



PSRN-Biodiversità – sottomisura 10.2, progetto Latteco2  
«Le razze bovine da latte per la definizione di modelli selettivi sostenibili»,  
ANAFIBJ Comparto Bovini latte



“Fondo europeo agricolo per lo sviluppo rurale: l’Europa investe nelle zone rurali”  
Autorità di gestione: MASAF Ministero dell’Agricoltura della Sovranità Alimentare e delle Foreste  
Spesa ammessa a contributo Progetto LATTeco2 ANAFIBJ: euro 12.535.931,95

# ATTITUDINE CASEARIA: LA CENTRALITÀ DEL DATO

di Ferdinando Galluzzo



**N**ell’edizione di marzo 2023 abbiamo introdotto la tematica dell’attitudine casearia: da allora ad oggi abbiamo lavorato allo sviluppo di un indice genetico che identificasse i **riproduttori con il maggior potenziale genetico per la produzione di un latte il più possibile adatto ad essere trasformato in formaggio**.

Come ogni indice genetico, anche questo si basa sui dati fenotipici (cioè ciò che viene registrato in stalla) e sulla capacità del modello statistico di estrarre la variabilità di essi non dovuta alla genetica, per identificarne la componente trasmissibile: il “cibo” del nostro modello statistico sono, quindi, i dati fenotipici di input.

Riprendendo il discorso iniziato a marzo, i fenotipi che determinano la caseificabilità del latte sono tre: **tempo di coagulazione (RCT), tempo di rassodamento (k20) e consistenza del coagulo (a30)**. Per quanto riguarda questi tre caratteri, dal 2017 ad oggi, disponiamo di circa 6.7 milioni di dati derivanti dal Progetto LEO dell’Associazione Italiana Allevatori.

Il passaggio di “pulizia” dei dati, *editing* dati, è fondamentale per l’accuratezza delle stime finali, ovvero per gli indici genetici dei riproduttori. Per garantire omogeneità fra i gruppi di animali, abbiamo escluso tutte le osservazioni oltre il quinto parto e al di fuori dell’intervallo 5-405 giorni di lattazione. Per quanto riguarda le proprietà di coagulazione del latte, **consideriamo errori ovvi tutte le osservazioni al di fuori del range 5-60 minuti per l’RCT, 1-20 minuti per il k20 e**

**5-60 millimetri per l’a30**. Al fine di **identificare ed escludere anomalie nelle procedure di laboratorio**, abbiamo implementato un algoritmo di *machine-learning* chiamato **isolation forest**: questo algoritmo ci permette di identificare tutti i valori troppo distanti da un riferimento ed escluderli. Il riferimento utilizzato sono le correlazioni fenotipiche fra le proprietà di coagulazione risultanti dallo studio “*Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows*” (Visentin et al, 2015) pubblicato sul Journal of Dairy Science. Le correlazioni fra i fenotipi in nostro possesso sono state calcolate entro gruppi di allevamento-anno-giorno di controllo, i cui campioni sono stati processati nello stesso periodo nello stesso laboratorio: tutte le classi di allevamento-anno-giorno di controllo risultate anomale sono state escluse dall’analisi. Infine, per garantire un’adeguata ed ef-

ficace comparazione, i **gruppi di contemporanee per allevamento-anno-stagione di controllo con meno di 20 osservazioni sono stati eliminati**. Alla fine della pulizia, il numero di osservazioni del *dataset* di lavoro era circa 4 milioni.

Tutto questo lavoro di “pulizia” dati è finalizzato alla selezione delle sole osservazioni accurate: riprendendo un articolo di Bianconero del gennaio 2021, **“non si migliora ciò che non si misura, ma nemmeno ciò che si misura male”**.

Dopo la fase di *editing*, siamo passati all’identificazione del modello statistico che meglio isolasse la componente genetica, senza diventare troppo dipendente dai dati inseriti (e quindi poco stabile, con minor capacità predittiva – fenomeno dell’*overfitting*). Il modello identificato è un **multiple-trait repeatability linear animal model**: un modello il cui effetto genetico è quello dell’animale stesso (*animal model*), caratterizzato dalla presenza di più osservazioni per animale entro e fra le lattazioni (*repeatability model*) e composto da quattro modelli statistici, uno per ogni carattere incluso (*multiple trait model*). **I quattro caratteri sono la caseina percentuale e le tre proprietà di coagulazione del latte: essendo geneticamente correlati, la stima simultanea ci permette di aumentare l’accuratezza degli indici per ognuno**.

Ora, la prossima domanda a cui rispondere è: **quanto sono ereditabili i caratteri dell’attitudine casearia? Che porzione del fenotipo è determinata dalla genetica? La risposta, sempre su Bianconero, nella prossima ed ultima puntata!**

