di classificazione delle aziende.

<sup>1</sup> <https://rica.crea.gov.it/classce/>

Esempio<sup>2</sup>:

Classificazione di un'azienda agricola che presenta le seguenti attività agricole:

Attività agricola	Rubrica	Dimensione	Unità	SO	SO totale
Frumento tenero e spelta	D01	7,53	ha	1.110	8.360
Mais	D06	15,13	ha	1.449	21.917
Orticole - all'aperto - in pieno campo	D14A	2,78	ha	13.293	36.955
Orticole - all'aperto - in orto industriale	D14B	0,32	ha	16.314	5.221
Prati avvicendati (medica, sulla, trifoglio, lupinella, ecc.)	D18B	0,37	ha	842	312
Prati permanenti e pascoli	F01	5,50	ha	640	3.522
Bovini maschi e femmine meno di 1 anno	J02	2,00	Capi	1.036	2.071
Bovine femmine da 1 a meno di 2 anni	J04	2,00	Capi	656	1.311
Bovini maschi d 2 anni e più	J05	1,00	Capi	325	325
Altre vacche (vacche nutrici, vacche da riforma)	J08	28,00	Capi	795	22.247

Codice attività produttive e soglie (in €)

<b>p45</b>	C_02_01 + C_02_03 + C_02_05 + C_02_06	3.382
<b>p46</b>	P45 + C_02_02 + C_02_04 + C_02_99	25.955
<b>GL</b>	C_01 + P46 + C_03_01_01 + C_03_01_99 + C_03_02_01 + C_03_02_99	25.955

<b>GL&gt;0</b>		
<b>FCP1</b>		0
<b>FCP4</b>	B_01_05 + B_01_09 + B_03_01 + B_03_02	3.834

<b>P1</b>		67.232
<b>P2</b>		5.221
<b>P3</b>		0
<b>P4</b>	GL + FCP4	27.717
<b>P5</b>		0

<b>SO</b>		100.170
<b>2/3 SO</b>		66.780

Definizione Orientamento tecnico economico (OTE)

<b>Polo 1</b>	P1 > (2/3 SO)	<b>SI</b>
<b>Ote 15</b>	(P15 + P16 + B_01_02) > (2/3 SO)	<b>NO</b>
<b>Ote 16</b>	(P15 + P16 + B_01_02) <= (2/3 SO)	<b>SI</b>
<b>Ote 161</b>	P17 > (2/3 SO)	<b>NO</b>
<b>Ote 162</b>	(P15 + P16 + B_01_02) > (1/3 SO) e P17 > (1/3 SO)	<b>NO</b>
<b>Ote 163</b>	B_01_07_01_01 > (2/3 SO)	<b>NO</b>
<b>Ote 164</b>	B_01_06_01 > (2/3 SO)	<b>NO</b>
<b>Ote 165</b>	B_01_06_03 > (2/3 SO)	<b>NO</b>
<b>Ote 166</b>		<b>SI</b>

Calcolo Standard output aziendale (SO totale)

<b>P1 + P2 + P3 + P4 + P5</b>	<b>100.170</b>
-------------------------------	----------------

<sup>2</sup> Per approfondimento:

[https://rica.crea.gov.it/public/download/normativa/Reg%20CE%201242\\_08.pdf](https://rica.crea.gov.it/public/download/normativa/Reg%20CE%201242_08.pdf)

## 2 Il campo di osservazione

Secondo quanto stabilito nel Reg. CE n. 1242/2008 e successivi aggiornamenti, l'universo di riferimento per l'estrazione del campione RICA è costituito dall'ultima indagine sulla struttura delle aziende agricole disponibile, al fine di assicurare la rappresentatività del campo di osservazione.

In Italia le indagini sulla struttura delle aziende agricole sono condotte dall'Istituto Nazionale di Statistica (ISTAT) in due forme:

- 1) Censimento dell'agricoltura italiana, indagine di tipo censuario con cadenza decennale<sup>3</sup>;
- 2) Indagine sulla Struttura e Produzione delle Aziende agricole (SPA), indagine campionaria con cadenza triennale, realizzata nel periodo intercensuario.

Entrambe le indagini rispondono all'attuazione di diverse normative comunitarie, di cui la più recente è il Reg. N. 1166/2008 del Parlamento Europeo e del Consiglio del 19 novembre 2008.

La disponibilità di una nuova indagine strutturale consente di svolgere un'analisi sulle principali caratteristiche del settore agricolo e di valutare l'esistenza di eventuali cambiamenti significativi nel contesto di riferimento e, qualora ciò si verifici, si procede alla definizione di un nuovo piano di campionamento. La normativa prevede inoltre, che sia sempre garantita una percentuale di copertura per alcune variabili definite<sup>4</sup>, quali la Superficie Agricola Utilizzata (SAU), la Produzione Standard (PS) o Standard Output (SO) e le Unità di Bovino Adulto (UBA).

A partire dal campione RICA 2014, l'universo di riferimento è costituito dalle aziende rilevate dal Censimento agricolo 2010 ed è stata fissata una soglia minima di ingresso pari a 8.000 euro di PS, che garantisce tutte le coperture richieste dal regolamento comunitario. Infatti, il campione rappresenta circa il 37% delle aziende agricole italiane, l'85% della SAU, il 95% dello Standard Output o Produzione Standard totale e il 99% delle UBA (Figura x). Successivamente, attraverso le informazioni raccolte dalle indagini di struttura SPA 2013 e 2016, è stata realizzata un'analisi del contesto dell'agricoltura italiana, ma i cui risultati non hanno suggerito la necessità di procedere alla definizione di un nuovo piano di campionamento.

---

<sup>3</sup> Il prossimo Censimento generale dell'agricoltura sarà svolto nel 2020, al quale parteciperanno tutte le aziende agricole italiane, con l'obiettivo di descrivere un quadro informativo statistico sulla struttura dell'intero sistema agricolo.

<sup>4</sup> Per maggiori informazioni si può fare riferimento ai documenti RI/CC 1519 oppure RI/CC 1524 e RI/CC 1526 della Commissione Europea.

**Figura 2 - Copertura del campione RICA, a partire dall'anno contabile 2014**

Classe di SO	Limite inferiore (in €)	Limite superiore (in €)	Numero di aziende	Inversa cumulata (%)	Superficie Agricola Utilizzata	Inversa cumulata (%)	Standard Output	Inversa cumulata (%)	Unità di Bovino Adulte (UBA)	Inversa cumulata (%)
1		2.000	518.385	100,0	578.843	100,0	512.254.676	100,0	10.062	100,0
2	2.000	4.000	263.773	68,0	523.959	95,5	765.093.762	99,0	24.266	99,9
3	4.000	8.000	236.338	51,7	796.708	91,4	1.354.603.954	97,4	62.835	99,7
4	8.000	15.000	177.023	37,2	981.797	85,2	1.949.711.788	94,7	129.876	99,0
5	15.000	25.000	119.505	26,2	1.018.546	77,6	2.320.878.795	90,7	204.087	97,7
6	25.000	50.000	128.590	18,9	1.718.703	69,7	4.557.314.122	86,0	517.790	95,6
7	50.000	100.000	88.655	10,9	2.058.613	56,3	6.227.506.358	76,8	896.858	90,4
8	100.000	250.000	59.436	5,5	2.557.983	40,3	9.054.610.179	64,2	1.547.441	81,4
9	250.000	500.000	17.410	1,8	1.214.261	20,4	5.988.971.178	45,9	1.204.871	65,7
10	500.000	750.000	5.062	0,7	491.858	10,9	3.074.596.760	33,8	739.032	53,6
11	750.000	1.000.000	2.125	0,4	244.078	7,1	1.834.088.918	27,6	465.392	46,1
12	1.000.000	1.500.000	2.009	0,3	241.238	5,2	2.429.247.995	23,9	709.868	41,4
13	1.500.000	3.000.000	1.679	0,2	259.774	3,3	3.426.054.114	19,0	1.062.737	34,2
14	3.000.000		894	0,1	169.687	1,3	5.965.397.110	12,1	2.325.681	23,5
<b>Totale</b>			1.620.884		12.856.048		49.460.329.710		9.900.796	

## 3 Il Piano di selezione

### 3.1 Piano di campionamento stratificato

Per l'indagine RICA si utilizza un piano di campionamento stratificato, secondo il quale la popolazione delle aziende agricole italiane viene suddivisa in H strati, omogenei rispetto ad una o più variabili di stratificazione che, in questo caso, sono utilizzate quali variabili ausiliarie.

Utilizzare un piano di campionamento stratificato consiste nel suddividere la popolazione di riferimento in sottopopolazioni, sulla base delle informazioni ausiliarie. Ciò consente di aumentare l'omogeneità delle unità statistiche all'interno degli strati e di conseguenza, ottenere stime più efficienti di quelle che si possono avere adottando un campione casuale semplice (Nicolini et al., 2013).

Il primo aspetto che viene valutato riguarda, quindi, le variabili di stratificazione, regione, OTE e dimensione economica, che garantiscono la rappresentatività della popolazione da parte del campione e, di conseguenza, i domini di studio sui quali viene fissata la precisione delle stime delle variabili strategiche (Cardillo et al. 2010).

In generale si può procedere all'aggregazione delle variabili di stratificazione a seconda di quanto si ritenga più opportuno, ad esempio gli OTE si possono utilizzare come poli principali oppure raggruppando le specializzazioni, oppure le classi di dimensione economica possono essere raggruppate in un numero minore rispetto alle 14 previste dal regolamento di riferimento, sempre però rimanendo nell'ambito di quanto fissato dalle linee guida della Commissione. Di seguito si riportano come esempio le modalità di campionamento messe in atto per il campione 2014, che al momento è quello più aggiornato.

Per il campione RICA 2014, le variabili di stratificazione sono state:

- 19 Regioni italiane e 2 Province autonome (Trento e Bolzano);
- Gruppi di OTE<sup>5</sup> particolari, che tengono conto della rilevanza che ogni coltura e/o allevamento riveste in ciascuna regione;
- 6 classi di dimensione economica, espressa in termini di Produzione Standard, comuni a tutte le regioni.

La forte eterogeneità che caratterizza l'agricoltura italiana a livello regionale, come ad esempio la presenza di aziende molto piccole in alcune regioni, un'alta specializzazione di alcune tipologie di aziende o l'importanza di determinati OTE, ha reso necessario definire un piano di campionamento specifico per ciascuna regione e provincia autonoma. Tale aspetto, ha consentito all'indagine sia di rispondere a nuovi bisogni informativi come la stima per piccole aree e, quindi, l'analisi dei dati relativi alle politiche dello sviluppo rurale.

---

<sup>5</sup> L'OTE di un'azienda rappresenta il sistema di produzione o l'attività prevalente dell'azienda ed è determinato, come si è detto, dal contributo relativo delle diverse attività alla PS totale dell'azienda, in particolare la classificazione proposta dalla normativa comunitaria prevede tre diversi livelli di OTE:

- 9 OTE generali, incluso quello per le aziende non classificate;
- 21 OTE principali;
- 62 OTE particolari.

Come già detto, in ogni regione i gruppi di OTE sono il risultato di un'aggregazione di OTE particolari, dipendente dalla copertura della PS generata dalla singola tipologia sul totale della PS della regione. Il criterio usato per l'aggregazione è stato, quindi, quello della copertura di almeno il 5% della PS regionale. Gli OTE al di sotto di questa soglia sono stati aggregati per tipologie quanto più possibile omogenee.

Per quanto riguarda la dimensione economica, sono state considerate 6 classi di dimensione economica, comuni a tutte le regioni e province autonome:

- 1) 8.000 - 25.000 euro di PS;
- 2) 25.000 - 50.000 euro di PS;
- 3) 50.000 - 100.000 euro di PS;
- 4) 100.000 - 500.000 euro di PS;
- 5) 500.000 – 1.000.000 euro di PS;
- 6) >1.000.000 euro di PS.

Tali classi, sono il risultato di un'aggregazione delle 14 classi elementari previste dal regolamento.

**Figura 3 Classi di dimensione economica adottate per il campione RICA**

CLASSI DI PRODUZIONE STANDARD	RAGGRUPPAMENTO IN CLASSI	DESCRIZIONE (euro)
I		meno di 2.000
II		2.000 – 4.000
III		4.000 – 8.000
IV	Classe 1	8.000 – 15.000
V		15.000 – 25.000
VI	Classe 2	25.000 – 50.000
VII	Classe 3	50.000 – 100.000
VIII	Classe 4	100.000 - 250.000
IX		250.000 – 500.000
X	Classe 5	500.000 – 750.000
XI		750.000 – 1.000.000
XII	Classe 6	1.000.000 – 1.500.000
XIII		1.500.000 – 3.000.000
XIV		maggiore o uguale 3.000.000

## 3.2 Variabili strategiche e numerosità del campione

Le variabili strategiche sono quelle per le quali è stato fissato un livello di precisione relativa per le stime nei domini di studio pianificati, espresso in termini di coefficiente di variazione CV<sup>6</sup>.

Per il campione RICA, la scelta delle variabili strategiche viene fatta sulla base di quelle disponibili nelle indagini strutturali, ovvero di una variabile economica, la Produzione Standard e di due variabili strutturali quali la SAU e le UBA. Tali coefficienti, fissati a livello nazionale e regionale, costituiscono i domini di studio pianificati del campione RICA (Cardillo et al. 2010).

<sup>6</sup> Il coefficiente di variazione (CV) è definito come:

$$CV = \frac{\sqrt{\widehat{Var}(\hat{Y})}}{\hat{Y}}$$

dove  $\sqrt{\widehat{Var}(\hat{Y})}$  è la deviazione standard della stima  $\hat{Y}$ .

**Precisione in termini di CV del campione RICA**

<b>DOMINIO</b>	<b>CV (%)</b>
Nazionale - Italia	3,0
Regionale	5,0

Una volta definiti i coefficienti di variazione, si procede all'estrazione della numerosità campionaria per ciascun singolo strato del campione, attraverso un'estensione del metodo di allocazione di Neyman al caso di più variabili e adottando come metodo di risoluzione una generalizzazione della proposta di Bethel (1989). L'algoritmo, generalizzato al caso multivariato consente il calcolo dell'allocazione campionaria ottima per disegni campionari stratificati. Di conseguenza, la dimensione delle unità campionarie nei diversi strati avviene tenendo in considerazione sia i vincoli imposti sulla precisione delle variabili strategiche in termini di variabilità, che la numerosità della popolazione nello strato. In questo modo si realizza un'allocazione ottima multivariata e multi-dominio.

### **3.4 Liste per l'estrazione delle aziende**

Uno degli aspetti fondamentali per la formulazione di un piano di campionamento è rappresentato dalla disponibilità di una lista con contenuti informativi corretti ed aggiornati, da cui poter estrarre le unità statistiche da rilevare.

Fino ad oggi, l'unica base informativa disponibile ed esaustiva per rappresentare la popolazione delle aziende agricole italiane è stata il Censimento Generale dell'agricoltura, che contiene per ciascuna azienda sia informazioni anagrafiche che strutturali.

Tuttavia, come già detto in precedenza, il censimento è svolto con cadenza decennale, quindi pur venendo aggiornato nel corso del decennio con altre indagini campionarie in agricoltura l'agricoltura, già dopo qualche anno dalla sua uscita risulta spesso obsoleto. Infatti, le aziende nel tempo possono aver cambiato denominazione, conduzione o, semplicemente aver cessato l'attività. Questo comporta maggiori difficoltà durante la fase di rilevazione, per la loro individuazione e contatto da parte dei rilevatori. Inoltre, il campione RICA ha una soglia minima da rispettare in termini di numero di aziende rilevate da consegnare alla DGAgri, pari all'80% della numerosità teorica. Tale vincolo, spesso, obbliga a far ricorso ad un'ulteriore estrazione di nominativi di aziende casuali oppure ad aziende rilevate su base volontaria, spesso già selezionate in passate indagini RICA (Rondinelli, 2011). Inoltre, risulta abbastanza intuitivo che, per attenuare eventuali distorsioni, le aziende volontarie utilizzate in sostituzione di aziende casuali non rispondenti devono necessariamente avere caratteristiche strutturali ed economiche simili a quelle delle aziende casuali sostituite.

Per questi motivi il campione RICA ha assunto sempre di più la struttura di un panel con rotazione annua del 25% delle unità statistiche. Tale decisione, non solo consente di facilitare il raggiungimento della soglia minima, ma anche di svolgere analisi longitudinali più accurate.

La disponibilità del Farm Register, il registro delle aziende agricole implementato da ISTAT e costruito sulla base di informazioni provenienti da varie fonti amministrative ed aggiornato annualmente, costituirà la nuova base informativa per la costruzione della lista della popolazione delle aziende agricole italiane. Esso consentirà di disporre di un universo di aziende sempre aggiornato con cui ricostruire la popolazione e selezionare le unità statistiche del campione.

# 4 Le stime

## 4.1 Pesi del campione<sup>7</sup>

La stima delle variabili strategiche di un'indagine campionaria si ottiene utilizzando dei coefficienti moltiplicativi, i "pesi", quali fattori moltiplicativi della variabile campionaria in oggetto. Per un campione stratificato in H strati ( $H=1, \dots, H$ ), a ciascuna unità campionaria viene associato un peso definito "di base":

- 1) Per ogni strato si calcola il peso di base  $w_h = \frac{N_h}{n_h}$ , dove  $N_h$  è la numerosità dell'universo di riferimento delle aziende nello strato e  $n_h$  è la numerosità delle aziende del campione estratto nello stesso strato;
- 2) Ad ogni azienda campionaria  $j$  appartenente allo strato  $h$ , viene associato il peso base  $w_h$ .

Tutte le aziende  $j$  appartenenti allo stesso strato  $h$  presentano, quindi, lo stesso peso "di base".

Una volta conclusa la rilevazione, i pesi "di base" vengono corretti per il fenomeno della mancata risposta, quindi il peso  $w_h$  viene riproporzionato con l'effettiva numerosità delle aziende rilevate ottenendo, quindi, i pesi "finali" ( $w_h'$ ). Questi pesi finali consentono il riporto all'universo a livello regionale e nazionale.

Indicando con:

- $H$  : numero di strati in cui è suddivisa la popolazione ( $h=1, \dots, H$ );
- $n_h'$  : numerosità campionaria dello strato  $h$ -esimo;
- $w_h'$  : peso "finale" dell'azienda RICA  $j$ -esima;
- $y_j$  : valore della variabile d'interesse dell'azienda RICA  $j$ -esima.

La stima del totale di popolazione della generica variabile  $Y$  è data da:

$$T(\hat{Y}) = \sum_{h=1}^H \sum_{j=1}^{n_h} w_h' y_{hj} \quad [1]$$

dove  $w_h' = \frac{N_h}{n_h'}$

Analogamente, la stima della media di popolazione sarà data dal rapporto tra la stima del totale e la somma dei pesi finali campionari:

$$m(\hat{Y}) = \frac{\sum_{h=1}^H \sum_{j=1}^{n_h} w_h' y_{hj}}{\sum_{h=1}^H w_h'} \quad [2]$$

## 4.2 Calcolo degli errori<sup>8</sup>

Quando si vuole estendere la stima campionaria di una variabile d'interesse all'universo, bisogna considerare quali sono le variabili strategiche definite nel piano di campionamento. Infatti, la numerosità campionaria e la distribuzione delle unità negli strati viene calcolata per contenere la variabilità delle stime di queste variabili. Per questo motivo, i pesi garantiscono una stima affidabile

<sup>7</sup> Per approfondimenti sulle stime per un campionamento stratificato si consiglia il lavoro di Daniela Cocchi (2011).

<sup>8</sup> Per approfondimenti sul calcolo degli errori campionari si consiglia il lavoro di Mario Mazzocchi (2008).

solamente per le variabili strategiche nei domini di studio pianificati per il campione. Proprio per la loro natura, i pesi, quali fattori moltiplicativi, consentono, con un certo grado di affidabilità che va tuttavia valutato ogni volta, di estendere all'universo anche alcune variabili fortemente correlate<sup>9</sup> a quelle strategiche.

Viceversa, l'utilizzo dei pesi per variabili non correlate a quelle strategiche, come ad esempio dati specifici su allevamenti e coltivazioni, risulta fortemente sconsigliato.

Come già detto, i dati RICA consentono di fornire valide informazioni degli aggregati di popolazione. Per questo motivo, l'utilizzo dei pesi per estendere i dati campionari ai corrispondenti valori di popolazione, da origine a delle stime e non a valori "veri" di popolazione.

Purtroppo, non essendo a conoscenza del valore reale di una variabile per la popolazione di riferimento, non è possibile calcolare puntualmente l'errore commesso con la nostra stima. Tuttavia, dal momento che il campione è casuale, possiamo stimare l'entità di tale errore.

Generalmente, per stimare l'errore di uno stimatore in un'indagine statistica, vengono utilizzate due grandezze:

- 1) L'errore di campionamento relativo (chiamato anche coefficiente di variazione, CV, già precedentemente accennato nel paragrafo 3.2), o anche precisione relativa, che essendo un indice relativo e adimensionale (privo di unità di misura), consente un confronto degli errori sia tra variabili che per domini diversi. In formula:

$$\hat{\varepsilon}(\hat{Y}) = \frac{\sqrt{\widehat{Var}(\hat{Y})}}{\hat{Y}} \quad [3]$$

Dove  $\widehat{Var}(\hat{Y})$  e  $\hat{Y}$  sono, rispettivamente, la stima della varianza e la stima della generica variabile  $Y$ . Nel caso di una stima del totale, l'errore di campionamento relativo sarà dato dal rapporto tra la stima della varianza dello stimatore del totale<sup>10</sup> e la stima del totale;

- 2) La stima dell'errore di campionamento assoluto, o precisione:

$$\hat{\sigma}(\hat{Y}) = \sqrt{\widehat{Var}(\hat{Y})} \quad [4]$$

generalmente utilizzato per calcolare gli intervalli di confidenza

$$\hat{Y} - z_{\frac{\alpha}{2}} \hat{\sigma}(\hat{Y}) \leq Y \leq \hat{Y} + z_{\frac{\alpha}{2}} \hat{\sigma}(\hat{Y}) \quad [5]$$

Poiché la variabile stima del totale  $\hat{Y}$  si distribuisce ancora come una variabile aleatoria normale è possibile calcolare le soglie dell'intervallo di confidenza [5] per, ad esempio, un livello fiduciario

<sup>9</sup> La letteratura suggerisce soglie di forte correlazione a partire da valori oltre 0.70 (Bryman & Cramer, 1990; Tabachnick & Fidell, 1996); inoltre valori oltre la soglia di 0.85 sono indicativi di altissima correlazione tali da non garantire la distinzione tra le variabili coinvolte (Kline, 2005). Quindi per variabili fortemente correlate si intende tutte quelle variabili con correlazione oltre 0.70; meglio se vicine, od oltre, 0.85.

<sup>10</sup> La stima della varianza dello stimatore del totale  $T(\hat{Y})$  è calcolata come:

$$\widehat{Var}(\hat{Y}) = \sum_{h=1}^H N_h * \frac{1 - \frac{n_h}{N_h}}{n_h} * s_h^2$$

$$s_h = \sqrt{\frac{\sum_{j=1}^{n_h} w_h * y_{hj} - m_h(y)^2}{N_h}}$$

del 95%, il valore di  $z_{\alpha/2}$  è pari al noto valore di 1,96. In definitiva, fissare la precisione della stima [4] ed il livello fiduciario equivale a definire l'ampiezza massima accettabile dell'intervallo di confidenza per garantire una certa accuratezza della stima rispetto al valore vero  $Y$  nella popolazione.

### 4.3 Calibrazione dei pesi

Un'indagine campionaria, oltre alle stime della popolazione, può fornire anche stime relative ad alcune sottopopolazioni di interesse. Ad esempio, come già detto, la RICA essendo stratificata a livello regionale, per OTE e classi di dimensione economica, garantisce una certa affidabilità delle stime entro questi domini di studio. Tuttavia, le diverse esigenze informative a cui risponde l'indagine RICA, possono comprendere anche la stima di parametri di interesse definiti in domini di studio non previsti nel piano di campionamento.

Se si vogliono ottenere delle stime per domini di studio non pianificati è necessario inizialmente post-stratificare secondo la stratificazione scelta. Purtroppo, questa post-stratificazione darà inevitabilmente luogo a degli errori nelle stime che possono rivelarsi ben al di sopra di quelli accettabili, perché appunto non sono stati pianificati nel piano iniziale di campionamento.

Per ovviare a questa problematica e mantenere, quindi, una coerenza tra le stime prodotte in domini di studio non pianificati con quelle del dominio pianificato si è deciso di applicare un processo di calibrazione dei pesi (Deville & Särndal, 1992) che rispetti le seguenti due condizioni adattate a questa esigenza per la RICA:

- 1) I pesi calibrati si discostino il meno possibile dai pesi ottenuti dalla post-stratificazione scelta;
- 2) la stima del totale di ciascuna variabile di interesse ottenuta con questi pesi calibrati deve coincidere con la stima del totale noto ottenuta con i pesi relativi ai domini pianificati a livello territoriale (regionale e nazionale) in modo da mantenere una uniformità ed una coerenza.

In altri termini, tale procedura consiste nella soluzione di un problema di minimo vincolato (Nicolini et al., 2013), ovvero:

$$\left\{ \begin{array}{l} \min_{w_j} \sum_{j \in S} G_j(w_j; a_j) \\ \text{sotto il vincolo } \sum_{j \in S} w_j y_j = \hat{Y} \end{array} \right. \quad [6]$$

dove  $w_j$  e  $a_j$  rappresentano rispettivamente il nuovo peso incognito e quello iniziale noto ottenuto dalla prima post-stratificazione da calibrare,  $G_j$  una funzione di distanza tra i due pesi che va minimizzata sotto il vincolo che la stima del totale noto,  $\hat{Y}$ , per ciascuna variabile di interesse  $Y$ , sia uguale a quella ottenuta con i nuovi pesi.

Una volta ottenuti i pesi calibrati è necessario ricalcolare anche gli errori delle stime, secondo quanto già esposto nel paragrafo 4.2, per valutare l'accuratezza della nuova stratificazione di dominio scelta<sup>11</sup>.

<sup>11</sup> Per l'utilizzo a fini pratici dei pesi, anche nel caso di domini non pianificati, si rimanda alla nota pratica disponibile in BDR on-line al momento dell'accesso.

## 5 Gli outlier

Spesso, in un'indagine campionaria è possibile trovare valori anomali (o outlier), che possono contribuire a distorcere le stime pianificate e condurre a risultati fuorvianti (sottostima o sovrastima del fenomeno). In generale, un dato si definisce anomalo quando risulta considerevolmente diverso dalla maggior parte dei dati presenti in una distribuzione poiché ha un valore estremamente elevato, o estremamente basso, rispetto al resto della distribuzione tale da costituire un'anomalia, ossia un caso molto raro, isolato. In tal senso, il dato anomalo ha una duplice natura: una plausibile ed una implausibile. Nel primo caso potrebbe essere considerato ancora un dato valido a tutti gli effetti per il collettivo di riferimento, seppur, come detto, raro, ed andrebbe trattato a parte; mentre nel secondo caso sarà considerato un errore, perché classificato come un valore appunto implausibile per il collettivo di riferimento e va escluso. Un dato anomalo, tuttavia, non deve essere confuso con un dato estremo. I dati estremi sono dati *potenzialmente* anomali, ma sempre plausibili in una distribuzione perché hanno valori elevati o bassi, ma non così elevati o così bassi da essere considerati propriamente anomali.

Pertanto, il primo passo da affrontare prima di iniziare un'analisi dei dati è quello di impostare un metodo di trattamento preliminare degli stessi al fine di acquisire un'adeguata conoscenza della loro natura e minimizzare eventuali distorsioni dei risultati dovuti ad anomalie.

Nella scelta di tale metodo, occorre tenere bene in considerazione sia alcuni aspetti caratteristici del campione RICA quali, ad esempio, la scarsa numerosità delle unità del collettivo delle sottopopolazioni del campione e la forma spesso molto asimmetrica della loro distribuzione che alcune variabili di natura economica che rivestono una notevole importanza nelle analisi del settore quali i ricavi totali aziendali (RTA), i costi correnti (CC) e il reddito netto (RN).

In questi anni, tra le tecniche univariate<sup>12</sup> utilizzate nel trattamento preliminare dei dati RICA, la più efficace è risultata essere l'analisi grafica non parametrica dei boxplot, originariamente proposta da J. W. Tukey (1977). Tale tecnica, partendo dalla costruzione del boxplot della distribuzione dei dati, definisce due baffi (i.e., *whiskers*), uno inferiore e uno superiore, identificati dai seguenti due intervalli (i.e., *fence*):

$$[Q_1 - 1,5 IQR; Q_3 + 1,5 IQR] \quad [1]$$

$$[Q_1 - 3 IQR; Q_3 + 3 IQR] \quad [2]$$

dove  $IQR = Q_3 - Q_1$  rappresenta lo scarto interquartile (*interquartile range*, IQR). I dati esterni all'intervallo [1] sono definiti come estremi, mentre quelli esterni all'intervallo [2] sono definiti come anomali. Dall'esempio in figura 5.1a si nota come per una distribuzione abbastanza simmetrica<sup>13</sup> (i.e., indici di asimmetria/skewness = 0,535 e curtosi/kurtosis = 0,377), la tecnica del boxplot di Tukey individua abbastanza bene i valori estremi (vedi figura 5.1b; non sono presenti valori anomali).

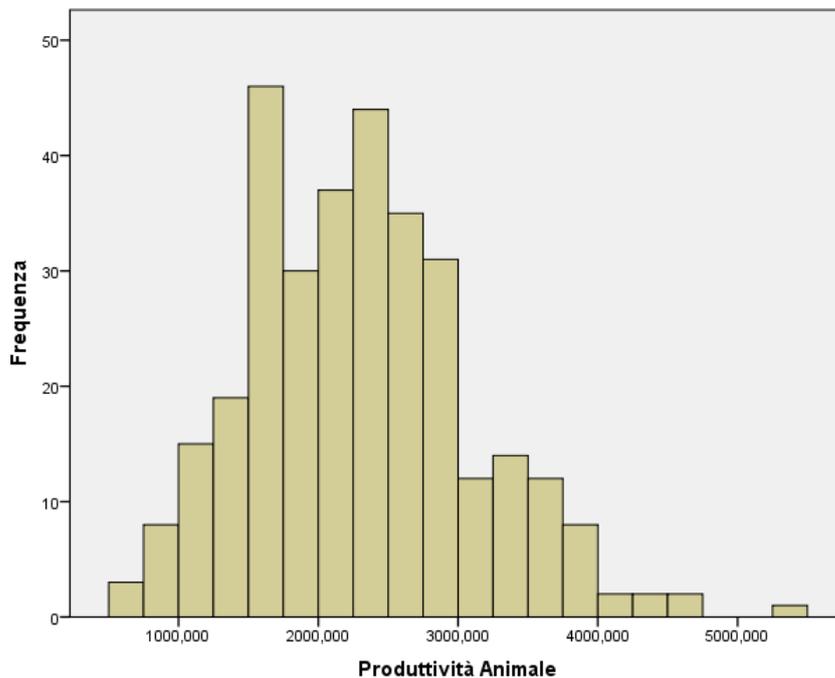
---

<sup>12</sup> Il trattamento degli outlier nei dati RICA attualmente avviene solo in modo univariato, ossia per singole variabili, e non multivariato, ossia per più di una variabile simultaneamente. La motivazione della scelta di un approccio univariato è stata di tipo conservativo a causa della mancanza di univocità nelle ipotesi relazionali tra le variabili strategiche rilevate tale da supportare un approccio multivariato.

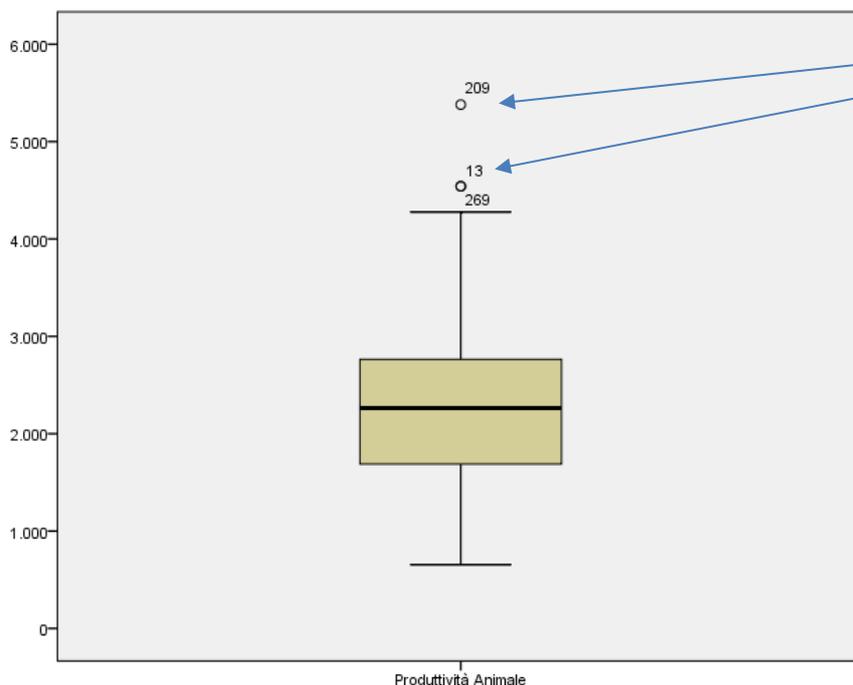
<sup>13</sup> con valore zero, o vicino allo zero, per entrambi gli indici di asimmetria e curtosi la distribuzione è perfettamente simmetrica, mentre con valori rispettivamente  $<|2|$  e  $<|7|$  la distribuzione è leggermente asimmetrica, mentre con valori  $>|2|$  e  $>|7|$  risulta fortemente asimmetrica (Boosma & Hoogland, 2001; West et al., 1995).

Figura 5.1. – Output SPSS - Produttività Animale anno 2013, campione di 321 aziende zootecniche, vacche da latte, nella pianura padana (Emilia-Romagna, Lombardia, Veneto).

a. istogramma



b. Box plot di Tukey, i piccoli cerchi rappresentano i valori estremi.



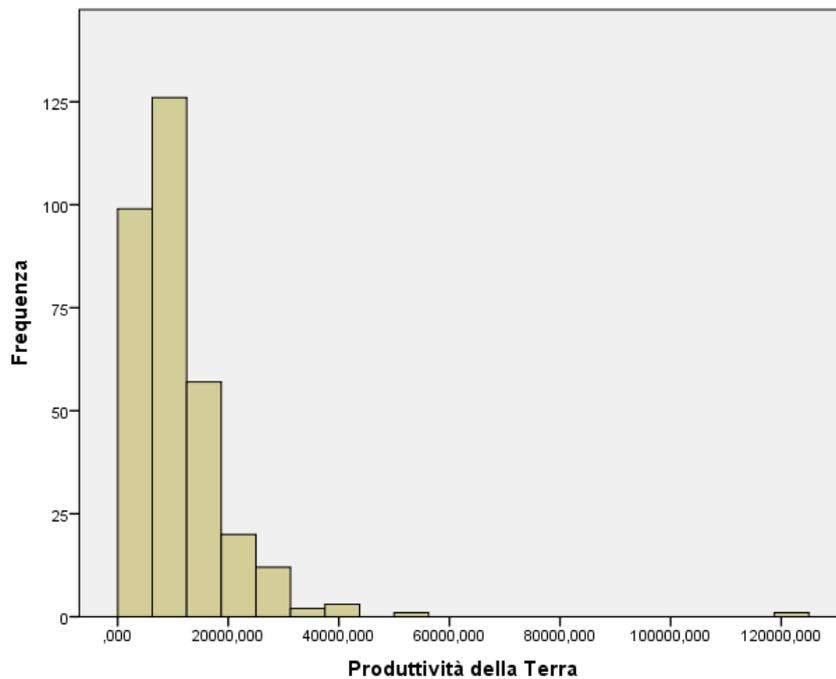
valori estremi, visualizzati con dei cerchi nell'output di SPSS.

Tuttavia, quando la distribuzione dei dati risulta essere particolarmente asimmetrica, questa tecnica con difficoltà può identificare e distinguere valori estremi da quelli anomali. Dall'esempio in figura 5.2a si nota come per una distribuzione molto asimmetrica (i.e., indici di asimmetria/skewness = 5,466 e curtosi/kurtosis = 54,643), la tecnica del boxplot di Tukey può trovare difficoltà nell'individuare e distinguere con chiarezza i valori estremi da quelli anomali. In

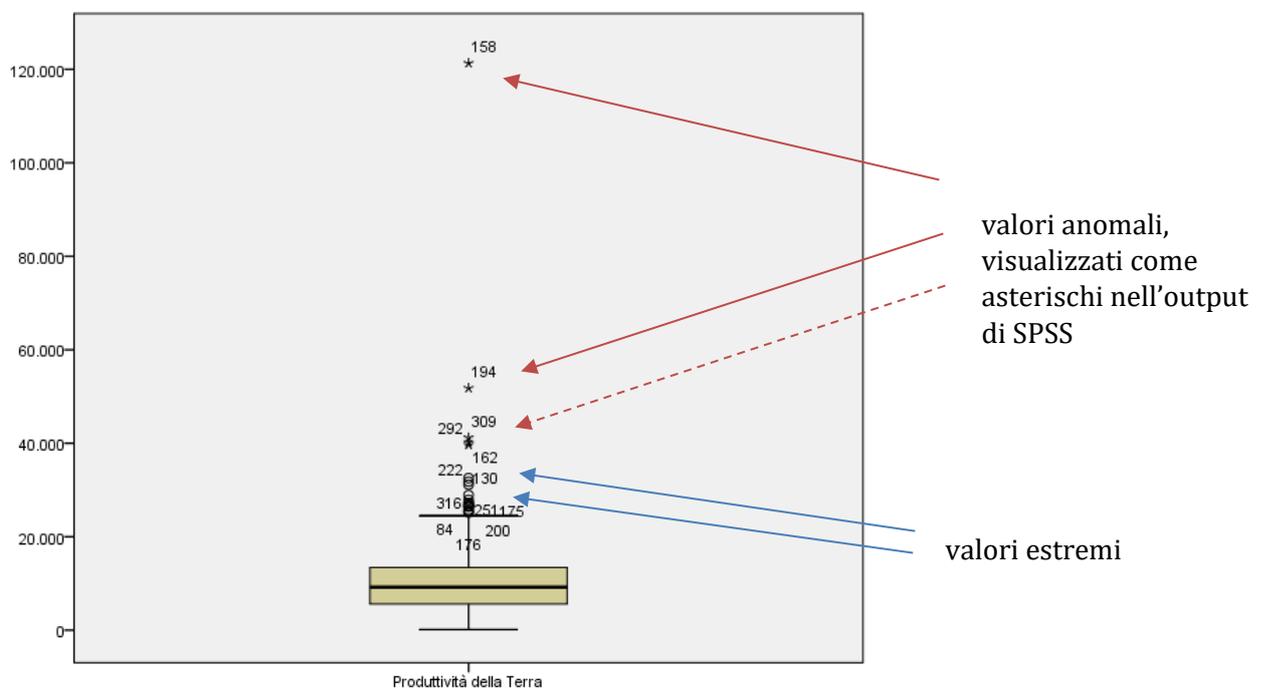
figura 5.2b si evince come la tecnica di Tukey non riesce sempre chiaramente a distinguere gli estremi dagli anomali (soprattutto per quei valori indicati con la linea tratteggiata in rosso).

Figura 5.2. – Output SPSS - Produttività della Terra anno 2013, campione di 321 aziende zootecniche, vacche da latte, nella pianura padana (Emilia-Romagna, Lombardia, Veneto).

a. istogramma



b. Box plot di Tukey



Per questo motivo, la tecnica qui proposta per il trattamento preliminare dei dati RICA è quella di un boxplot corretto per distribuzioni asimmetriche (*Adjusted Boxplot for Skewed Distributions*; Hubert & Vandervieren, 2007) che tiene in considerazione la misura dell'asimmetria basata sul concetto di mediana (indice di centralità statistica notoriamente robusto nell'individuare gli outlier

in quanto il suo valore non viene modificato in loro presenza) per la determinazione dei baffi. Inoltre, tale metodo può essere applicato senza dover effettuare alcuna ipotesi sulla natura probabilistica della distribuzione in esame, mantenendo, quindi, la peculiarità non-parametrica della tecnica originaria di Tukey.

Questo boxplot corretto utilizza come misura della asimmetria il *medcouple* (MC) (Brys et al., 2004), o mediana a coppia, definito come:

$$MC = \text{med}_{x_i \leq Q_2 \leq x_j} h(x_i, x_j) \quad [3]$$

dove *med* sta per mediana (i.e., *median*),  $Q_2$  rappresenta la mediana del campione per ogni  $x_i \neq x_j$  e  $h$  rappresenta la funzione chiamata *kernel* ed è data da:

$$h(x_i, x_j) = \frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i} \quad [4]$$

Tale funzione [4] *kernel* misura, in generale ed in maniera standardizzata<sup>14</sup>, la distanza di tutte le coppie di valori  $x_i \neq x_j$  della distribuzione dalla mediana della distribuzione stessa  $Q_2$ . La MC, nello specifico, rappresenta la mediana di tutti i valori della funzione *kernel*  $h(x_i, x_j)$ , non per ogni  $x_i \neq x_j$ , ma solo di quelli che si trovano rispettivamente al di sotto (i.e.,  $x_i \leq Q_2$ ) e al di sopra (i.e.,  $x_j \geq Q_2$ ) della mediana della distribuzione iniziale  $Q_2$ . In altre parole, la MC fornisce una misura del grado di asimmetria della distribuzione oggetto di studio esaminandone tutta la forma confrontando, a coppie, i valori più distanti, sia in difetto che in eccesso, dalla mediana iniziale  $Q_2$ . Intuitivamente, questo metodo assicura una misurazione più robusta del grado di asimmetria perché tiene conto delle distanze tra tutte le coppie possibili di valori potenzialmente elevati e, di conseguenza, consente una rappresentazione più accurata dei dati, migliorando l'efficacia del processo di identificazione dei valori anomali in una distribuzione.

Ora, la modifica proposta per "correggere" il boxplot standard di Tukey consiste proprio nell'introdurre questo concetto di *medcouple* nella formula degli intervalli di Tukey [1]:

$$[Q_1 - h_l(MC)IQR; Q_3 + h_u(MC)IQR] \quad [5]$$

dove  $h_l(MC)$  e  $h_u(MC)$  sono rispettivamente due funzioni *kernel* che includono l'iniziale costante<sup>15</sup> (i.e., 1,5) introdotta da Tukey per definire i baffi rispettando la condizione di simmetria dove per  $MC=0$  si ha che  $h_l(0) = h_u(0) = 1,5$  e la [5] eguaglierà la [1].

Hubert e Vandervieren (2007) nelle loro simulazioni hanno riscontrato come la miglior forma funzionale per definire  $h_l(MC)$  e  $h_u(MC)$  sia quella ottenuta con modello esponenziale in base al valore ottenuto di MC:

$$[Q_1 - 1,5e^{-4MC}IQR; Q_3 + 1,5e^{3MC}IQR] \quad ; \text{ con } MC \geq 0 \quad [6]$$

$$[Q_1 - 1,5e^{-3MC}IQR; Q_3 + 1,5e^{4MC}IQR] \quad ; \text{ con } MC < 0 \quad [7]$$

<sup>14</sup>  $h$  varia da -1 e +1; inoltre più i valori sono vicini alla mediana  $Q_2$  e più la asimmetria si attenua e la funzione *kernel* restituisce valori prossimi allo zero.

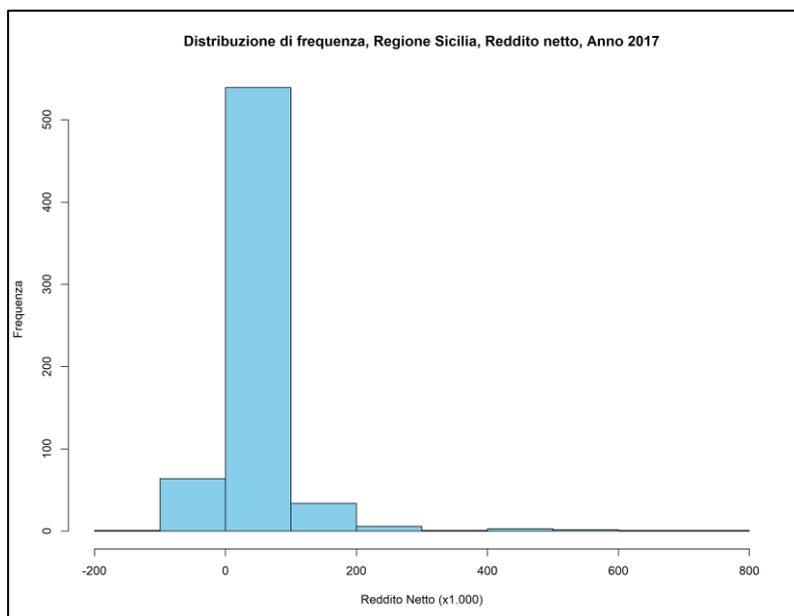
<sup>15</sup> Da notare che solo la costante 1,5 viene considerata nelle due funzioni *kernel*, questo perché rappresentano i valori soglia dai quali partire per ridefinire i baffi corretti per distribuzioni asimmetriche.

Le formule [6] e [7] rappresenteranno, quindi, gli intervalli utilizzati per il calcolo del boxplot corretto<sup>16</sup> per l'individuazione degli outlier nei dati RICA relativi a tre variabili economiche quali ricavi totali aziendali, costi correnti e reddito netto. Un'azienda è stata considerata outlier se sia risultata tale per tutte e tre le variabili e nella BDR è presente un flag che individua l'azienda come outlier.

In figura 5.3 viene rappresentato un esempio numerico relativo alla variabile Reddito Netto per le aziende agricole in Sicilia nell'anno 2017. Dall'istogramma della variabile in figura 5.3a si evince come la distribuzione sia molto asimmetrica a destra (i.e., indici di asimmetria/skewness = 5,4 e curtosi/kurtosis = 42,7). Inoltre, in questo caso ci troviamo in presenza anche di valori negativi, i quali, presumibilmente, dovrebbero costituire di per sé un'eccezione, un'anomalia per una variabile come il Reddito Netto, che solitamente assume valori positivi. Nella figura 5.3b è facile notare come il boxplot di Tukey non sia in grado di distinguere i valori estremi da quelli anomali, sovrastimandoli nel baffo superiore (i.e., ne individua di più di quanti in realtà non siano poiché molti si trovano indistintamente oltre il baffo superiore) e sottostimandoli in quello inferiore (i.e., ne individua di meno di quanto in realtà non siano poiché molti si trovano indistintamente entro il baffo inferiore); al contrario il boxplot corretto tiene invece conto della forma della distribuzione, e quindi della sua forte asimmetria, riuscendo a definire correttamente gli intervalli e quindi tenere fuori solo i valori potenzialmente anomali.

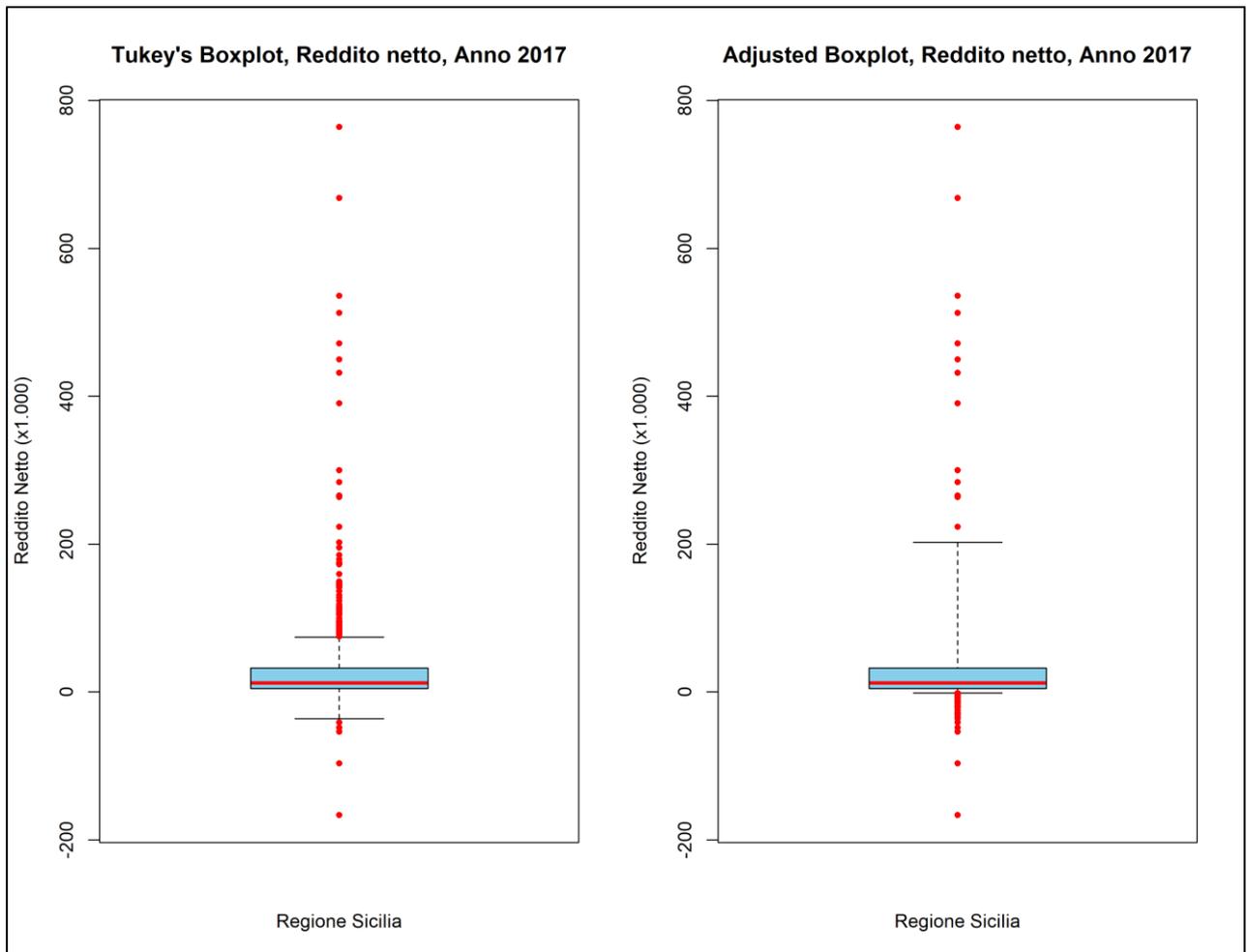
**Figura 5.3 - Output R – Reddito Netto anno 2017, campione di 652 aziende, regione Sicilia**

**a. istogramma**



<sup>16</sup> Il calcolo automatico dei boxplot corretti è stato effettuato tramite il pacchetto open source R (<https://cran.r-project.org/web/packages/robustbase/robustbase.pdf>)

**b. Boxplot**



# Bibliografia

- **Ballin M.** (a cura di) (2004) "Indagine sui Risultati Economici delle Aziende Agricole RICA-REA". Anni 2002-2003-2004. Istruzioni per la rilevazione delle variabili REA. ISTAT, Roma
- **Bethel J.** (1989): "Sample allocation in multivariate survey", Survey Methodology, vol. 15 n. 1
- **Boomsma, A., & Hoogland J. J.** (2001). The Robustness of LISREL Modelling Revisited. In R. Cudeck, S. Du Toit, D. Sörbom (Eds.), Structural Equation Modeling: Present and Future. Chicago, IL: Scientific Software International Inc.
- **Bryman, A., & Cramer, D.** (1990). Quantitative data analysis for social sciences. London: Routledge
- **Brys, G., Hubert, M., Struyf, A.** (2004). A Robust Measure of Skewness. Journal of Computational and Graphical Statistics, Vol 13, pp. 996-1017.
- **Cardillo C., De Leo S., Gabrieli G., Zanolli A.** (2010), "Il campione RICA: metodologia di calcolo dei pesi e analisi dell'affidabilità delle stime" Collana Metodi RICA, INEA.
- **Cocchi, D.** (2011). Teoria dei Campioni. Dispense didattiche. Dipartimento di Scienze Statistiche "Paolo Fortunati". Università degli studi di Bologna. Accessibile presso: <https://studylibit.com/doc/6539708/indice>
- **Commissione Europea** (1982), Regolamento (CEE) N. 1859/82 del 12 luglio 1982, Scelta delle aziende contabili ai fini della constatazione dei redditi nelle aziende agricole.
- **Commissione Europea** (2008), Regolamento (CE) N. 1242/2008 dell'8 dicembre 2008, Istituzione di una tipologia comunitaria delle aziende agricole
- **Commissione Europea** RI/CC1519 e successive revisioni (2008): Implementation of the new typology, FADN field of survey, 2010 selection plan working programme
- **Commissione Europea** RI/CC 1524 (2009): Implementation of the new typology, New thresholds
- **Commissione Europea** RI/CC 1526 (2009): Guidelines for selection plans
- **Consiglio Europeo** (1965), Regolamento N. 79/65/CEE del del 15 giugno 1965, Istituzione di una rete d'informazione contabile agricola sui redditi e sull'economia delle aziende agricole nella Comunità Economica Europea.
- **Commissione Europea** (2014) Regolamento Delegato (UE) N. 1198/2014 del 1 agosto 2014 che integra il regolamento (CE) n. 1217/2009 del Consiglio relativo all'istituzione di una rete d'informazione contabile agricola sui redditi e sull'economia delle aziende agricole nell'Unione europea.
- **Commissione Europea** (2015) Regolamento di esecuzione (UE) 2015/220 del 3 febbraio 2015 recante modalità di applicazione del regolamento (CE) n. 1217/2009 del Consiglio relativo all'istituzione di una rete d'informazione contabile agricola sui redditi e sull'economia delle aziende agricole nell'Unione europea.
- **Consiglio Europeo** (2009) Regolamento (CE) N. 1217/2009 del del 30 novembre 2009, Istituzione di una rete d'informazione contabile agricola sui redditi e sull'economia delle aziende agricole nella Comunità europea
- **Deville J-C & Sarndal C-E.** (1992). Calibration Estimators in Survey Sampling, Journal of the American Statistical Association Vol. 87, No. 418 (Jun. 1992), pp. 376-382
- **Hubert M., & Vandervieren, E.** (2007) An adjusted boxplot for skewed distributions. Computational Statistics and Data Analysis, 52(12), pp.5186-5201
- **ISTAT** (2008): Nota metodologica sulla Indagine SPA 2007

- **ISTAT** (2008): Statistiche in breve, Indagine sulla Struttura e produzione delle aziende agricole anno 2007
- **Jean-Claude Deville and Carl-Erik Sarndal**, Calibration Estimators in Survey Sampling, Journal of the American Statistical Association Vol. 87, No. 418 (Jun., 1992), pp. 376-382
- **Kline, B., R.** (2005). Principles and Practice of Structural Equation Modeling. (Second Edition). The Guilford Press. NY
- **Mari F., Rossi R.** (a cura di) (2010) I Redditi Lordi Standard e gli Standard Output “2004” delle attività produttive agricole italiane, INEA, Roma
- **Mazzocchi, M.** (2008). Statistics for Marketing and Consumer Research. Sage Publications.
- **Nicolini G., Marasini D., Montanari G.E., Pratesi M., Ranalli M.G., Rocco E.** (2013), Metodi di stima in presenza di errori non campionari, Springer-Verlag
- **Parlamento Europeo e Consiglio Europeo (2013)**, Regolamento (UE) N. 1318/2013 del 22 ottobre 2013, Modifica del regolamento (CE) n. 1217/2009 del Consiglio, relativo all’istituzione di una rete d’informazione contabile agricola sui redditi e sull’economia delle aziende agricole nella Comunità europea.
- **Rondinelli, V.** (2011). Il piano di selezione delle aziende agricole RICA nel campione RICA-REA 2010. ISTAT Working Papers, N. 16. ISBN 88-458-1703-2.
- **Tabachnick, B. G., & Fidell, L. S.** (1996). Using multivariate statistics. New York: Harper & Row
- **West, S., G., Finch, J., F., & Curran, P., J.,** (1995). Structural Equation Models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), Structural equation modelling: Concepts, issues and applications (pp. 56.75). Newbury Park, CA: Sage

# Appendice

## Glossario

**AAL:** Altre Attività Lucrative. Sono definite anche attività complementari o connesse. Esse identificano le attività connesse a quella agricola primaria di produzione di beni. Rientrano tra queste attività l'agriturismo, le fattorie didattiche, la trasformazione dei prodotti aziendali ecc.

**CC:** Costi Correnti. Comprendo tutti i costi variabili, inclusi i reimpieghi aziendali, per l'acquisizione dei mezzi tecnici a logorio totale e dei servizi necessari per realizzare le attività messe in atto dall'azienda, siano esse prettamente agricole sia per realizzare prodotti e servizi derivanti dalle attività complementari.

**OTE:** Orientamento Tecnico Economico. Esso definisce l'indirizzo produttivo dell'azienda ed il suo grado di specializzazione, sulla base dell'incidenza percentuale, in termini di PS, delle diverse attività produttive sul totale della Produzione standard aziendale.

**PS:** Produzione Standard (o Standard Output SO). Valore monetario della produzione vegetale o animale che include le vendite, i reimpieghi, l'autoconsumo e i cambiamenti nello stock dei prodotti, al prezzo franco azienda (fanno eccezione i prodotti per i quali è impossibile la vendita senza il confezionamento). La Produzione Standard non include i pagamenti diretti, l'IVA e le tasse sui prodotti.

**RN:** Reddito Netto. Rappresenta la remunerazione dell'imprenditore agricolo nelle sue diverse forme giuridiche. Nel Bilancio riclassificato RICA\_INEA è ottenuto come differenza tra il RO e gli oneri finanziari e straordinari (in diminuzione), e gli aiuti pubblici in conto capitale e quelli in conto esercizio del 2 Pilastro della PAC.

**RTA:** Ricavi Totali Aziendali. Rappresentano i ricavi complessivi aziendali per la cessione di prodotti e servizi, costituiti a sua volta dai ricavi delle attività primarie agricole e zootecniche (la cosiddetta PLV), e i ricavi derivanti dalle Attività Complementari, conosciute anche come attività connesse (multifunzionalità).

**SAU:** Superficie Agricola Utilizzata. Rappresenta la superficie agricola utilizzata per realizzare le coltivazioni di tipo agricolo, escluse quindi le coltivazioni per arboricoltura da legno (pioppeti, noceti, specie forestali, ecc.) e le superfici a bosco naturale (latifoglie, conifere, macchia mediterranea). Dal computo della SAU sono escluse le superfici delle colture intercalari e quelle delle colture in atto (non ancora realizzate). La SAU comprende invece la superficie delle piantagioni agricole in fase di impianto.

**UBA:** Unità Bovine Adulte. La consistenza degli allevamenti viene determinata attraverso le UBA. Tali unità di misura convenzionale derivano dalla conversione della consistenza media annuale delle singole categorie animali nei relativi coefficienti definiti nel Reg. CE 1974/2006. Sono esclusi da calcolo gli animali allevati in soccida.

**ULT:** Unità di Lavoro annue. Le unità di lavoro sono rappresentate dalla manodopera familiare e salariata. Le ULT vengono calcolate secondo il parametro 2.200 ore/anno/persona. Per tutti i componenti della manodopera sia familiare che retribuita (avventizi esclusi) le UL vengono calcolate per ogni soggetto dividendo il numero di ore prestate nel corso dell'esercizio contabile per il

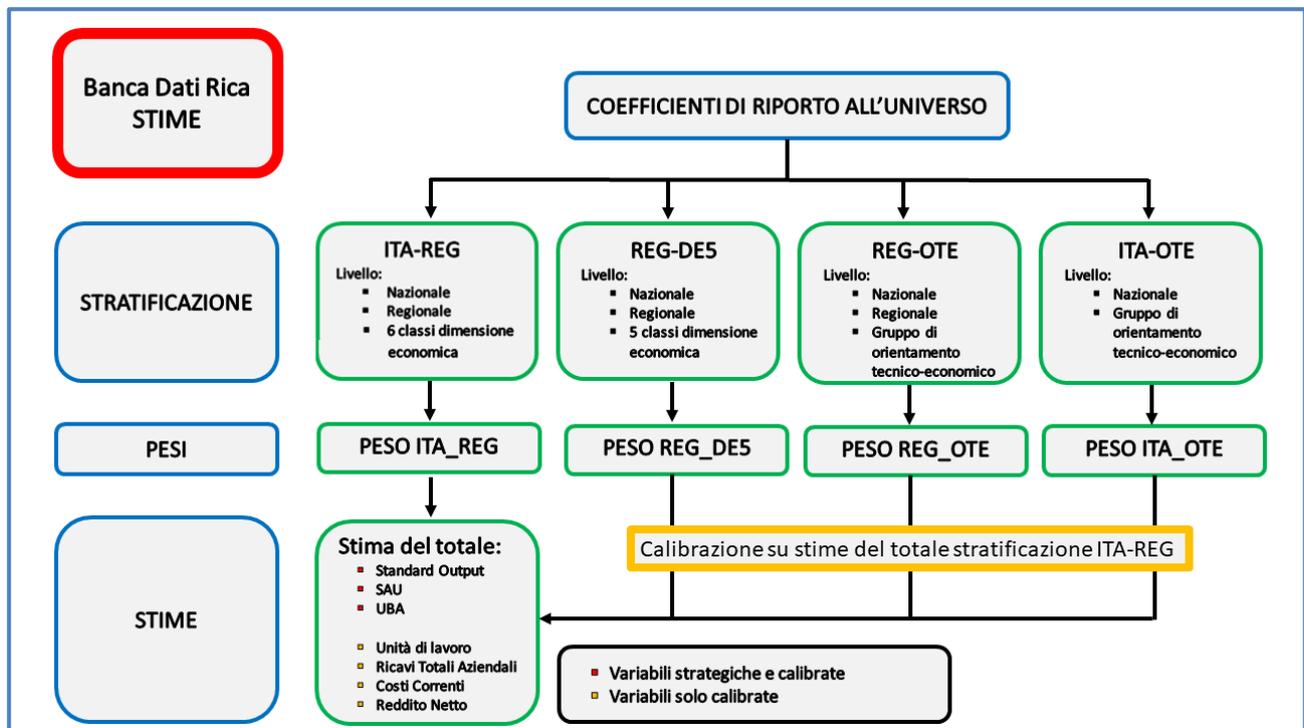
parametro 2.200. Nel caso in cui il numero di ore prestate da un singolo componente è superiore alle 2.200 ore/anno la UL sarà uguale a 1, mentre nel caso in cui invece il numero di ore è inferiore a 2.200 allora la UL sarà proporzionale alle ore effettivamente prestate. La sommatoria delle UL dei singoli componenti la manodopera così calcolate vengono sommate alle UL della manodopera avventizia, determinata dal rapporto delle ore prestate dai gruppi di avventizi per il parametro 2.200. Dalle UL aziendali sono escluse le ore prestate dalla manodopera derivante dai servizi di contoterzismo passivo. Nel calcolo delle UL è compreso invece lo scambio della manodopera tra aziende agricole limitrofe.

## Utilizzo dei Pesi

Il campione RICA è un campione casuale stratificato e consente, quindi, l'utilizzo di coefficienti di riporto all'universo (pesi) per estendere i risultati di alcune variabili del campione, definite strategiche, a quelli della popolazione di riferimento secondo il dominio di stima pianificato. Definire un dominio di stima consente di fissare l'affidabilità delle stime di alcune variabili all'interno dello stesso. Nel campione RICA, il dominio di stima è stato definito a livello Nazionale e Regionale (ITA-REG) e le variabili strategiche scelte sono la Produzione Standard (PS), la Superficie Agricola Utilizzata (SAU) e le Unità di Bovino Adulto (UBA).

Per poter estendere le stime del campione RICA anche in domini di stima non pianificati, sono stati implementati coefficienti di riporto all'universo specifici per altri domini di stima (REG-DE5, REG-OTE e ITA-OTE).

Si è proceduto, quindi, dapprima a post-stratificare l'universo di riferimento secondo i nuovi domini di stima non pianificati e poi a calibrare le stime del totale ottenute in ciascun dominio sulle stime del totale del dominio pianificato. Questo garantisce sia una coerenza tra le stime dei diversi domini che un miglioramento dell'accuratezza delle stime.

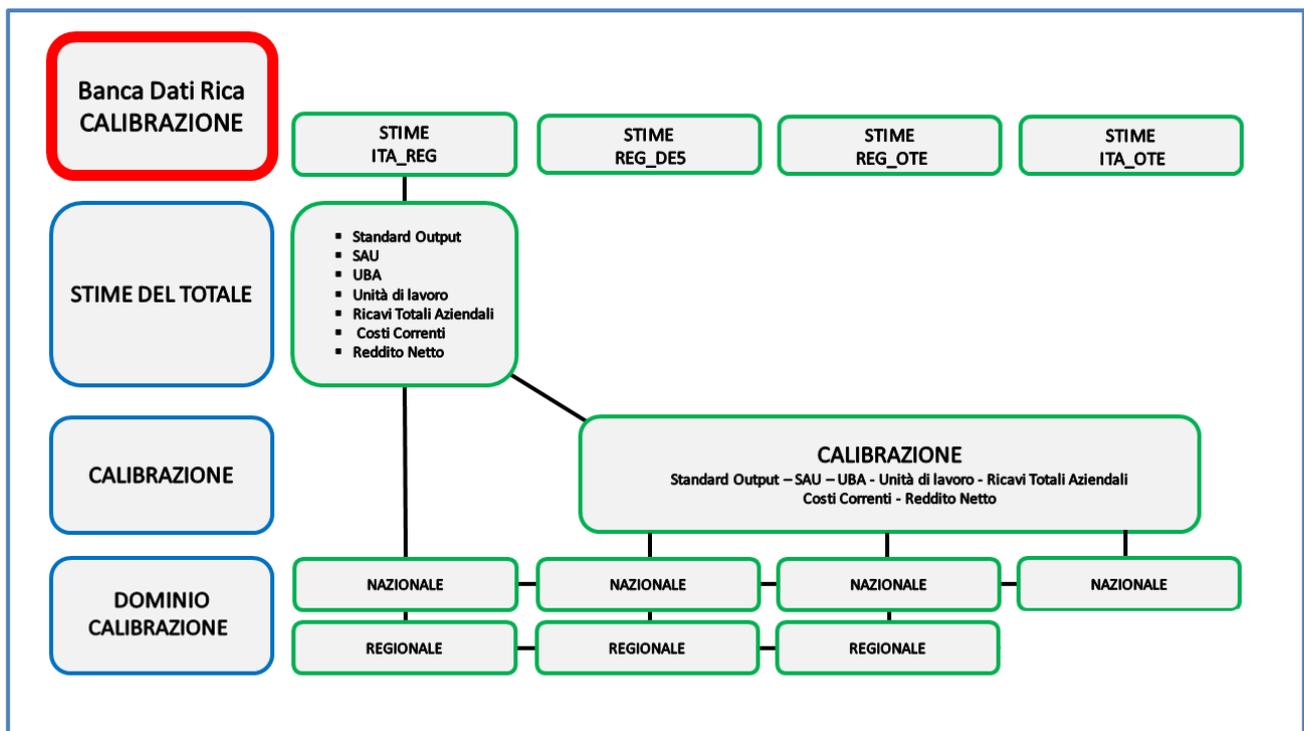


Il grafico sopra riportato sintetizza la metodologia adottata in ambito RICA riguardo ai coefficienti di riporto all'universo appena descritta.

Per il processo di calibrazione, sono state utilizzate 7 variabili:

- 1) Le 3 variabili strategiche, quali Standard Output, Superficie Agricola Utilizzata e Unità di Bovino Adulto;
- 2) Altre 4 variabili ritenute di interesse per le analisi in ambito agricolo-economico, quali Unità di Lavoro Totali, Ricavi Totali Aziendali, Costi Correnti e Reddito Netto.

Dal momento che le stime estese all'universo nel dominio ITA-REG, essendo state pianificate nel piano di campionamento in termini di affidabilità, rappresentano i valori attendibili su cui calibrare le altre, non pianificate. Quindi, le stime del totale di queste 7 variabili nei domini non pianificati, quali REG-DE5, REG-OTE e ITA-OTE convergeranno tutte alle stime del dominio pianificato ITA-REG, secondo lo schema sotto riportato:



Nell'immagine sotto riportata risulta ben visibile l'impatto della calibrazione sulle stime nei domini non pianificati quali, ad esempio, il dominio REG-OTE. Infatti, per quest'ultimo dominio non pianificato, le stime a livello nazionale differiscono molto da quelle del dominio ITA-REG. Questo accade in quanto il campione RICA, essendo stato pianificato a livello nazionale e regionale, nel processo di post-stratificazione per regione e polo produttivo, non ha un numero idoneo in termini di caratteristiche di aziende per adeguarsi a questa nuova stratificazione provocando una distorsione delle stime ottenute. Per questo motivo, si rende necessaria una calibrazione dei risultati, per migliorare l'affidabilità delle stime per domini non pianificati, facendo convergere le stime del totale dei domini non pianificati ai totali del dominio ITA-REG.

CALIBRAZIONE STIMA DEL TOTALE		ANNO 2014 – REGIONE VALLE D'AOSTA			
		Produzione standard	Superficie agricola utilizzata	Unità Bovino Adulto	Unità di lavoro
DOMINIO PIANIFICATO	ITA-REG	57.574.766	43.351	29.726	2.259
DOMINIO NON PIANIFICATO	REG-OTE Pesi originali post-stratificazione	106.400.853	106.427	52.671	3.052
CALIBRAZIONE PESI REG-OTE SU STIME DEL TOTALE ITA-REG					
DOMINIO NON PIANIFICATO	REG-OTE Pesi calibrati	57.574.795	43.351	29.726	2.259

<b>Dominio di stima pianificato</b>	
<b>Stratificazione ITA-REG</b>	
▪ Dominio di stima	Livello nazionale e regionale
▪ Stratificazione	Regione, gruppo di orientamento-tecnico-economico e 6 classi di dimensione economica (6 classi)
<b>Classi di dimensione economica</b>	
<b>CLASSE</b>	<b>LIMITI (€)</b>
Classe 1	8.000-25.000
Classe 2	25.000-50.000
Classe 3	50.000-100.000
Classe 4	100.000-500.000
Classe 5	500.000-1.000.000
Classe 6	≥ 1.000.000
<b>Domini di stima non pianificati</b>	
<b>Stratificazione REG-DE5</b>	
▪ Dominio di stima	Livello nazionale e regionale
▪ Dominio di calibrazione	Livello nazionale e regionale
▪ Stratificazione	Regione e classi di dimensione economica (5 classi)
<b>Classi di dimensione economica</b>	
<b>CLASSE</b>	<b>LIMITI (€)</b>
Classe 1	8.000-25.000
Classe 2	25.000-50.000
Classe 3	50.000-100.000
Classe 4	100.000-500.000
Classe 5	≥500.000
<b>Stratificazione REG-OTE</b>	
▪ Dominio di stima	Livello nazionale e regionale per polo produttivo (10 classi)
▪ Dominio di calibrazione	Livello nazionale e regionale
▪ Stratificazione	Regione e polo produttivo (10 classi)
<b>Poli produttivi (riclassificazione CREA-PB)</b>	
<b>CODICE</b>	<b>DESCRIZIONE</b>
100	Seminativi
110	Cerealicoltura
200	Ortofloricoltura
310	Viticultura
320	Olivicoltura
330	Fruttiferi
400	Erbivori
410	Bovini da latte
500	Granivori
800	Miste coltivazioni ed allevamenti
<b>Stratificazione ITA-OTE</b>	
▪ Dominio di stima	Livello nazionale per polo produttivo (10 classi)
▪ Dominio di calibrazione	Livello nazionale

- Stratificazione Polo produttivo (10 classi)

**Poli produttivi (riclassificazione CREA-PB)**

<b>CODICE</b>	<b>DESCRIZIONE</b>
100	Seminativi
110	Cerealicoltura
200	Ortofloricoltura
310	Viticoltura
320	Olivicoltura
330	Fruttiferi
400	Erbivori
410	Bovini da latte
500	Granivori
800	Miste coltivazioni ed allevamenti