



UNIVERSITA' DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Scienze Animali

**DOTTORATO DI RICERCA IN:
CONSERVAZIONE, GESTIONE E MIGLIORAMENTO
DELLE RISORSE GENETICHE ANIMALI**

CICLO XIX

**VALUTAZIONE GENETICA DEL CARATTERE LONGEVITÀ DIRETTA
MEDIANTE SURVIVAL ANALYSIS
E VERIFICA DELL'APPLICABILITÀ
NEL PIANO DI SELEZIONE DELLA RAZZA BRUNA
ATTRAVERSO LA CREAZIONE DI UN INDICE COMBINATO.**

Coordinatore e Supervisore : Ch.mo Prof. Paolo Carnier

Dottorando : Omar Bonetti
matr. n° 963234

31 gennaio 2008

SUMMARY:

Objective of this study is to improve the estimation of survival breeding values for Italian Brown Swiss Dairy Cattle. The genetic evaluation was carried out using survival analysis (Survival Kit). The trait “longevity” was never investigated with Survival Analysis in Italian Brown Swiss Dairy Cattle. The Survival Analysis is the powerful methodology for investigate survival data. The trait investigated were the length of productive life (LPL) defined as the period between the first calving and culling. Data on 323'849 Brown Swiss daughters of 11'076 Brown Swiss Bulls were analyzed under a frailty proportional hazard model assuming a Weibull distribution for the baseline hazard function and including sire and herd-year as random effects. Estimated hazard ratios (HR) indicated that age at first calving, herd, parity, deviation (%) from the EVM herd-year mean and the interaction between herd size and the linear regression of herd yearly size based on 4 years were significant on LPL. The estimated variance for herd-year effect were 0,21 and for sire effect were 0,04. The effective heritability estimated was 12% and the equivalent heritability was 10% with 25% of censoring. The genetic (co)variance components and predicted breeding values among LPL and other traits analyzed by Italian Brown Swiss Cattle Breeders Association of about 1500 Bulls with already know Iiterbull information were obtained using an Expectation-Maximization REML procedure based on the MACE equations. Eight different MT analysis were improved. Mean correlation between direct and combined breeding values was 97% ranging from 94,5% to 99,2% Mean of reliability improvement obtained was 0,05 ranging from 0,03 to 0,06.

RIASSUNTO:

Obiettivo di questo studio è di sviluppare un procedura per la stima dei valori riproduttivi per il carattere LPL nei bovini di razza Bruna. La valutazione genetica è stata ottenuta impiegando l'analisi di sopravvivenza (survival Kit). La survival analysis la miglior metodologia utilizzabile per l'approccio a dati di sopravvivenza. Il carattere analizzato è la lunghezza della vita produttiva compreso tra la data di primo parto e la riforma. Sono stati analizzati 323'849 record relative alle carriere produttive delle figlie di 11'076 tori di razza Bruna. Per il modello "frailty" impiegato si è assunta una distribuzione di Weibull per la funzione di rischio basale. I rischi relativi stimati indicano che l'età al primo parto, l'ordine di parto l'allevamento, la deviazione dalla media EVM di allevamento anno e l'interazione tra dimensione annua dell'allevamento e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo sono significativi nel determinismo del rischio di riforma. La varianza stimata per l'effetto allevamento anno è 0,21, mentre per l'effetto genetico legato al toro è 0,04. L'ereditabilità effettiva ottenuta è del 12%, quella equivalente del 10% con un 25% di record incompleti. La stima delle componenti di varianza – covarianza tra LPL e gli altri caratteri analizzati dall'Associazione Nazionale degli Allevatori di razza Bruna ha coinvolto circa 1'500 tori ed è stata ottenuta con la metodologia MACE basata sull' Expectation-Maximization REML multiple trait analysis. Sono state testate 8 differenti analisi. La correlazione media tra EBV di LPL diretta ed LPL combinata è del 97% variabile tra un 94,5% ed un 99,2%. L'incremento medio di attendibilità ottenuto è di 5 punti percentuali variabili da 0,03 a 0,06.

A chiunque creda ancora che la libertà non misuri in gradi...

LA STATISTICA
(Trilussa)

*Sai ched'è la statistica? È 'na cosa
che serve pe' fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che sposa.*

*Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, l'ì, la media sempre eguale
puro co' la persona bisognosa.*

*Me spiego, da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:
e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perché c' un antro che se ne magna due*

Indice:	
Summary	1
Riassunto	2
Indice	4
Indice delle figure	6
1. INTRODUZIONE	8
1.1 La longevità nella razza Bruna	8
1.2 I caratteri funzionali	9
1.3 Caratteri morfologici	12
1.4 Espressione della longevità	14
1.5 Longevità diretta e indiretta	17
1.6 La situazione internazionale	18
1.7 Progresso genetico della razza Bruna	22
1.8 Survival analysis	24
1.8.1 Definizione	24
1.8.1.1 Analisi di sopravvivenza: caratteristiche specifiche	26
1.8.1.2 Censoring	27
1.8.2 Distribuzioni	29
1.8.2.1 Distribuzione esponenziale	30
1.8.2.2 Distribuzione di Weibull	31
1.8.2.3 Distribuzione log-normale	31
1.8.3 Stima parametrica	32
1.8.4 Stima non parametrica	34
1.8.5 Modelli di regressione	35
1.8.6 Bontà di stima e test delle ipotesi (bootstrapping)	39
1.8.7 Frailty models	41
1.9 Bibliografia dell'introduzione	44
2. OBIETTIVI	48
3. MATERIALE E METODI	49
3.1 Base dati	49
3.2 Editing	50
3.3 Survival analysis	51
3.4 Bibliografia di materiale e metodi	57
4. Primo contributo sperimentale: ANALISI DEL RISCHIO E STIMA DEI PARAMETRI GENETICI	58
4.1 Introduzione	58
4.2 Materiale e metodi	59
4.3 Risultati e discussione	61
4.3.1 Effetti non genetici	63
4.3.2 Componenti di varianza e parametri genetici	67
4.4 Bibliografia del primo contributo sperimentale	71
5. Secondo contributo sperimentale: L'INDICE GENETICO LUNGHEZZA DELLA VITA PRODUTTIVA E LE SUE CORRELAZIONI GENETICHE CON GLI ALTRI CARATTERI	73
5.1 Introduzione	73
5.2 Materiale e metodi	74
5.3 Risultati e discussione	76
5.4 Bibliografia del secondo contributo sperimentale	80
6. CONCLUSIONI E PROSPETTIVE FUTURE	82

6.1 Bibliografia di conclusioni e prospettive future	84
Tabelle del secondo contributo sperimentale	85
Figure	

Indice delle figure:

- Figura 1: Numero medio annuo di lattazioni nelle 3 principali razze da latte.
- Figura 2: Confronto tra la dinamica della numerosità totale di primi parti e quelli dei quali si perdono informazioni dal 1990 al 2006.
- Figura 3: Proiezione (linea grigia) della dinamica di riforma delle vacche dal 1990 al 2006.
- Figura 4: Confronto tra il numero medio di lattazioni per anno di primo parto ottenuto dalla proiezione di riforma delle vacche (linea rossa tratteggiata) e quello ottenuto dall'analisi dei dati (linea grigia tratteggiata e barrata). Fino al 1999 la media è comune.
- Figura 5: composizione percentuale della numerosità di parti raggiunti per anno di primo parto e numerosità di parti annui (linea rossa).
- Figura 6: illustrazione del modello ideale di una vacca di razza Bruna.
- Figura 7a: trend genetico dell'Indice Totale Economico della razza Bruna in Italia (espresso su base genetica pari alla media ITE dei tori nati nel 2000 con media 0 e d.s. pari a quella del carattere kg di latte).
- Figura 7b: trend genetico del carattere longevità funzionale (indiretta) della razza Bruna in Italia. (espressa con media 100 e d.s. 12).
- Figura 7c: trend genetico del carattere kg di latte nella popolazione di razza Bruna in Italia.
- Figura 7d: trend genetico del carattere kg di proteine nella popolazione di razza Bruna in Italia.
- Figura 8a: composizione percentuale dell'Indice Totale Economico della razza Bruna fino a maggio 2006. In evidenza il carattere longevità.
- Figura 8b: composizione percentuale attuale dell'Indice Totale Economico della razza Bruna. In evidenza il carattere longevità.
- Figura 9: schematizzazione di dati di sopravvivenza.
- Figura 10: esempio di curva di sopravvivenza dell'esempio in figura 9 (linea continua) e la stessa ottenuta da un'analisi a time-point (doppia linea tratteggiata).
- Figura 11a (a sinistra): esempi di funzioni di sopravvivenza in una distribuzione esponenziale.
- Figura 11b (a destra): esempi di funzioni di distribuzione di probabilità in una distribuzione esponenziale.
- Figura 12a (a sinistra): esempi di funzioni di sopravvivenza in una distribuzione di Weibull.
- Figura 11b (a destra): esempi di funzioni di distribuzione di probabilità in una distribuzione di Weibull.
- Figura 13a (a sinistra): esempi di funzioni di sopravvivenza in una distribuzione log-normale.
- Figura 13b (a destra): esempi di funzioni di distribuzione di probabilità in una distribuzione log-normale.
- Figura 14: dinamica di $h_0(t)$ al variare del parametro di inclinazione ρ .
- Figura 15a (a sinistra): rappresentazione grafica della verifica dell'assunzione di Weibull con bootstrapping dei dati sulla curva di sopravvivenza. (Casellas, 2006). la curva di sopravvivenza è riportata in grassetto, le curve degli intervalli di confidenza del calcolati sono raffigurate dalle linee più sottili.

Figura 15b (a destra): rappresentazione grafica dei *P-values* ottenuti con bootstrapping dei dati (Casellas, 2006). Il limite di significatività è rappresentato dalla linea tratteggiata.

Figura 16: rappresentazione grafica esemplificativa dell'effetto "Regressione per la dimensione dinamica dell'allevamento nel quadriennio successivo".

Figura 17: Curve di sopravvivenza di LPL e LV.

Figura 18: funzione del rischio di LPL e LV.

Figura 19: verifica dell'assunzione di Weibull per LV e LPL. È riportata la retta di regressione lineare e il valore di R^2 di approssimazione alla retta.

Figura 20: verifica dell'assunzione di Weibull stratificando LPL stratificata per ordine di parto.

Figura 21: Hazard ratio e relativi intervalli di confidenza al 95% dell'effetto età al primo parto.

Figura 22: Hazard ratio e relativi intervalli di confidenza al 95% dell'effetto ordine di parto.

Figura 23: funzione di sopravvivenza costruita sulle informazioni uncensored dell'ordine di parto e relativa variazione percentuale tra ordini di parto successivi.

Figura 24: Hazard ratio e relativi intervalli di confidenza al 95% dell'effetto scostamento percentuale dalla media EVM di allevamento/anno. I rischi relativi sono stati espressi su due diverse scale per apprezzarne meglio le variazioni tra classi successive, soprattutto per i livelli relativi a produzioni EVM superiori alla media. Figura 25: Hazard ratio e relativi intervalli di confidenza al 95% dell'effetto interazione tra dimensione annua dell'azienda e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo.

Figura 26: distribuzione di frequenza dei coefficienti di regressione stimati e dei relativi Hazard ratio dell'effetto fisso *time-independent* allevamento.

Figura 27: distribuzione di frequenza dei coefficienti di regressione stimati e dei relativi Hazard ratio dell'effetto casuale *time-dependent* allevamento*anno.

Figura 28: distribuzione di frequenza dei coefficienti di regressione stimati e dei relativi Hazard ratio dell'effetto casual *time-independent* del toro.

Figura 29: trend genetico espresso in *Hazard ratio* del carattere LPL.

Figura 30: distribuzione di frequenza dell'attendibilità degli EBV.

Figura 31: attendibilità media e massima del carattere longevità per anno di nascita del toro.

Figura 32: andamento dell'attendibilità dei tori con più di 20 figlie e con più di 20 figlie uncensored per anno di nascita del toro.

Figura 33: dinamica generale dell'attendibilità della stima in funzione del numero di figlie e della relativa percentuale di uncensored. Non sono valorizzate le combinazioni con R^2 inferiore al 50%.

1. INTRODUZIONE

1.1. La longevità nella razza Bruna

Questo studio nasce in seguito ad un'analisi del carattere longevità nella popolazione e dall'esigenza degli allevatori di avere strumenti sempre nuovi, efficaci e efficienti per poter migliorare le proprie vacche.

La longevità degli animali, nell'arco degli anni è cambiata. La figura 1 mostra il cambiamento del numero medio di lattazioni per soggetto all'interno delle principali razze bovine allevate in Italia (AIA, 2005). A partire dal 1997, per quanto riguarda la razza Bruna, emerge una riduzione di circa 0,05 lattazioni all'anno che, dal punto di vista pratico e considerando una lattazione standard a 305 gg, si quantifica in 15 giorni in meno all'anno (un mese al biennio). Questi valori sono la media del numero di lattazioni per anno di controllo, un primo empirico, ma significativo, approccio alla conoscenza del comportamento di questo carattere nelle popolazioni bovine. Da questa considerazione è nata l'esigenza di verificare il reale andamento della longevità delle bovine di razza Bruna e di studiare una possibile strategia volta a migliorare la permanenza in stalla delle vacche. Una prima verifica ha confermato, utilizzando dati più aggiornati, la statistica ottenuta con i dati AIA. Si è cercato in seguito di valutare la media del numero di lattazioni per anno di primo parto. Il limite di questo tipo di analisi è legato alla necessità di non considerare informazioni relative a soggetti ancora in vita e quindi ancora in produzione, per evitare la sottostima della reale longevità degli anni più recenti. Allo stesso modo le informazioni dei soggetti nati negli ultimi 3-4 anni e riformati precocemente, anche se sono indiscutibilmente gli unici dati certi, si riferiscono unicamente agli animali poco longevi sottostimando, anche in questo caso, la reale longevità della popolazione: i soggetti riformati che hanno partorito la prima volta nel 2005 non possono avere più di 3 lattazioni e non più di 2 per quelli del 2006.

Per tentare di ovviare a questa situazione e cercare di ottenere una statistica attendibile anche per gli anni più recenti, è stata utilizzata una media storica della "dinamica di riforma". Osservando la distribuzione delle informazioni archiviate nel passato (figura 2) si nota che fino al 1999 la numerosità dei soggetti con informazioni

complete (vacche riformate) rimane relativamente stabile. Allo stesso modo si nota un incremento progressivo delle informazioni incomplete (soggetti senza informazioni), dovuto a vacche ancora vive. Sulla base di queste informazioni e solo al fine di un'indagine esplorativa la percentuale eccedente il normale andamento dei soggetti per i quali non sono disponibili informazioni complete (circa 5000 all'anno) è stato distribuito proporzionalmente su ipotetiche lattazioni successive. A partire dal 1999 quindi, la correzione della dinamica dei soggetti riformati è stata forzata a mantenere un andamento paragonabile a quello ottenuto dall'analisi storica (figura 3). Sulla base di questa ipotetica distribuzione è stato possibile confrontare il numero di lattazioni medio ottenibile da vacche che hanno avuto il primo parto in anni successivi al '99 (figura 4). Da questo grafico è possibile osservare che le primipare del 1995, mediamente, raggiungevano 3,74 lattazioni, mentre vacche di primo parto del 2005 raggiungeranno mediamente 3,06 lattazioni con una perdita netta di 0,68 lattazioni in 10 anni, pari a 20 giorni di vita produttiva in meno all'anno.

In figura 5 infine è possibile osservare come si stia riducendo nell'arco degli anni la numerosità media di parti nell'arco degli anni (linea rossa), mentre relativamente ai dati "proiettati", gli istogrammi riportano la composizione percentuale di numero di parti raggiunti suddividendo gli animali per anno di primo parto. Partendo dalle primipare e salendo all'incrementare dell'ordine di parto, si nota che fino circa al 2000 il 50% dei soggetti raggiunge un ordine di parto superiore al quarto. Questa percentuale si riduce al 45% nel 2005. È possibile inoltre evidenziare la sostanziale riduzione delle vacche che superano il sesto ordine di parto, dal 25% relativo vacche che hanno partorito la prima volta nel 1985 rispetto al 17% delle primipare del 2005.

1.2. I caratteri funzionali

In un panorama internazionale caratterizzato dalla riduzione complessiva della numerosità dei bovini allevati e in regime di quote latte penalizzanti le sovrapproduzioni, al fine di garantire agli allevatori il mantenimento di un'attività imprenditoriale competitiva è fondamentale porsi degli obiettivi selettivi rivolti a caratteri che possano incidere sui costi di gestione riducendoli. È noto da anni come, un aumento della longevità delle bovine, comporti una significativa diminuzione dei

costi di allevamento delle manze dovuta ad una riduzione della quota di rimonta (Vicario, 2005). Stimando un costo medio della fase improduttiva da manza, al netto dei ricavi della vacca da riforma, di circa 500,00 €, risulta estremamente significativo il numero di lattazioni su cui questo costo viene diviso ($/2 = 250,00\text{€}$, $/3=165,00\text{€}$). Cambia sensibilmente la situazione per i bovini di razze a duplice attitudine, considerando che il valore delle vacche da riforma acquisisce un valore aggiunto per la massa muscolare sviluppata. Allevando animali più longevi si riduce inoltre il fabbisogno di manze per la rimonta. Nell'ultimo decennio il fabbisogno di manze per coprire la quota di rimonta, raramente viene soddisfatto con gli animali aziendali. Questo comporta che tutte le vitelle nate che raggiungono lo stato di manza sono necessarie per mantenere stabile la consistenza dell'allevamento. Una riduzione di questa quota permetterebbe di poter incrementare l'*intensità di selezione* rivolgendo l'attenzione verso i soggetti migliori determinando, a cascata un incremento del *progresso genetico*. Negli ultimi anni inoltre si sta affacciando nel campo della ricerca il settore della genetica molecolare, rivolto all'individuazione dei loci e dei geni responsabili dell'espressione dei caratteri. Questa metodologia sembrerebbe rappresentare il futuro del miglioramento genetico, anche se solo in seguito ad una riduzione dei costi operativi di laboratorio rendendola accessibile all'applicazione su intere popolazioni. La Bruna Italiana è tra una delle popolazioni bovine su cui si è sempre ricercata la strategia selettiva che permettesse un incremento del progresso genetico. Il programma di questa razza è stato tra i primi a adottare una rudimentale selezione assistita da marcatori; dal 1990 infatti seleziona i propri riproduttori per le varianti alleliche al locus della K-caseina e identifica, congiuntamente a rilevazioni in campo di figli malati, i soggetti portatori di patologie autosomiche recessive quali Weaver, Sindrome di malassorbimento, aracnomelia e la sindrome da dismielinizzazione del midollo spinale, note col termine di tare genetiche.

Congiuntamente alla longevità, altri caratteri sono indagati nel loro determinismo genetico. Considerando che animali "scarsamente funzionali" hanno anche una maggior probabilità di essere riformati i legami fenotipici di questi caratteri con la durata di un animale in allevamento sono evidentemente stretti. Neerhof et al (2000) con la presenza di mastiti, Sölkner e Petschina (1999) e Larroque e Ducrocq (2001) con i caratteri morfologici lineari, Caraviello et al (2004 e 2005) con il contenuto

di cellule somatiche nel latte e Lopez de Maturana et al (2007) con la facilità di parto, Gröhn et al (1998) con lo stato di salute, verificano la presenza di un effetto, dovuto da alcuni di questi caratteri in diverse razze bovine da latte, nel rischio di riforma introducendoli direttamente nel modello di analisi. Ognuno di essi è stato scelto in funzione di un presunto legame con la lunghezza della vita in stalla di un animale:

- Somatic Cell Score: in una situazione di mastiti ricorrenti l'allevatore preferisce riformare la vacca non appena possibile, onde evitare continue spese per terapie antibiotiche che, tra il resto, comportano l'esclusione dal conferimento del latte prodotto dalla vacca fino a ristabilimento delle condizioni sanitario e rispetto dei tempi di sospensione dei farmaci utilizzati. (Neerhof et al 2000)
- Velocità di mungitura e eiezione assistita: al fine di evitare la riforma di animali con problemi sanitari è importante che la mungitura avvenga in modo rapido e completo. Economicamente inoltre è vantaggioso riuscire a "sincronizzare" la durata della mungitura degli animali, evitando momenti di stress da costrizione in sala di mungitura, nonché un dispendioso prolungamento del tempo dedicato a questa manovra. Anche in questo caso, vacche problematiche, che richiedono somministrazioni frequenti di ossitocina, o che occupano la sala di mungitura per tempi considerati troppo lunghi, sono fortemente penalizzanti nella gestione razionalizzata di un'azienda al punto che un allevatore potrebbe essere incentivato ad anticiparne la riforma (Ral et al 1990).
- Fertilità: probabilmente è il carattere il cui determinismo genetico è più complesso e con il quale sono stati maggiormente ricercati i legami con la longevità (Strapack, 2005 - Lopez de Maturana, 2007). Attualmente, a tale scopo, l'associazione nazionale della razza Bruna stima l'indice del carattere BCS (Body Condition Score) relativo alla condizione corporea durante la lattazione (Edmonson, 1989) quale indicatore della fertilità degli animali.

La prima difficoltà nella stima di valori riproduttivi dei caratteri funzionali è sicuramente insita nella difficoltà di rilevazione e raccolta dei dati. Gli sforzi delle associazioni nazionali di allevatori bovini sono rivolte in questi ultimi anni proprio alla sensibilizzazione del settore nella raccolta accurata anche di queste informazioni: per la fertilità attraverso una registrazione puntuale delle fecondazioni, delle date di

parto, di asciutta, ecc., per longevità, sono di primaria importanza la corretta rilevazione della data e del motivo dell'eliminazione per avere a disposizione un dato fenotipico corretto e completo. Nella moderna zootecnica, questi caratteri rivestono ormai un ruolo fondamentale per il successo economico degli allevamenti (Sölkener, 2002).

1.3. Caratteri Morfologici

La valutazione morfologica viene fatta sulla singola bovina e consiste in una descrizione lineare del suo aspetto. Si quantifica in questo modo la misura biologica di ogni carattere o di gruppi di essi (punteggio finale e complessivi) esprimendo il grado di perfezione di ogni animale rispetto al modello ideale (figura 6). La loro importanza nell'analisi della longevità risiede in legami prima di tutto fenotipici (Forabosco, 2004) vincolanti lo stato sanitario delle vacche. Dal punto di vista selettivo sono state inoltre indagate anche le correlazioni genetiche con questi caratteri (Tarrés, 2006 - Pérez-Cabal e Alenda, 2002 - Pérez-Cabal et al., 2006)

Un altro aspetto rilevante legato alla rilevazione e stima degli indici genetici è legato al loro impiego: attualmente le nazioni che stimano dei valori riproduttivi per il longevità ai fini di uno scambio internazionale di informazioni, inviano la combinazione degli *estimated breeding values* (EBV's) ottenuti mediante analisi di sopravvivenza con quelli di alcuni caratteri morfologici o funzionali.

Assemblando tutte queste fonti di variabilità è possibile delineare un quadro generale delle cause di riforma volontaria, che verranno comunque illustrate in dettaglio nel primo contributo sperimentale in merito alla scelta dei fattori da includere nel modello quali fattori di rischio ambientali, verificando inoltre la significatività (Welgel, 2003) per la successiva stima dei BV e delle componenti di varianza.

La descrizione lineare è espressa su una scala di valori compresi tra 1 a 50 e rappresenta una stima indicata da un esperto di razza di ciascuna caratteristica dell'animale. Si precisa che, per difficoltà pratiche ed economiche, si tratta sempre di stime e non di misure con metro o nastro misuratore pur se vi sono parametri ben precisi da verificare con misure da effettuare a campione. Per alcuni caratteri

l'ottimale si trova in corrispondenza del valore intermedio degli estremi biologici, mentre per altri è rappresentato dall'estremo superiore.

Tabella 1.1: confronto tra l'attuale scheda di rilevazione dei tratti lineari e quella ottenuta con l'armonizzazione europea (A.N.A.R.B., 2007)

Carattere rilevato	Cambiamento del carattere	Raggruppamento		
		Attuale	Nuovo	
Altezza al garrese	altezza alla groppa	Capacità corporea (15%)	Struttura, forza e capacità produttiva (25%)	
Forza e vigore				
Profondità addominale		Caratteri da latte (15%)		
Angolosità	non più rilevato			
Linea dorsale		Struttura (15%)	Groppa (10%)	
Angolo della groppa				
Larghezza della groppa	Ampiezza degli ischi			
Lunghezza della groppa	riprendere le rilevazioni sospese nel 1996			
	Coxo-femorale	--		
Arti visti di lato		Arti e piedi (15%)	Arti e piedi (25%)	
Qualità dei garretti				
Forza delle pastoie				
Altezza di talloni				
	Forza anteriore mamm.	--		
Lungh. Attacco anteriore mammella		Apparato mammario (40%)	Apparato mammario (40%)	
Largh. Attacco posteriore mammella				
Altezza attacco posteriore mammella				
Profondità mammella				
Forza legamento centrale mammella				
Equilibrio nel piano della mammella	riprendere le rilevazioni sospese nel 1996			
Lunghezza dei capezzoli (anteriori e posteriori)				
Posizione dei capezzoli (anteriori e posteriori)	riprendere le rilevazioni sui posteriori sospese nel 1996			
	Diametro dei capezzoli			--

I singoli tratti lineari sono descritti in sintesi da 5 caratteri complessivi che a loro volta sono riuniti in un punteggio unico (Punteggio Finale) che esprime il giudizio complessivo dell'animale. Quest'ultimo così come i 5 complessivi è espresso su una scala che ha valori compresi tra il 65 e il 100.

Nel panorama europeo questo tipo di rilevazioni subirà, a partire dal 2008 un processo di armonizzazione internazionale ridistribuendo il peso di alcuni caratteri nei relativi complessivi e la loro rilevanza nel punteggio finale. La tabella 1.1 riporta la situazione attuale e le variazioni che saranno adottate nella scheda di rilevazione dei caratteri morfologici.

1.4. Espressione della longevità

Per dare un'unica definizione al carattere longevità è necessario identificare definire il carattere: “numero di giorni che intercorrono tra la nascita e eliminazione o la data di troncamento dell'informazione” (Smith e Quaas, 1984). Come per ogni carattere l'espressione è stata successivamente analizzata anche utilizzando un carattere discreto fino ad assumere anche la connotazione di carattere binario (Vollema, 1998). In particolare Vollema la suddivide in 4 categorie:

1. lunghezza della vita: sia misurata in modo continuo che discreto
 - 1.1. intervallo nascita-eliminazione
 - 1.2. intervallo data di primo parto- eliminazione
 - 1.3. quantità di latte prodotto nella carriera produttiva
 - 1.4. numero di giorni produttivi in tutta la carriera
 - 1.5. numero di lattazioni
2. presenza in allevamento (durabilità): carattere ad espressione binaria
 - 2.1. presenza / assenza ad intervalli prestabiliti a partire dalla nascita o dalla data di primo parto
 - 2.2. presenza / assenza all'inizio di una nuova lattazione
3. altre forme di espressione:
 - 3.1. numero di mesi produttivi a determinati intervalli di tempo
 - 3.2. probabilità di sopravvivenza da una lattazione alla successiva

4. longevità funzionale: qualora ciascuno di questi caratteri viene corretto per il livello produttivo, il carattere prende il suffisso di funzionale.

In generale la longevità può essere definita come “*l’abilità di una vacca a non essere riformata*” (Ducrocq, 1994). Ciascuna di queste forme manifesta pregi e difetti nella stima delle componenti di varianza genetica. Come per qualsiasi carattere, la continuità dell’espressione permette di apprezzare le differenze tra variazioni di minore entità, dell’ordine di mesi o giorni, d’altra parte necessita di verifiche e controlli nei dati d’entrata (editing) più accurati rispetto a espressioni relative a caratteri espressi sulla base di informazioni fenotipiche discrete.

Un limite all’approccio con modelli misti lineari classici risiede nel ritardo con il quale si rendono disponibili le informazioni relative all’eliminazione delle vacche (Van Raden, 1993): per avere un’informazione completa di lunghezza della vita, qualora sia analizzata con i classici modelli lineari misti e considerando l’intera carriera dell’animale, è necessario attenderne la riforma. Questa attesa si riflette in ambito della ricerca del determinismo genetico del carattere in un allungamento dell’intervallo di generazione con conseguente riduzione del progresso genetico ottenibile (Smith e Quaas, 1984). Van Raden (1993) riduce la perdita di informazioni non considerando la data di eliminazione come termine delle osservazioni, ma utilizza come variabile dipendente il numero di mesi produttivi all’età di 84 mesi (4 anni) escludendo quindi dall’analisi unicamente i soggetti di età inferiori a quella richiesta. Una rilevazione discreta della “durabilità” viene espressa come presenza/assenza della vacca in azienda a intervalli prestabiliti di tempo delimitati da “*time-point*” come ad esempio a 36, 72 mesi d’età oppure 12, 24, 36 mesi dal primo parto ecc...(Vollema, 1998) rendendo completa ogni informazione “troncata” all’interno di ciascun intervallo. Un soggetto che entra in un intervallo, ma non raggiunge il *time-point* finale fornisce un’informazione troncata in merito a questo intervallo anche se non è possibile determinare a priori di quanto la sua vita possa prolungarsi e eventualmente fornire un’informazione completa in merito allo stesso intervallo di tempo. Questo tipo di informazione non può essere introdotta nell’analisi con modelli lineari.

In questo studio si è scelto di analizzare la lunghezza della vita produttiva compresa tra la data di primo parto e la data di eliminazione. La scelta è stata orientata verso questo modo di espressione del carattere per diversi motivi:

- l'interesse al carattere non è rivolto ad un'analisi delle cause di eliminazione nel periodo improduttivo, ma unicamente al periodo successivo al primo parto.
- l'elevata incidenza della lunghezza del periodo antecedente il primo parto sull'intera vita di una vacca: ad esempio per una vacca riformata al termine della sua prima lattazione della durata di 10 mesi e che ha partorito a 24 mesi (età 34 mesi), ha solo un 30% della lunghezza della propria vita con informazioni "utili" riguardanti le cause di riforma volontaria (Ducrocq, 1994).
- l'età al primo parto riflette eventuali "problemi" di fertilità e stato sanitario relativi al periodo antecedente la prima lattazione. Questi sono fattori che incidono molto, come è stato illustrato nel capitolo "Caratteri funzionali", sulle cause di riforma volontaria (Neerhof et al, 2000).

Smith e Quaas (1984) sono tra i pionieri nell'applicazione della *failure time analysis* introducendo la possibilità di analizzare anche le informazioni incomplete definite col termine *censored* (Hald, 1949 - Kaplan e Meier, 1958). Con lo sviluppo sull'applicabilità della survival analysis nello studio della longevità (Ducrocq, 1994 – Louzada-Neto, 1998 – Yazdi, 2002 – Casellas, 2006) e con l'implementazione di software idonei a questo tipo di applicazioni (Ducrocq, 2000) è ora possibile indagare le cause di riforma volontaria determinando, nella variabilità residua rappresentante la quota di riforma involontaria legata alla componente genetica determinante il rischio di eliminazione (Ducrocq e Casella, 1996). Un'ulteriore possibilità di esprimere il valore genetico dei tori analizzando la lunghezza della vita produttiva è appunto il coefficiente di rischio relativo di riforma. Dato il legame esistente tra funzione di sopravvivenza e funzione di rischio, come verrà illustrato nel paragrafo relativo alla survival analysis, tale coefficiente esprime con valori inversamente proporzionali alla lunghezza della vita produttiva il valore riproduttivo dei tori per il carattere longevità (Cox, 1972 – Ducrocq e Casella, 1996). Il legame che permette di paragonare funzioni di distribuzione di sopravvivenza e di rischio, apre quindi alla possibilità, con opportune trasformazioni, di esprimere l'EBV dei tori contemporaneamente su diverse scale: giorni/mesi/anni di vita/vita produttiva, numero di lattazioni, rischio di eliminazione e, per le sue proprietà nel modello di analisi, il suo logaritmo (coefficiente di regressione).

1.5. Longevità diretta e indiretta

La principale e sostanziale differenza esistente tra un piano di miglioramento genetico basato sulla selezione diretta per un carattere, rispetto all'impiego di caratteri ad esso correlati è rappresentata dalla riduzione del progresso genetico annuo ottenibile dovuto all'introduzione del coefficiente di correlazione genetica tra i due (o più) caratteri:

$$\Delta G_y = \frac{r_{gg} * i * (\sigma_{xGA} * corr_{xy})}{N}$$

dove ΔG_y è il progresso genetico annuo del carattere y , r_{gg} è l'accuratezza della stima, i l'intensità di selezione, σ_{xGA} e $corr_{xy}$ rispettivamente la varianza genetica additiva del carattere secondario e la sua correlazione con il carattere obiettivo di selezione e N l'intervallo di generazione. Tanto maggiore è la correlazione tra i due caratteri, tanto più simili saranno i valori di ΔG_y tra l'impiego della selezione diretta e quella indiretta (Bittante, 1990).

Come accennato precedentemente, l'impiego di modelli misti lineari con stime delle componenti di varianza genetica additiva tramite REML BLUP animal model, per la longevità diretta presenta alcuni limiti. Una figlia ancora in vita, fino a quando non viene riformata non contribuirà alla valutazione genetica del padre e rappresentando un forte vincolo nell'attendibilità del valore riproduttivo soprattutto per i tori giovani che per la maggior parte sono i soggetti che nel piano di selezione genetica sono rappresentati dai tori in prova di progenie. Le problematiche nelle quali è possibile imbattersi sono la mancanza di informazioni precoci e la sottostima del loro valore genetico (o sovrastima nel caso del rischio di riforma) in quanto le prime informazioni uncensored disponibili sono relative alle figlie *minus-varianti*. L'approccio con modelli misti lineari è stato studiato in passato sulla razza Bruna e sarà illustrato nel paragrafo "*Progresso genetico della razza Bruna*" di questo capitolo. L'obiettivo perseguito dal precedente studio era limitato ad un'indagine esplorativa delle componenti di varianza genetica additiva del carattere longevità funzionale e le relative correlazioni genetiche con i caratteri morfologici al fine di poter calcolare un indice composto, di derivazione lineare appunto, che potesse orientare gli allevatori ed

i centri genetici verso l'impiego di riproduttori miglioratori anche per questo carattere (Cassandro, 1998).

L'applicazione della survival analysis, inizialmente applicata al settore zootecnico come *failure time analysis* da parte di Smith e Quaas (1984) ha dato il via ad un nuovo modo di analizzare le informazioni fenotipiche disponibili relative a dati di sopravvivenza, basandosi sui rapporti intercorrenti tra le informazioni ottenibili dell'intera popolazione, dai soggetti riformati *uncensored* e dai soggetti con informazioni troncate denominate *censored* (Ducrocq, 2000). Un limite, in fase di studio, nella stima del determinismo genetico del carattere longevità è la difficoltà nell'applicazione di un animal model, per la stima contemporanea del valore riproduttivo dell'intera popolazione. A questo proposito l'interesse internazionale si sta orientando verso lo sviluppo di metodologie per la stima approssimata di BV. Ducrocq, (2001) ha ottenuto correlazioni tra EBV superiori al 99,9% per i maschi e del 99,8% per le vacche tra stime ottenute con la survival analysis utilizzando un modello a regressione di Weibull e un modello misto lineare che utilizza il BLUP animal model.

1.6. La situazione internazionale

In un contesto più ampio, molti sono i Paesi che si sono apprestati a stimare le componenti di varianza genetico additiva per il carattere longevità e per gli altri caratteri funzionali.

La tabella 1.2 riporta le nazioni Interbull che calcolano gli indici genetici per il carattere longevità, la razza, il metodo utilizzato e l'ereditabilità stimata del carattere (Interbull, 2007).

Tabella 1.2: Informazioni riguardo le valutazioni nazionali dei paesi che aderiscono a Interbull per il carattere longevità diretta.

Nazione	Razza/e	h ²	Metodo	Espressione
Australia	RDC, GUE, JER, HOL	2,50 %	AM-RP	BV+
Belgio	HOL	10,60 %	AM-RR	BV+
Canada	RDC, BSW, GUE, JER, HOL	9,75 %	AM-MT	TA+
Danimarca	RDC, RHOL, JER	11,70 %	SM-SA	BV+
Finlandia	RDC, HOL	12,00 %	SM-SA	TA+
Francia	HOL	10,80 %	SM-SA	BV+
	MON	13,00 %	SM-SA	BV+
	BSW, SIM, + razze minori	14,00%	SM-SA	BV+
Germania	RDC, HOL	16,60 %	SM-SA	BV-
Ger. E Austria	BSW	12,00 %	SM-SA	BV+
Ungheria	HOL	5,00 %	SM-SA	BV+
Irlanda	HOL	1,60 %	AM-MT	TA+
Israele	HOL	11,00 %	AM-ST	TA-
Italia	HOL	9,70 %	SM-SA	TA+
Nuova Zelanda	RDC, BSW, JER, GUE, HOL	5,50 %	AM-MT	BV+
Spagna	HOL	11,50 %	SM-SA	TA+
Svezia	RDC, HOL	8,00 %	SM-MT	BV+
Svizzera	HOL	13,85 %	SM-SA	BV+
	BSW	18,10 %	SM-SA	BV+
Paesi bassi	HOL, SIM, JER, BSW	10,00 %	SM-SA	BV+
Regno unito	RDC, HOL, GUE, JER	6,00 %	AM-MT	TA+
U.S.A.	RDC, HOL, BSW, JER, GUE	8,00 %	AM-ST	TA+

Legenda:

Razza	
RDC	Red Dairy Cow
HOL	Frisona
BSW	Bruna
JER	Jersey
RHOL	Red Holstein
MON	Montbéliarde
GUE	Guernsey
SIM	Simmental

Espressione	
BV +/-	Breeding value di merito/ demerito
TA +/-	Trasmitting Ability di merito /demerito

Metodo	
Prima parte	
SM	Sire model
AM	Animal model
Seconda parte	
ST	Single trait
MT	Multiple trait
RP	Repeatable trait
RR	Random regression
SA	Survival analysis

Le nazioni che utilizzano la survival analysis per la stima dei BV procede a combinarlo con altri caratteri ad esso geneticamente correlati, esprimendolo il risultato in una longevità definita, appunto, combinata.

La combinazione con altri caratteri permette di ovviare ad alcuni limiti dell'analisi di sopravvivenza del carattere longevità: bassa ereditabilità, caratteristica comune a tutti i caratteri funzionali, disponibilità tardiva, per i motivi già evidenziati precedentemente (Smith, 1983). In particolare la preponderante quantità di informazioni di tipo *censored* (incomplete) dei tori giovani si riflette su un'attendibilità ridotta dell'indice genetico (Yazdi, 2002). La combinazione con altre fonti di informazione più precoci, permette di incrementare l'attendibilità dell'indice composto con un notevole risparmio di tempo nel conoscere il valore di un toro, sia a fini selettivi sia commerciali (Druet, 1999). La tabella 1.3 riporta il carattere analizzato dalle diverse nazioni che calcolano gli EBV nella popolazione di Bruna e gli effetti inclusi nei modelli dai vari paesi. Le informazioni sono ottenute dalle schede (Form GE) inviate dalle nazioni a Interbull.

Tabella 1.3: modelli utilizzati a livello internazionale per il calcolo degli EBV per il carattere longevità

	Carattere	Effetti fissi		Effetti casuali	
		Time indep	Time dep.	Time indep	Time dep.
Canada	Surv3	HY + regEP + regEP ² + regGr + regGr ² + regGr ³ + regPr + regPr ² + regPr ³ + regPr*(R*H*S)		R*H*S + Animale	
Francia	LPL	EPP	$\Delta Hs + Y*R + \Delta 305La*OP*S$ L + $\Delta 305Gr,Pr*S$ L	Animale	H*Y
Germania e Austria	LPL	EPP	R*Y*S + H* $\Delta 305Gr,Pr$ $\Delta Hs + Alp$	Animale	H*Y*S
Nuova Zelanda	Surv4+ 1OP+ 2OP	--	--	--	--
Svizzera	LPL	EPP	OP*gg1P*SL + OP + Alp + $\Delta 305La,Gr,Pr$	Animale	H*Y*S
Paesi bassi	LPL strat.=1-2- >3OP	EPP + eterosi + ricomb.	OP + OP*Y*S + $\Delta Hs + \Delta 305La,Gr,Pr$	Animale	H*Y*S
U.S.A.	LPL	GGr + inbreeding		H*animale	
	*LPL strat.=area geografica	regEPP	OP*SL + H*Y* $\Delta 305Gr,Pr$	Animale	H*Y*S

*Caraviello et al. 2004

Surv_n: sopravvivenza da un parto al successivo fino al n^{esimo} parto

nOP: n^{esimo} ordine di parto / **LPL** – lunghezza della vita produttiva

EPP – età al primo parto / **EP** – età al parto / **gg1P** giorni dal primo parto

reg – inserimento dell'effetto come covariata

H – allevamento / Δ **Hs** – variazione della dimensione dell'allevamento

Y – anno / **S** – stagione / **R** – regione

OP – ordine di parto / **SL** – stadio di lattazione

Δ **305** – variazione su lattazione corretta a 305 giorni di:

La – latte / **Gr** – grasso / **Pr** – proteina

eterosi / **ricombinazione** – effetto in classi crescenti di eterosi / ricombinazione

GGr – gestione del gruppo (non meglio definito) / **giudizio** – opinione dell'allevatore

La tabella 1.4 invece riporta, dove riportati, i caratteri correlati utilizzati nel calcolo della longevità combinata.

Tabella 1.4: caratteri con i quali viene combinata la longevità diretta per lo scambio internazionale degli indici genetici

	Produttivi	Morfologici	Funzionali
Canada	--	--	--
Francia	latte, grasso, proteina	--	SCS, fertilità
Ger. E Austria	--	A&P, car_mamm, ADL, Leg, PrM LuC	--
Nuova Zelanda	1OP(Pr)	1OP(A&P, car_lat, car_mamm)	Surv(1-2°; 1-3°; 1°-4°) 1OP(BCS, giudizio, VMU, SCS) 2OP:(CE*S)
Svizzera	--	--	SCS, % di non ritorno, gg dal I servizio
Paesi bassi	--	AnG, PrA, PrM, A&P	SCS, interparto
U.S.A.	latte ($r_{PL-La}=0,08$) grasso ($r_{PL-Gr}=0,08$) proteina ($r_{PL-Pr}=0,10$)	car_mamm ($r_{PL-Udder}=0,30$) A&P ($r_{PL-F\&L}=0,19$) Taglia ($r_{PL-Size}=-0,16$).	SCS ($r_{PL-SCS}=-0,38$) tasso di gravidanza ($r_{PL-DPR}=0,51$) facilità di parto toro ($r_{PL-SCE}=-0,25$) facilità di parto vacca ($r_{PL-DCE}=-0,30$)

BCS – Body Condition Score / **SCS** – Somatic Cell Score

VMU – velocità di mungitura / **CE** – facilità di parto

A&P – arti e piedi / car_mamm – complessivo mammella / **ADL** – arti visti di lato

Leg – Legamento sospensorio della mammella / **PrM** – profondità della mammella

LuC – lunghezza dei capezzoli / **car_lat** – caratteri lattiferi / **AnG** – angolo della groppa
PrA – profondità addominale / **Taglia** – dimensione complessiva dell'animale

1.7. Progresso genetico della razza Bruna

La scelta di A.N.A.R.B. di inserire l'indice longevità tra i propri obiettivi di selezione è il risultato delle indagini svolte che hanno evidenziato la crescente richiesta da parte del settore di ridurre sensibilmente il peso della componente produttiva e puntare alla riduzione dei costi di gestione. Questo processo è attuabile analizzando le voci dei costi di un'azienda e individuando i caratteri sui quali è possibile agire efficacemente attraverso la selezione genetica. L'assenza di selezione per determinati caratteri è rilevabile dal punto di vista selettivo attraverso l'andamento del trend genetico. La figura 9b mostra chiaramente un'assenza di trend genetico del carattere longevità fino alla fine degli anni 90, periodo in cui questo carattere è stato inserito tra gli obiettivi di selezione e ha acquisito una certa rilevanza anche nell'Indice Totale Economico (I.T.E.) del quale si riporta il trend nel periodo 1989-2000 in figura 9a. Come illustrato precedentemente l'analisi genetica quantitativa dei caratteri funzionali ed in particolare della longevità è oggetto di studio da molti anni (Smith, 1984 - Ducrocq, 1987, 1996 e 1997; Cassandro et al 1998; Vanraden, 1995; Vukasinovic, 1997; Louzada-Neto, 1998) ma solo dopo il 2000 è iniziato un lavoro strutturato e orientato al perfezionamento della metodologia rivolta la stima dei parametri genetici e dei valori riproduttivi anche per il carattere sopravvivenza (Ducrocq, 2001; Yazdi, 2002; Roxsötrm, 2003). A partire dal 2003 poi, con l'introduzione del regime a quote latte, l'importanza relativa, anche a livello internazionale, del livello produttivo è andata progressivamente riducendosi. Va sottolineato che mediamente oltre il 50% del peso relativo dei vari indici di selezione internazionali è attribuito ancora al livello produttivo. Dal punto di vista scientifico si è assistito ad un incrementando della quantità di pubblicazioni relative allo studio del carattere longevità (Caraviello, 2004; Forabosco, 2005; Gonzales-Recio, 2006; Terawaki, 2006; Damgaard, 2006; Phocas, 2006; Chirinos, 2006), alle sue correlazioni con i caratteri produttivi e morfologici (Sölkner, 1999 - Larroque, 2001; Tarrés, 2006) e anche con gli altri caratteri funzionali (Neerhof, 2000; Leclerc, 2005; Druet). Con lo sviluppo della genetica molecolare sono state sviluppate anche prove di selezione assistita da marcatori (Moreno, 2005).

In merito al progresso genetico delle razze bovine, in Italia il compito monitorare e migliorare le caratteristiche genetiche della popolazione è affidato agli uffici centrali dei libri genealogici delle associazioni nazionali allevatori delle specifiche razze. La tabella 1.5 riporta un breve excursus sull'evoluzione degli obiettivi di selezione perseguiti per la razza Bruna.

Tabella 1.5: Evoluzione degli obiettivi di selezione nella razza Bruna.	
Anno / periodo	Obiettivo
Primi '900	Latte Carne Lavoro
1930	Latte Carne (anno di fondazione del Libro genealogico)
1968	Latte
1970	Latte, proteina (kg)
1990	Indice di selezione composto (I.T.E.): proteina (kg e %)
1995	Longevità funzionale (indiretta)
2005	Mungibilità e S.C.S.
2006	Forza delle pastoie e Punteggio finale

I risultati ottenuti nel corso degli anni sono verificabili dal trend genetico dell'I.T.E. riportato in figura 8a. La composizione dell'indice totale economico ha subito non pochi cambiamenti, oltre che dal punto di vista dei caratteri che lo compongono anche per quanto riguarda i pesi relativi di ciascun carattere al suo interno. Attualmente, gli indici morfologici inseriti nel 2006 hanno ridotto ulteriormente la quota relativa agli indici produttivi, già ridimensionata con l'inserimento alla fine degli anni '90 della longevità indiretta e nel 2005 dai caratteri funzionali mungibilità e Somatic Cell Score. Nell'ultima variazione dei pesi dei caratteri inseriti nell'indice di selezione la longevità ha visto un incremento dal 12 al 18% (Figura 8b).

La continua selezione diretta sui kg di proteine (59% dell'I.T.E.) ed indiretta sui kg di latte (è risaputo l'elevato legame genetico tra i due caratteri) si evidenzia nei rispettivi trend genetici (figura 9c e 9d).

1.8. Survival analysis

1.8.1. Definizione

L'analisi di sopravvivenza chiamata anche analisi storica degli eventi o analisi della durata o ancora analisi di transizione (Buis, 2006) ha l'obiettivo di indagare il determinismo di un evento, definito *failure*, che definisce l'ampiezza di un periodo di osservazione. L'estensione di questo intervallo di tempo definisce la variabile dipendente oggetto di studio (Ducrocq, 2000a). Analizzando il carattere longevità nella specie bovina, l'evento ricercato non è la morte spontanea dell'animale, ma la scelta di riforma fatta dall'allevatore. Difficilmente infatti una vacca giunge al termine della propria vita biologica per vecchiaia, ma generalmente viene avviata al macello in seguito ad una riduzione delle proprie performance produttive e/o funzionali (Vollema, 1998). L'intervallo di tempo è compreso tra uno *start-point* o momento di inizio osservazioni che può essere rappresentato dalla data di nascita (Smith e Quaas, 1984) o, molto più frequentemente, di primo parto (Ducrocq, 1994 – VanRaden, 1995 – Chirinos, 2006) e da un *end-point* o momento in cui termina l'osservazione. Questo valore assume un diverso significato nel caso in cui l'evento si manifesti (*failure time*) rispetto a un'informazione incompleta che termina per mancanza di informazioni successive ad un determinato momento (*sensor-time*). Al fine di focalizzare l'analisi sul determinismo dell'evento, è assunta sempre più frequentemente come ampiezza dell'intervallo di osservazione l'intero periodo che caratterizza la lunghezza della vita produttiva (LPL) di una vacca detratto dagli intervalli privi di informazioni come già fecero Smith e Quaas (1984) e, più recentemente Chirinos (2006). Analogamente è possibile perseguire il determinismo dell'evento riforma descrivendo la variabile dipendente con valori discreti quali l'ordine di parto o gli anni (Yazdi, XXXX) e, molto più raramente, binari come la presenza/assenza all'inizio di una nuova lattazione o a intervalli di tempo arbitrariamente scelti (Yazdi, 2002). La figura 9 riporta queste tipologie di rilevazione di intervalli di tempo che sono riassunti nella tabella 1.6, dove $n^{\circ} failure$ cumulativi identifica il numero di soggetti che ha manifestato l'evento fino a al momento (t_1, t_2, t_3, \dots) in cui si verifica l'evento o ad ogni *time-point* definito arbitrariamente (a, b, c, \dots); $n^{\circ} censored$ cumulativo identifica il numero di soggetti per i quali si perdono informazioni aggiuntive dopo un preciso momento (cfr. Censoring). Nel caso di rilevazione discreta a time-point è possibile vedere che si verifica un

ritardo nell'identificazione degli eventi e che nel caso in cui il tempo $T=t$ non rappresentasse un *time-point* (-) di rilevazione di presenza/assenza, 2 sarebbero i soggetti con informazioni troncate e 2 quelli con evento (*failure*) identificato.

Tabella 1.6: esempio di dati di sopravvivenza ottenuti dalla figura 9

<i>Failure-time</i>	<i>failure cumulativi</i>		N° <i>cens. cumulativi</i>	<i>Time-point</i>	<i>failure cumulativi</i>		N° <i>censored cumulativi</i>
	N°	%			N°	%	
t1	1	0,25	0	a	0	0,00	0
t2	2	0,50	0	b	1	0,25	0
t3	3	0,75	0	c	2	0,50	0
T	3	0,75	1	t / -	3 / 2	0,75/0,50	1 / 2

Il rapporto tra il numero di *failure* cumulativo e la numerosità totale dei soggetti presenti nella popolazione definisce ad ogni istante (t1, t2 e t3 oppure a, b e c) la proporzione dei soggetti riformati, mentre il suo complementare ad 1 (figura 10) esprime una proporzione o funzione di sopravvivenza.

L'approccio più immediato nell'analisi dei dati di sopravvivenza è l'analisi della *life-table*. In questo approccio si lascia che "i dati parlino di loro stessi" (Buis, 2006). Con questo tipo di approccio la funzione impiegata per descrivere la dinamica nella popolazione dell'evento è la stessa funzione di sopravvivenza. Questa tecnica è tra le prime sviluppate (Berkson e Gage, 1950; Cutler e Ederer, 1958; Gehan, 1969) e può essere considerata come una tabella di distribuzione di frequenza specifica per dati di sopravvivenza. La procedura prevede di suddividere arbitrariamente il periodo di osservazione in intervalli di tempo, al termine dei quali è possibile calcolare le proporzioni tra i soggetti con *failure-time*, soggetti *censored* e il resto della popolazione, in particolare è possibile definire:

- numero dei soggetti a rischio: ottenuto per differenza tra il numero dei soggetti entranti nell'intervallo e la metà dei soggetti *censored*.
- proporzione di evento: ottenuto dal rapporto, entro intervallo, tra il numero di *failure* ed il numero di soggetti a rischio.
- proporzione di sopravvivenza: è il complementare della proporzione di evento.

- proporzione cumulativa di sopravvivenza (funzione di sopravvivenza): è la sommatoria delle proporzioni tra i soggetti sopravvissuti alla fine dei rispettivi intervalli e i soggetti totali; assumendo l'indipendenza della sopravvivenza tra intervalli, la si ottiene moltiplicando la probabilità di sopravvivenza di un intervallo per tutti quelli temporalmente antecedenti. La funzione risultante è definita appunto funzione di sopravvivenza o *survivorship*. Un esempio, riferito ai dati in tabella 1.6 , riportato in grafico è rappresentato dalla doppia linea tratteggiata di figura 10.
- densità di probabilità: è la stima della probabilità di ogni intervallo di tempo che si verifichi l'evento e si ottiene rapportando la differenza tra la proporzione cumulativa di sopravvivenza stimata all'inizio e alla fine dell' *i*esimo intervallo, rispettivamente P_i e P_{i+1} , all'ampiezza dell'intervallo h_i :

$$F_i = (P_i - P_{i+1}) / h_i$$

- hazard rate (Barlow, 1963): definito come la probabilità, per unità di tempo, che un soggetto verifichi l'evento nell'intervallo. In particolare il calcolo si esegue dividendo il numero di eventi verificatisi nell'intervallo di tempo per il numero medio dei soggetti che non verificano l'evento a metà del tempo dell'intervallo stesso:

$$HR_i = failure_i / n_{i/2}$$

- tempo mediano di sopravvivenza: rappresenta il valore di tempo al quale la funzione cumulativa di sopravvivenza è pari al 50%. Allo stesso modo è possibile determinare valori per altri percentili (25%-75%...). Vista la presenza di informazioni *censored*, il *time-point* del cinquantesimo percentile della funzione cumulativa di sopravvivenza non corrisponde obbligatoriamente al *time-point* in cui è sopravvissuto il 50% della popolazione.

1.8.1.1. Analisi di sopravvivenza: caratteristiche specifiche

La tipologia di dati che si analizzano (*survival data*) in un'analisi di sopravvivenza è strettamente vincolato allo scorrere del tempo (Cox, 1972). A differenza delle informazioni raccolte per un'analisi con i classici modelli lineari il susseguirsi di eventi determina una variazione nei rapporti sulla numerosità della popolazione, andando a descrivere una dinamica che non è apprezzabile con i modelli

misti lineari. È contemplata anche la presenza di informazioni “troncate” o *censored*, le cui proprietà sono illustrate nel prossimo paragrafo. Un altro importante motivo (non strettamente vincolante la scelta di questo tipo di analisi) è la mancanza, generalmente, di normalità nella distribuzione di frequenza dei dati di sopravvivenza, violando una delle assunzioni vincolanti nell’approccio diretto con metodi di analisi della varianza basati su modelli misti lineari (Ducrocq, 1994). Ciò che caratterizza quindi l’analisi storica degli eventi è il cambio dei rapporti di numerosità della popolazione nel tempo, nonché la sua adattabilità ad altre tipologie di dati come potrebbe essere l’analisi di alcuni caratteri legati alla fertilità (Schneider, 2005).

Come verrà dettagliatamente illustrato in seguito, la stretta relazione tra funzione di sopravvivenza e funzione di rischio, e quindi anche delle relative funzioni di distribuzione cumulativa, permettono di accomunare i modelli misti lineari ed i modelli di analisi di sopravvivenza, seppur con modalità diverse, nella stima della distribuzione marginale per la determinazione delle componenti di varianza, aprendo anche in questo settore la possibilità di indagare il determinismo genetico intrinseco nei dati di sopravvivenza (Cox, 1972). Scegliere la survival analysis è vincolante quando si rende necessario l’utilizzo di dati incompleti, che altrimenti non sarebbe possibile analizzare se non con approssimazioni artificiose nella definizione del carattere oggetto di studio, per verificarne il determinismo nella loro manifestazione (Comin, 2005).

1.8.1.2. Censoring

L’uso di questa metodologia si è sviluppato inizialmente nel settore medico-biologico, acquisendo in seguito importanza anche nelle scienze sociali economiche, ingegneristiche e zootecniche (Statsoft, 2003). La scelta dell’evento che determina l’ampiezza dell’intervallo di sopravvivenza è determinante nella discriminazione dell’informatività di una rilevazione. Riprendendo l’esempio di figura 9, scegliendo di analizzare un evento diverso dalla riforma, quale ad esempio la manifestazione di un caso clinico di mastite e presupponendo che solo l’ultimo soggetto manifesti un episodio clinico in corrispondenza di un tempo $t_{2a} \leq t$ la definizione dei dati di sopravvivenza cambierebbe drasticamente, identificando ai tempi t_1 t_2 e t_3 non più la manifestazione dell’evento, ma la perdita di informazioni successive alla riforma, non potendo escludere che dopo tale momento, qualora la scelta di eliminare il capo sia di

tipo volontario, non si sarebbe verificato un caso di mastite clinica. Secondo questo punto di vista, i primi tre soggetti acquisiscono la codifica di dato troncato, definito comunemente *censored* (Hald, 1949). La loro informatività è parziale, e vincolata all'intervallo di tempo nel quale erano presenti nella popolazione. Ipotizzando che l'evento mastite si verificasse a un tempo t_{2a} compreso tra t_2 e t_3 la tabella dei dati di sopravvivenza potrebbe essere descritta (solo in merito ad una rilevazione continua delle informazioni) come in tabella 7, invertendo la proporzione tra informazioni complete (o *uncensored*) e incomplete (o *censored*).

Tabella 1.7: esempio di dati di sopravvivenza: cambiando il tipo di evento cambia la definizione di un dato *censored*.

<i>Failure-time</i>	N° <i>failure cumulativi</i>		N° <i>censored cumulativi</i>
t1	0	0,00	1
t2	0	0,00	2
T_{2a}	1	0,25	2
t3	1	0,25	3
t	1	0,25	3

I dati *censored* possono essere di tre tipologie: tipo I quando il vincolo, come nell'esempio appena riportato, è il tempo ovvero stabilendo a priori una lunghezza massima del periodo di osservazione ci saranno soggetti che manifesteranno l'evento quantificando un'ampiezza della sopravvivenza corrispondente al *failure-time* (T_i). Un di una certa proporzione di osservazioni al termine del periodo di osservazioni invece non avrà manifestato l'evento. In questo caso viene attribuito loro un *censored-time* (C_i) pari all'ampiezza dell'intero intervallo di osservazione.; tipo II quando il vincolo viene posto sul numero, o la proporzione, di *failure-time*, ad esempio si raggiunge il termine delle osservazioni quando il 50% delle osservazioni ha a disposizione un *failure-time* (T_i). I valori di C_i dei restanti record sono troncati in corrispondenza del valore di tempo (C_i) pari all'ultima osservazione assegnando loro la codifica *censored*. La terza tipologia è definita *random censoring*: il periodo di osservazione è prefissato, ma i soggetti vi entrano a tempi diversi e vengono distinti per ogni i^{esimo} soggetto *failure-time* (T_i) rilevato al momento della manifestazione dell'evento dai *censored-*

time (C_i) identificati dal momento oltre il quale non è possibile estendere il periodo di osservazione per il soggetto (Hougaard, 1999).

Un'altra caratteristica dei dati censored è l'estremo temporale al quale l'informazione è assente. È possibile infatti che oltre ad avere assenza di informazione all'estremo temporale maggiore (*right-censored*), non siano disponibili nel periodo antecedente l'inizio delle osservazioni definendo quindi un estremo temporale minore troncato (*left-censored*). Ulteriore possibilità, dipendente dal tipo di dato rilevato, è la possibilità di avere dati con *censoring* singolo o multiplo. Nel caso della rilevazione della sopravvivenza in seguito ad una terapia, di pazienti ospedalizzati, la dimissione rappresenta il punto di troncamento di informazioni. Qualora un soggetto subisce ripetuti ricoveri, ogni evento "dimissione" identifica un punto di *censoring* (Hougaard, 1999).

1.8.2. Distribuzioni

L'analisi di sopravvivenza quindi indaga congiuntamente la *distribuzione cumulativa di probabilità* che ad ogni istante del periodo di osservazione si verifichi l'evento ricercato e il suo valore complementare a 1 ovvero la *distribuzione di probabilità di sopravvivenza*. In funzione del tipo di informazioni raccolte è possibile quindi descrivere funzioni di sopravvivenza continue (minuti, giorni...) o discrete (anni, ordini di parto...). Entrambe sono definite matematicamente rispettivamente dalla *funzione di distribuzione cumulativa di probabilità* $F(t)$:

$$F(t) = \text{Prob}[T < t]$$

la quale esprime la probabilità di sopravvivenza o "durabilità" ad ogni istante T nell'intervallo di tempo analizzato. La funzione di sopravvivenza $S(t)$ rappresenta invece la funzione complementare a 1 della *funzione di distribuzione cumulativa di probabilità*:

$$S(t) = \text{Prob}[T \geq t] = 1 - F(t)$$

Maggiore è la probabilità che un evento si verifichi, minore sarà la probabilità di sopravvivenza. Al pari della funzione della distribuzione normale o Gaussiana,

derivando la funzione di distribuzione cumulativa di probabilità è possibile ottenere informazioni in merito alla *densità di distribuzione* $f(t)$ dei dati analizzati:

$$S'(t) = f(t) = \lim_{\partial t \rightarrow 0} \frac{\text{Prob}[t \leq T < t + \partial t]}{\partial t} = \frac{dF(t)}{dt}$$

L'analisi della *life-table* fornisce indicazioni sommarie sulla dinamica dell'evento nel tempo e si riferisce ad intervalli di tempo definiti arbitrariamente.

Al fine di ottenere informazioni predittive con l'analisi di sopravvivenza è possibile confrontare la funzione densità di distribuzione $f(t)$ osservata con delle funzioni matematiche note utilizzate nella descrizione della dinamica di eventi nel tempo. Nella survival analysis le funzioni di sopravvivenza e le relative funzioni di densità di probabilità maggiormente utilizzate per definire la distribuzione empirica di dati di sopravvivenza sono: la distribuzione esponenziale, la distribuzione di Weibull e la distribuzione log-normale (Ducrocq, 1994).

1.8.2.1. Distribuzione esponenziale

La distribuzione esponenziale è tra le più semplici distribuzioni di probabilità utilizzate ed è caratterizzata da un rapporto tra *failure-time* costante nel tempo espresso nella funzione di densità di distribuzione $f(t)$ dal coefficiente ν , valore assunto dalla funzione al tempo $t=0$. Graficamente è possibile rappresentare la funzione di distribuzione della densità di probabilità come illustrato in figura 11b. Parallelamente la relativa funzione di sopravvivenza $S(t)$ ha un andamento decrescente come mostra la figura 11a. È possibile definire la funzioni di sopravvivenza e la sua derivata prima come segue:

$$S(t) = \exp^{-\nu t} \quad \text{e} \quad f(t) = \nu \exp^{-\nu t}$$

Le caratteristiche di questa funzione di distribuzione sono:

- è una funzione continua
- assume valore $f(t)=\nu$ al tempo $t=0$
- il valore ν definisce quanto velocemente la funzione $S(t)$ tende a 0
- ν deve assumere un valore strettamente maggiore di 0

1.8.2.2. Distribuzione di Weibull

È sicuramente una delle distribuzioni più usate nell'ambito dell'analisi di sopravvivenza (Ducrocq, 1994). La funzione dalla quale deriva questo tipo di modello è definita da 2 parametri λ e ρ , il primo definisce l'ampiezza della distribuzione, il secondo varia il gradiente. Il modello di Weibull viene ampiamente utilizzato per la sua elevata flessibilità. In particolar modo è possibile individuare 5 precisi momenti in relazione al valore assunto da ρ (figura 11b):

- $0 \leq \rho \leq 1$: la funzione di distribuzione di probabilità è monotona decrescente con doppio asintoto: per $t \rightarrow 0$ $f(t) \rightarrow \infty$ e per $t \rightarrow \infty$ $f(t) \rightarrow 0$;
- $\rho = 1$: la funzione di distribuzione di probabilità assume una struttura assimilabile ad una funzione di tipo esponenziale.

Con valori di ρ superiori a 1 la distribuzione assume un andamento a tendenza normale con le seguenti caratteristiche:

- $1 < \rho \leq 2.6$: funzione con skewness positiva (coda a destra). Con $\rho = 2$ la funzione coincide con un'altra funzione di distribuzione denominata distribuzione di Rayleigh.
- $2.6 < \rho < 3.7$: curva assimilabile ad una distribuzione di tipo normale con skewness nulla.
- $\rho \geq 3.7$: funzione con skewness negativa (coda a sinistra).

La funzione di sopravvivenza (figura 11a) e di densità di distribuzione (figura 11b) utilizzando la regressione di Weibull sono definite come segue:

$$S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^\rho\right)$$

$$f(t) = \frac{\rho}{\lambda} \left(\frac{t}{\lambda}\right)^{\rho-1} \exp\left(-\left(\frac{t}{\lambda}\right)^\rho\right) \quad \text{con } \rho \neq 1 \quad \text{e} \quad f(t) = \frac{1}{\lambda} \exp\left(-\frac{t}{\lambda}\right) \quad \text{con } \rho = 1$$

1.8.2.3. Distribuzione log-normale

Probabilmente la funzione più complessa nella sua definizione. La distribuzione dei dati è assimilabile ad una distribuzione log-normale quando il logaritmo dei *failure-time* si distribuisce normalmente.

La funzione di densità di distribuzione è definita dai parametri di media μ e deviazione standard σ del logaritmo naturale dei *failure-time*. La figura 11b riporta un esempio grafico di come essa si distribuisce evidenziando che a valori di $t \rightarrow 0$ $f(t) \rightarrow 0$, l'incremento del tempo ne determina un aumento progressivo fino al valore che corrisponde alla moda della funzione per poi decrescere progressivamente in una lunga coda. La definizione della funzione di sopravvivenza non avendo una determinazione specifica, ma riportando l'annotazione dell'integrale da 0 a t della funzione di densità di distribuzione, non viene riportata.

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right\}$$

Le figure 11a e 11b, come per i precedenti casi riportano l'andamento della funzione di sopravvivenza e della sua derivata prima nel tempo. Come si può vedere nella figura 11b si presenta un'accentuata skewness positiva (coda a destra) della distribuzione.

1.8.3. Stima parametrica

L'assimilazione della funzione di sopravvivenza (e quindi delle sue funzioni correlate) ad una funzione di distribuzione permette quindi di definire un andamento lineare continuo della dinamica di *failure* nella popolazione. Il principale metodo di verifica utilizzato per verificare l'applicabilità di una particolare distribuzione ai dati di sopravvivenza è di tipo grafico. Un rapporto di linearità tra i valori di $\log[S(t)]$ e t è identificativo dell'applicabilità di un modello a regressione esponenziale, mentre un rapporto di linearità tra i valori di $\log\{-\log[S(t)]\}$ e $\log(t)$ che la distribuzione del rischio basale dei dati segue la regressione di Weibull (Ducrocq, 1997).

La stima dei parametri delle diverse distribuzioni viene effettuata in funzione delle proprietà della funzione attraverso la stima della funzione di verosimiglianza:

La funzione di verosimiglianza della distribuzione esponenziale è definita come:

$$L(\nu) = \prod_{k=1}^n \nu \exp^{-\nu t_k} = \nu^n \exp^{-\nu \sum_{k=1}^n t_k} = \nu^n \exp^{-\nu n \bar{t}} \quad \text{dove} \quad \bar{t} = \frac{1}{n} \sum_{k=1}^n t_k$$

derivando il logaritmo della funzione si ottiene

$$\hat{\sigma} = \frac{1}{\bar{t}}$$

La distribuzione log-normale permette una stima diretta di μ e σ con il metodo della massima verosimiglianza utilizzato per la distribuzione normale, grazie alla proprietà che lega le due funzioni:

$$f_L(t; \mu, \sigma) = \frac{1}{t} f_N(\ln t; \mu, \sigma)$$

dove f_L e f_N sono rispettivamente le funzioni di distribuzione log-normale e normale con media μ e deviazione standard σ . I parametri sono così calcolabili:

$$\hat{\mu} = \frac{\sum_{k=1}^n \ln t_k}{n} \quad \text{e} \quad \hat{\sigma}^2 = \frac{\sum_{k=1}^n (\ln t_k - \hat{\mu})^2}{n}$$

Secondo la distribuzione di Weibull infine la stima dei parametri ρ e λ può avvalersi della funzione di verosimiglianza risolvendo contemporaneamente le derivate parziali $\frac{\partial \Lambda}{\partial \lambda} = 0$ e $\frac{\partial \Lambda}{\partial \rho} = 0$ dove Λ definisce il logaritmo naturale della funzione di verosimiglianza $L(t; \rho, \lambda)$.

$$L(t; \rho, \lambda) = \prod_{k=1}^n \left[\frac{\rho}{\lambda} \left(\frac{t_k}{\lambda} \right)^{\rho-1} \exp \left(- \left(\frac{t_k}{\lambda} \right)^\rho \right) \right]$$

Suddividendo la popolazione sottoposta a rilevazione di dati di sopravvivenza in gruppi omogenei secondo un fattore di classificazione è (maschi vs femmine, primipare vs pluripare ecc) possibile eseguire un test delle ipotesi tra i gruppi a confronto, verificando l'ipotesi nulla in merito all'assenza di differenze significative tra essi. Il confronto tra funzioni di sopravvivenza e di densità definite utilizzando le informazioni relative a ciascun gruppo è possibile mediante test del χ^2 permettendo di apprezzare l'entità delle differenze tra i gruppi consentendo di accettare o rifiutare l'ipotesi nulla.

1.8.4. Stima non parametrica

Come illustrato nel paragrafo relativo alla definizione dell'analisi di sopravvivenza, le rilevazioni del numero di soggetti che manifestano l'evento può essere condotta rilevando in modo discreto e suddividendo arbitrariamente l'intervallo di tempo analizzato, per una successiva descrizione tramite analisi della life-table, oppure registrando puntualmente i *failure-time* relativi ad ogni soggetto, applicare un'analisi indipendente dall'ampiezza dell'intervallo di tempo che intercorre tra due eventi consecutivi. L'analisi con gli estimatori di Kaplan Meier (1958) permette, dopo aver ordinato le informazioni per *failure-time* crescenti, questo tipo di approccio definendo la funzione di sopravvivenza e la relativa distribuzione di densità cumulativa in modo puntuale:

$$S_{KM}(t) = \prod_{j=1}^t \left(1 - \frac{d_j}{n_j} \right)$$

dove: $S_{KM}(t)$ è la funzione di sopravvivenza stimata con il metodo di Kaplan Meier, d e n indicano rispettivamente il numero di informazioni *failure* e *censored* al momento j rilevato puntualmente per ciascuna delle due tipologie di informazione. Come si può notare questo tipo di approccio non contempla nessuna parametrizzazione della funzione di sopravvivenza così come nemmeno della sua derivata. Dallo sviluppo di questa funzione Kalbfleisch e Prentice (1980) hanno sviluppato le funzioni per la stima della varianza degli estimatori e dei relativi intervalli di confidenza, nonché delle funzioni ad essa dipendenti. In figura 10, relativamente all'esempio riportato all'inizio del paragrafo 1.8 la linea continua riporta i valori di sopravvivenza utilizzando gli estimatori di Kaplan Meier.

L'applicazione di quest'analisi generalmente viene indicata per confrontare tra loro pochi gruppi di una popolazione. L'affermazione della significatività della differenza tra i gruppi può essere testata mediante log-rank test utilizzando il test di Wilcoxon.

L'analisi grafica della funzione di sopravvivenza tramite estimatori di Kaplan Meier rende possibile individuare se la distribuzione empirica dei dati di sopravvivenza segue qualche particolare tipo di distribuzione parametrica

permettendo sviluppare in questo senso l'analisi dei dati con tutti i vantaggi risultanti da un approccio di questo tipo (Falconer, 1996).

1.8.5. Modelli di regressione

Dal rapporto tra la funzione di *densità di distribuzione* nel tempo $f(t)$ e la sua primitiva $S(t)$ si ottiene la funzione di rischio $h(t)$ (Ducrocq, 1997):

$$h(t) = \frac{f(t)}{S(t)} = \frac{d \log S(t)}{dt} \quad \text{e} \quad S(t) = \exp \left\{ - \int_0^t h(u) du \right\}$$

Questa relazione diretta tra *hazard function* $h(t)$ e *density function* $f(t)$ implica che determinare la distribuzione della funzione di densità è sostanzialmente paragonabile a compiere la stessa operazione sulla funzione di rischio $h(t)$. Osservando la relazione qui descritta è possibile notare inoltre che quando la funzione di rischio assume valore 1 (rischio nullo), la derivata prima della funzione di sopravvivenza, ovvero la *density function* è identica alla *funzione di sopravvivenza* $S'(t) = f(t) = S(t)$.

Da questo assunto sono stati sviluppati una considerevole varietà di modelli destinati alla descrizione della funzione di rischio. I modelli utilizzati quasi esclusivamente quali descrittori della funzione di rischio sono quello a rischio proporzionale, secondo i quali il rapporto tra i rischi di due soggetti ($h_A(t)$ e $h_B(t)$) è costante e indipendente dal tempo (Cox, 1972). Definendo con x un vettore identificativo delle covariate esplicative del rischio e β il relativo vettore dei coefficienti di regressione è possibile descrivere la funzione di rischio $h(t)$ (*hazard function*) nei modelli a rischio proporzionale come il prodotto tra la funzione del rischio "basale" $h_0(t)$ che definisce la distribuzione del rischio comune (probabilità di *failure*) a tutti i soggetti presenti nel set di dati e $\exp^{x\beta}$ quantifica l'effetto delle covariate esplicative x :

$$h(t) = h_0(t) \exp^{x\beta}$$

La proporzionalità del rischio è verificabile rapportando due funzioni relative a soggetti identificati in gruppi di rischio differenti:

$$\frac{h_A(t)}{h_B(t)} = \frac{h_0(t)\exp^{x_A\beta}}{h_0(t)\exp^{x_B\beta}} = \exp[(x_A - x_B)' \beta] = K_{AB}$$

dove K_{AB} è costante e indipendente dal tempo tra A e B

Viste le relazioni esistenti tra funzione di rischio e di distribuzione di densità, analogamente a quanto descritto per l'individuazione di una funzione matematica che possa descrivere la distribuzione empirica del rischio, la prima scelta effettuabile nella definizione di un modello di regressione per l'analisi del rischio riguarda la distribuzione associata alla *baseline hazard function*. La funzione del rischio di base $h_0(t)$ esprime la variazione della distribuzione del rischio di *failure* nel tempo (Ducrocq, 1997) mentre il termine moltiplicativo $\exp^{x\beta}$ incrementa, riduce o mantiene inalterato (valori rispettivamente superiori, inferiori o uguali a 1) questo valore di rischio basale. Volendo fare un'analogia con i modelli lineari è possibile "paragonare" la funzione del rischio di base come la media del campione, mentre $\exp^{x\beta}$ rappresenta la componente degli effetti esplicativi considerati che quantificano il livello di scostamento tra livelli dell'effetto. In particolare β è il vettore dei coefficienti di regressione degli effetti e x il relativo vettore identificativo.

Le tipologie di distribuzione della funzione del rischio basale ricalcano quelle già descritte precedentemente nel paragrafo 1.8.2, ovvero $h_0(t)$ può essere definita secondo una distribuzione esponenziale, di Weibull e log-normale. In aggiunta a questi tipi di regressione parametrici, è possibile associare una distribuzione definita semi-parametrica identificata come modello di Cox (1972).

Con questo tipo di approccio è possibile riuscire a stimare i coefficienti di regressione β , senza fare alcuna assunzione sulla distribuzione della *baseline hazard function*, il cui valore viene definito ponendo a zero il valore di tutte le covariate x .

La possibilità di definire attraverso una parametrizzazione la funzione di rischio basale dipende dalle proprietà intrinseche ad ogni tipo di regressione:

un rapporto di linearità tra i valori di $\ln[S_{KM}(t)]$ e il tempo t è identificativo dell'applicabilità di un modello a regressione esponenziale

un rapporto di linearità tra i valori di $\log\{-\log[S_{KM}(t)]\}$ e $\log(t)$ che la distribuzione del rischio basale dei dati segue la regressione di Weibull (Ducrocq, 1997) nel quale $h_0(t) = \lambda\rho(\lambda t)^{\rho-1}$ dove da λ e ρ identificano l'ampiezza e la pendenza della funzione. Sviluppando la funzione di rischio $h(t)$ utilizzando il modello di Weibull otteniamo:

$$h(t) = \lambda\rho(\lambda t)^{\rho-1} \exp^{x'\beta} \quad \text{ovvero} \quad h(t) = \rho t^{\rho-1} \exp^{\rho \log \lambda + x'\beta}$$

Il valore di $\rho \log \lambda$, pur mantenendo le stesse notazioni, è possibile incorporarlo nelle variabili esplicative semplificando l'espressione del modello, riducendola a

$$h(t) = \rho t^{\rho-1} \exp^{x'\beta}$$

Per questa sua proprietà la *baseline hazard function* secondo la distribuzione di Weibull ha le stesse caratteristiche analizzate nel paragrafo 1.8.2.2 ovvero andamento monotono decrescente con $\rho < 1$, monotono crescente con $\rho > 1$ e assume valore 1 (rischio basale nullo) con $\rho = 1$.

L'impiego dei modelli a rischio proporzionale (PHM) hanno subito nel corso del loro utilizzo ulteriori sviluppi. Il determinismo del rischio non risulta essere influenzato unicamente da effetti esplicativi nei quali è possibile identificare un soggetto durante l'intero periodo di osservazione, ma anche dalla potenziale variazione degli stessi nell'arco della rilevazione delle informazioni. Una potenzialità dei PHM è, appunto, la possibilità di determinare dei coefficienti di rischio relativi ad ogni livello di variazione nel tempo dell'effetto. Prendendo, ad esempio, l'ordine di parto quale variabile x identificativa del rischio di riforma di una vacca, è verosimile affermare che le primipare $x(t')$ hanno un livello di rischio di eliminazione diverso rispetto a vacche al secondo parto $x(t'')$. Questa distinzione ha il pregio di delineare con maggiore efficienza il determinismo del rischio di *failure* col trascorrere del tempo. Si distinguono quindi dagli effetti *tempo-indipendenti*, che non variano durante l'intero periodo di osservazione (per es. sesso, età al primo parto...), gli effetti di tipo *tempo-dipendente* (quali classi di ordine di parto, classi di lunghezza dell'interparto o della lattazione...). Gli effetti tempo dipendenti sono identificabili nell'espressione della

funzione di rischio, definendo il vettore della variabile identificativa come funzione del tempo t come segue:

$$h(t) = h_0(t) \exp^{x'_1 \beta_1 + x'_2(t) \beta_2}$$

dove $x'_1 \beta_1$ identifica il prodotto tra vettore dei coefficienti di regressione degli effetti tempo indipendenti e il relativo vettore identificativo, mentre $x'_2(t) \beta_2$ è il prodotto del vettore dei coefficienti di regressione degli effetti tempo dipendenti con il relativo vettore identificativo.

Ulteriori approcci a questa metodologia hanno permesso di stratificare il rischio basale. Qualora sia possibile identificare un effetto che non varia la natura della funzione di rischio di base, ma ne cambia la sua distribuzione, attribuendo in questo modo parametri diversi ad ogni strato della *baseline hazard function* è possibile definire nel modello distribuzioni caratterizzate da diversi parametri. Nel caso di analisi con stratificazione, è possibile identificare l'effetto di stratificazione nella funzione del rischio di base come l' s^{esimo} strato di $h_0(t)$:

$$h(t) = h_{0,s}(t) \exp^{x'_1 \beta_1 + x'_2(t) \beta_2}$$

In merito alla stima dei parametri delle distribuzioni utilizzate per definire la *baseline hazard function*, si rimanda al paragrafo relativo alle stime parametriche, evidenziando che la funzione di densità di distribuzione $f(t)$ sulla quale poter sviluppare la funzione di verosimiglianza è data dal prodotto della funzione di rischio $h(t)$ e quella di sopravvivenza $S(t)$:

$$L(t) = \prod_{k=1}^n [h(t_k) S(t_k)]$$

Relativamente ai PHM nell'analisi di sopravvivenza, in merito alla stima dei coefficienti di regressione, la funzione di verosimiglianza viene costruita dal prodotto dei singoli contributi che assumeranno diversa rilevanza in funzione se il tipo di informazione che riportano è di tipo censored o uncensored. Ducrocq (1997) riporta che il contributo parziale alla funzione di verosimiglianza del soggetto k è ottenuto dalla funzione di densità di distribuzione se l'informazione è di tipo uncensored ($\delta_k = 1$),

mentre se il dato fenotipico è parziale (dato *censored* $\delta_k = 0$) il contributo parziale è definito dalla funzione di sopravvivenza

$$\ell_k = [h(t_k; \rho, \lambda, \beta)]^{\delta_k} S(t_k; \rho, \lambda, \beta) \quad L(t) = \prod_{k=1}^n h(t_k)^{\delta_k} S(t_k)$$

La costruzione della funzione parziale di verosimiglianza è possibile quando non viene fatta alcuna assunzione sulla distribuzione dei dati *censored* e unicamente nel caso in cui i dati troncati siano di tipo *random censoring* (come illustrato nel paragrafo 1.8.1.2).

1.8.6. Bontà di stima e test delle ipotesi (bootstrapping)

Come illustrato nel paragrafo “1.8.3. stime parametriche”, la scelta della distribuzione che descrive le funzioni di sopravvivenza e le sue correlate può essere fatta graficamente. Ducrocq (1994) afferma che il metodo empirico grafico, per l’assunzione di una distribuzione esponenziale è verificare la linearità dell’andamento di $\ln[S(t)]$ su t , mentre per verificare l’applicazione della distribuzione di Weibull vanno messe in grafico il $\log\{-\log[S(t)]\}$ sull’asse delle ordinate $\log t$ sull’asse delle ascisse.

Casellas nel 2006 ha proposto un metodo, basato su simulazioni ottenute con metodo Monte-Carlo, ridurre la soggettività nella determinazione della linearità grafica per i modelli che utilizzano come *baseline hazard function* la distribuzione di Weibull. Il metodo fornisce valori numerici oggettivamente confrontabili che non lasciano adito ad interpretazioni soggettive e si divide in 3 passi operativi:

- 1) determinazione del modello da testare e relative stime dei parametri della *baseline hazard function* (λ e ρ) dei coefficienti di regressione (β) degli effetti tempo indipendenti e tempo dipendenti.
- 2) generazione di nuovi set di dati (variabile dipendente T_i), partendo dai risultati ottenuti nel punto precedente (λ , ρ e β) e dai vettori identificativi (x) relativi al set di dati di partenza.

Senza *time-dependent* :

$$T_i = \left(-\frac{\ln(1-\alpha)}{\exp[\hat{\rho} \ln(\hat{\lambda}) + x_i \hat{\beta}]} \right)^{\frac{1}{\rho}}$$

Nel caso di utilizzo di effetti tempo dipendenti viene introdotto il parametro di ricorsività k va da 1 a n incrementando di un'unità per ogni ciclo ricorsivo fino a quando T_i assume un valore minore uguale a τ_k . Lo stesso Casellas (2007) afferma in una comunicazione personale la non completa chiarezza nella ricorsività della formula per la determinazione di T_i , non potendo definire la sommatoria per j che va da 1 a $k-1$ nel momento in cui k assume valore 1. In questa situazione, che rappresenta la situazione di partenza nel calcolo di T_i , Casellas attribuisce alla sommatoria valore 0.

$$\text{Con } time\text{-dependent: } T_i = \left\{ \left[\frac{\ln(1-\alpha) + \exp(x_i^* \hat{\beta}^*) \sum_{j=1}^{k-1} \hat{\lambda}^{\hat{\rho}} (\tau_j^{\hat{\rho}} - \tau_{j-1}^{\hat{\rho}}) \exp(\hat{\beta}_j)}{\hat{\lambda}^{\hat{\rho}} \exp(\hat{\beta}_k) \exp(x_i^* \hat{\beta}^*)} \right] + \tau_{k-1}^{\hat{\rho}} \right\}^{\frac{1}{\hat{\rho}}}$$

α = numero casuale tra 0 e 1 ottenuto da una distribuzione normale.

$\hat{\lambda}$ = lambda della *baseline hazard function* stimato

$\hat{\rho}$ = rho della *baseline hazard function* stimato

β = vettori dei coefficienti di regressione

x_i^* = i esimo vettore dei fattori identificativi

$\beta_1, \beta_2, \dots, \beta_n$ cambiano con $\tau_1, \tau_2, \dots, \tau_n$

- 3) calcolo degli intervalli di confidenza della probabilità di sopravvivenza ad ogni momento t nei dati reali e nelle diverse simulazioni. Per questo è necessario calcolare le sopravvivenze attese:

$$S_0^*(t) = \exp \left[-\hat{\lambda}^{\hat{\rho}} (t^{\hat{\rho}} - \tau_{k-1}^{\hat{\rho}}) \exp(\hat{\beta}_k) - \sum_{j=1}^{k-1} \hat{\lambda}^{\hat{\rho}} (\tau_j^{\hat{\rho}} - \tau_{j-1}^{\hat{\rho}}) \exp(\hat{\beta}_j) \right] \text{ con } t \in (\tau_{k-1}, \tau_k)$$

nella quale t va da 1 a n incrementando di un'unità se $S(t|x) > \alpha$

La verifica può essere eseguita utilizzando i classici intervalli di confidenza del 2,5% (0.025-0.975). Delineate le curve di sopravvivenza superiore ed inferiore, è possibile confrontarle con la curva ottenuta dai dati a disposizione. L'assunzione di proporzionalità si verifica quando i *P-values* risultano essere significativi (figura 15b), ovvero, volendo rappresentare la situazione graficamente, quando la curva di sopravvivenza giace tra le due curve marginali, come nell'esempio presentato in figura

15a . In questo esempio, riportato sempre dal lavoro di Casellas (2006) la distribuzione di Weibull sarebbe applicabile unicamente nell'intervallo di tempo compreso nel rettangolo rosso della figura 15a.

Il confronto tra modelli per determinarne la bontà di predizione può invece essere condotto classicamente con il confronto dei valori di logaritmo della verosimiglianza, di ciascun modello, o suoi derivati quale ad esempio il valore di AIC (Akaike Information Criteria) $AIC = 2n - 2\ln(L)$ o ancora likelihood ratio test (Casellas, 2006)

1.8.7. Frailty models

Secondo lo stesso sviluppo dei modelli lineari, data la stretta relazione illustrata precedentemente con la *funzione di densità di distribuzione*, si tratta comunque di un'analisi che ricalca il metodo dei minimi quadrati (Gehan e Siddiqui, 1973), è possibile introdurre anche effetti casuali, anch'essi distinguibili in tempo indipendenti, come l'effetto della componente genetica, o tempo dipendenti ai quali viene associata una quota del rischio residuo. Modelli di survival analysis che includono questo tipo di effetti vengono definiti *frailty models* (modelli "imperfetti"):

$$h(t) = v_q h_{0,s}(t) \exp^{x'_1 \beta_1 + x'_2(t) \beta_2}$$

dove v_q può rappresentare ad esempio la componente "*frailty*" casuale che descrive la variabilità individuale nel rischio del gruppo di figlie del toro q . Applicando una trasformazione logaritmica in cui $z = \log v \Leftrightarrow v = \exp z$ è possibile includere la componente "imperfetta" tra le variabili esplicative del modello. La stessa considerazione è applicabile anche per effetti casuali o "*frailty*" di tipo tempo dipendente $w(t) = \log[\tau(t)] \Leftrightarrow \tau(t) = \exp w(t)$:

$$h(t) = h_{0,s}(t) \exp^{x'_1 \beta_1 + x'_2(t) \beta_2 + z'u + w'(t)q}$$

dove $h(t)$ rappresenta il rischio relativo del verificarsi dell'evento al tempo t . $h_{0,s}(t)$ identifica la baseline hazard function. $x'_1 \beta_1$ e $x'_2(t) \beta_2$ identificano, come sempre, il prodotto dei vettori dei coefficienti di regressione degli effetti fissi, rispettivamente tempo indipendenti e tempo dipendenti, per il relativo vettore delle variabili

identificative; $z'u$ e $w'(t)q$ sono il prodotto tra il vettore dei coefficienti di regressione degli effetti casuali, rispettivamente tempo indipendenti e tempo dipendenti, ed il relativo vettore delle variabili identificative.

In merito agli effetti casuali va sottolineato infine che anche la loro descrizione generalmente viene fatta attraverso la densità di distribuzione associata ad una funzione. La più utilizzata nella descrizione dei fenomeni zootecnici è la distribuzione log-gamma che corrisponde alla distribuzione di $\log(x)$ quando x segue una distribuzione di tipo gamma. Nessuna delle distribuzioni attribuibili agli effetti casuali risulta essere affine alla distribuzione multivariata normale ($MVN(0;A\sigma^2)$) usata comunemente nei modelli poligenici infinitesimali per la stima delle componenti di varianza genetico additiva in cui A identifica la matrice di parentela tra i soggetti. Nel campo del miglioramento genetico bovino però è stato dimostrato che la componente “frailty” v , utilizzando le stime ottenute per definire i parametri dalla sua distribuzione gamma, segue una distribuzione verosimilmente log-normale. Applicando la trasformazione $z = \log v$ vista ad inizio paragrafo, la distribuzione di z può quindi essere considerata di tipo normale. Integrando nel processo di stima dei parametri della distribuzione di v la matrice di parentela tra animali è possibile considerare, in seguito all’affermazione sopra descritta, che le stime di z assumano una distribuzione multivariata normale con la quale poter stimare le componenti di varianza genetica.

Integrando la distribuzione marginale a posteriori ottenuta combinando la funzione di verosimiglianza con l’assunzione a priori della normalità di distribuzione $N(0;\sigma_G^2)$ dell’effetto legato alla componente genetica è possibile ottenere infine i parametri che descrivono la distribuzione marginale a posteriori dell’effetto casuale genetico con i quali determinare le componenti di varianza. La scelta del tipo di integrazione dipende dalla tipologia di informazioni dei dati analizzati. L’integrazione corretta per la stima della distribuzione marginale a posteriori è quella algebrica che operativamente non è sempre applicabile (Ducrocq e Casella, 1996). Le possibili soluzioni riportate sono l’estrazione di campioni casuali sui quali poter eseguire l’integrazione algebrica (Korsgaard, 1996), eseguire un’integrazione numerica (Smith

et al 1985) o definire un metodo di approssimazione come l'integrazione di Laplace (Ducrocq e Casella, 1996).

1.9. Bibliografia dell'introduzione

- A.I.A. – 2005. <http://www.aia.it/bollettino/bollettino.htm>
- A.N.A.R.B. – 2007. Incontro europeo dei punteggiatori per l'armonizzazione delle valutazioni morfologiche. Zug (CH)
- Bittante G., Andrighetto I., Ramanzin M. – 1990. Fondamenti di zootecnica. Miglioramento genetico, nutrizione e alimentazione. LIVIANA EDITRICE, TORINO.
- Caraviello D. Z., Weigel K. A., Shook G. E., Ruegg P. L. – 2005. Assessment of the Impact of Somatic Cell Count on Functional Longevity in Holstein and Jersey Cattle Using Survival Analysis Methodology *J. Dairy Sci.* 88:804–811
- Caraviello D. Z., Weigel K. A., Gianola D. – 2004. Analysis of the Relationship Between Type Traits and Functional Survival in US Holstein Cattle using a Weibull Proportional Hazards Model *J. Dairy Sci.* 87:2677–2686
- Casellas J., Tarrés J., Piedrafita J., Varona L. – 2006. Parametric bootstrap for testing model fitting in the proportional hazards framework: An application to the survival analysis of Bruna dels Pirineus beef calves *J. Anim. Sci.* 2006. 84:2609–2616
- Casellas J. - 2007. Comunicazione personale
- Cassandro M., Ghiroldi S., Bagnato A., Gallo L., Carnier P., Santus E. – 1998. Genetic evaluation of longevity in Italian brown cattle breed. Proceedings international workshop on genetic improvement of functional traits in cattle (GIFT). Jouy-en-Josas (F)
- Chirinos Z., Hernández D., Diaz C., Carabaño M. J. – 2006. Survival analysis of longevity in Spanish Holstein-friesian cows on a lactation basis. 8th World Congress on Genetics Applied to Livestock Production. August 13-18 2006, Belo Horizonte (BRA).
- Comin A., Cassandro M., Povinelli M., Bittante G. – 2005. Genetic aspects of milk coagulation properties in Italian Holstein cows. *Italian J. Anim. Sci.* vol. 4(2): 10-12.
- Cox, D. – 1972. Regression models and life tables. *J. Royal Stat. Soc.* 34:187-20.
- Cox, D.R. and Oakes, D. – 1984. Analysis of survival data. Chapman and Hall, London, UK.
- Damgaard L.H., Korsgaard L.R., Simonsen J., Dalsgaard O., Andersen A.H. – 2006. The effect of ignoring individual heterogeneity in Weibull log-normal sire frailty models. *J. Anim. Sci.* 84:1338-1350.
- Druet T., Sölkner J., Groen Ab F., Gengler N. – 1999. Improved genetic evaluation of survival using MACE to combine direct and correlated information from yield and functional traits. Proceedings international workshop on genetic improvement of functional traits in cattle (GIFT). Jouy-en-Josas (F) *Interbull Bulletin* 21: 122-127.
- Ducrocq, V., Quaas R. L., Pollak G., Casella G. – 1988. Length of productive life in dairy cows. Variance component estimation and sire evaluation. *J. Dairy Sci.* 71:3071-3079

- Ducrocq V. – 1994. Statistical analysis of length of productive life for dairy cows of the Normande breed. *J. Dairy Sci.*, 77, 855-866.
- Ducrocq V., Casella G. – 1996. A Bayesian analysis of mixed survival models. *Genet.Sel. Evol.* 28:505-529
- Ducrocq V. – 1997. Survival analysis, a statistical tool for longevity data. 48th Annual Meeting of the European Association for Animal Production - 25- 28 August, 1997 Vienna, Austria
- Ducrocq, V. and Sölkner, J. – 1998. The Survival Kit – V3.0, a package for large analyses of survival data. 6th World Cong. Genet. Anim. Genetics and Breeding Unit, Univ. of New England, Armidale, Australia. *Appl. Livest. Prod.* 27 : 447-448
- Ducrocq V. – 2001. A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life. Proceedings of the 2001 Interbull Meeting in Budapest (H), *Interbull Bulletin* 27: 147 – 152.
- Edmonson A. J., Lean I. J., Weaver L. D., Farver T., Webster G. – 1989. A body condition scoring chart for Holstein dairy cows. *J. Dairy Sci.*, 72, 68-78.
- Falconer, D. S., and T. F. C. Mackay, - 1996. *Introduction to Quantitative Genetics*, Ed 4. Longmans Green, Harlow, Essex, UK.
- Forabosco F., Groen A. F., Bozzi R., Van Arendonk J. A. M., Filippini F., Böttcher P., Bijma P. – 2004. Phenotypic relationships between longevity, type traits, and production in Chianina beef cattle. *J. Anim. Sci.* 82: 1572-1580.
- Forabosco F., Bozzi R., Filippini F., Böttcher P., Van Arendonk J. A. M., Bijma P. – 2005. Confronto tra modello nella valutazione genetica dei tori di razza chianina per il carattere longevità. 4 th World Italian Beef Cattle Congress. 29th April – 01st May 2005 Italy.
- Gehan E. A., Siddiqui M. M. – 1973. Simple regression methods for survival time studies. *J of the Amer. Stat. Assoc.* 68:848-856
- Gonzales-Recio O., Chang Y.M., Gianola D., Weigel K. A. – 2006. Comparison of models using different censoring scenarios for days open in Spanish Holstein cows. *Anim. Sci.* 82:233-239.
- Gröhn Y. T., Eicker S. W., Ducrocq V., Hertl J. A. – 1998. Effect of Diseases on the Culling of Holstein Dairy Cows in New York State. *J Dairy Sci* 81:966–978
- Hald A. – 1949. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuar.* 32: 119-34
- Hougaard P. - 1999. Fundamentals of survival data. *Biometrics* 55: 13-22
- Kaplan, E. L. & Meier, P. – 1958. Nonparametric estimation from incomplete observations. *J. of the American Statistical Association*, 53: 457-48.
- Kalbfleisch J.D. and Prentice R.L. (1980). *The statistical analysis of failure time data.* John Wiley and Sons, New York (USA)

- Korsgaard I. R. – 1996. Genetic analysis of survival data – a Gibbs sampling approach. International workshop on genetic improvement of functional traits in cattle. Faculté universitaire des sciences agronomiques. Gembloux (B).
- Larroque H., Ducrocq V. – 2001. Relationships between type and longevity in the Holstein breed. *Genet.Sel. Evol.* 33:39-59
- Leclerc H., Fiske W. F., Ducrocq V. – 2005. Prospects of principal components and factor analysis for estimation of genetic correlations among countries for milk production traits. Interbull Meeting. June 2-4. Uppsala (SWE).
- Louzada-Neto F., Davison A. C. – 1998. A note on bayesian analysis of the poly-weibull model. <http://en.scientificcommons.org/41939>
- López de Maturana E., Ugarte E., González-Recio O. – 2007. Impact of Calving Ease on Functional Longevity and Herd Amortization Costs in Basque Holsteins Using Survival Analysis. *J. Dairy Sci.* 90:4451–4457
- Moreno C. R., Elsen J. M., Le Roy P., Ducrocq V. – 2005. Interval mapping methods for detecting QTL affecting survival and time-to-event Phenotypes. *Genet. Res Camb.* 85:139-149
- Neerhof H.J., Madsen P., Ducrocq V., Vollema A. R., Jensen J., Korsgaard R. – 2000. Relationships between mastitis and functional longevity in danish black and white dairy cattle estimated using survival analysis. *J. Dairy Sci.* 83:1064-1071
- Perez-Cabal M. A., Alenda R. – 2002. Genetic Relationships between Lifetime Profit and Type Traits in Spanish Holstein Cows. *J Dairy Sci.* 85(12): 3480 - 3491.
- Perez-Cabal M. A., Garcia C., Gonzalez-Recio O., R. Alenda – 2006. Genetic and Phenotypic Relationships Among Locomotion Type Traits, Profit, Production, Longevity, and Fertility in Spanish Dairy Cows. *J Dairy Sci.* 89(5): 1776 - 1783.
- Phocas F., Ducrocq V. – 2006. Discrete vs continuous time survival analysis of productive life of charolais cows. 8th World Congress on Genetics Applied to Livestock Production. August 13-18 2006, Belo Horizonte (BRA).
- Ral G., Berglund B., Philipsson J., Emanuelson U., Tengroth G. – 1990. Comprehensive experiments on traits affecting longevity in Swedish dairy cattle breeds. Teat, udder and milkability traits effects of breed and age repeatabilities and correlations. *Swedish. J. Agricult. Res.* 52: 89-95.
- Roxsötrm A., Ducrocq V., Strandberg E. – 2003. Survival analysis of longevity in dairy cattle on a lactation basis. *Genet. Sel. Evol.* 35: 305-318.
- Smith S.P., Quaas R. L. –1994. Productive life span of bull progeny groups: failure time analysis. *J. Dairy Sci.* 67:2999-3007.
- Sölkner J., Ducrocq V. – 1999a. The Survival Kit: a tool for analysis of survival data. Proceedings International Workshop on EU Concerted Action Genetic Improvement of functional traits in Cattle (GIFT) - Longevity. Jouy-en-Josas (F) Interbull Bulletin 21:11-16.

- Sölkner J., Petschina R. – 1999b. Relationship between type traits and longevity in Austrian Simmental cattle. Proceedings of the international workshop on genetic improvement of functional traits (GIFT). Interbull Bulletin, 21: 91–96.
- Sölkner, J., Fuerst, C. – 2002. Breeding for functional traits in high yielding dairy cows. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France, 29:107-114.
- Statsoft - 2003. <http://www.statsoft.com/textbook/stsurvan.html>
- Strapák P., Candrák J., Aumann J. – 2005. Relationship between longevity and selected production, reproduction and type traits. Czech J. Anim. Sci., 50(1): 1–6
- Tarrés J., Piedrafita J., Ducrocq V. – 2006. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. Genet. Sel. Evol. 38: 65–83
- Terawaki Y., Katsumi T., Ducrocq V. – 2006. Development of a survival model with piecewise Weibull vaseline for the analysis of lenght of productive life of holstein cows in japan. J. Dairy Sci. 89: 4058-4065.
- Vanraden P. M. – 1995. Inbreeding adjustments and effect on genetic trend estimates. Interbull Meeting. June 2-4. Uppsala (SWE).
- Vicario D. - 2005. La sopravvivenza delle figlie dei tori di razza pezzata rossa italiana. Meeting della sezione regionale allevatori di razza Pezzata Rossa del Veneto
- Vollema, A.R. – 1998. Doctoral thesis, Animal breeding and genetics group. Wageningen agricultural university. Wageningen (NL)
- Vukasinovic N. – 1997. Application of survival analysis in breeding for longevity. Proceedings international workshop on genetic improvement of functional traits in cattle (GIFT). Jouy-en-Josas (F)
- Welgel K.A., Palmer R. W., Caraviello D. Z. – 2003. Investigation of factors affecting voluntary and involuntary culling in expanding dairy herds in Winsconsin using serviva analysis. J. Dairy Sci. 86:1482-1486
- Yazdi M. H., Visscher P. M. Ducrocq V., Thompson R. – 2002. Heritability, reliability of genetic evaluationa and response to selection in proportional hazard models. J. Dairy Sci. 85:1563-1577.

2. OBIETTIVI

Questo progetto di ricerca nasce a fronte del crescente interesse da parte degli allevatori e di A.N.A.R.B. nei confronti dei caratteri cosiddetti secondari, ovvero di quella serie di caratteri che non comportano un incremento diretto del tornaconto economico dell'allevatorie, ma mirano ad una oculata gestione delle spese. L'ottenimento di un indice diretto o combinato per la longevità, comporterebbe, come illustrato nel capitolo precedente, non pochi vantaggi economici e gestionali negli allevamenti, senza però dimenticare che il progresso genetico ottenibile dai caratteri funzionali, a fronte di parametri genetici che in bibliografia risultano essere medio bassi, non può essere che di ridotta entità. L'obiettivo principale si compone di diversi step operativi che andranno a comporre i contributi sperimentali di questa tesi di dottorato.

L'obiettivo primario è quello di ottenere un indice genetico accurato e precoce della longevità ponderando, sulla base delle informazioni a disposizione, la combinazione della misura diretta del carattere "giorni di vita produttiva dal primo parto all'eliminazione" con i caratteri che mostrano nella combinazione con esso un buon grado di predittività dello stesso. L'impiego della survival analysis permetterà l'impiego anche dei dati troncati (*censored*).

Per ottenere questo la sequenza operativa sarà la seguente:

1. analisi del carattere longevità (LPL) nella popolazione di razza Bruna in Italia, ricerca dei fattori di rischio e determinazione del modello per la stima dei parametri genetici e di valori riproduttivi.
2. stima delle correlazioni genetiche del carattere oggetto di studio (LPL) con gli altri caratteri già oggetto di stima nella popolazione di razza Bruna in Italia: produttivi, morfologici, funzionali e successivo sviluppo di un indice composto.

3. MATERIALE E METODI

3.1. Base dati

I dati impiegati sono stati ottenuti estraendo le informazioni dalla base dati dell'archivio A.N.A.R.B. Il periodo di osservazione, è compreso tra il 01 gennaio 1955 e il 30 agosto 2007. La tabella 3.1 riporta il numero di tutti i record estratti dall'archivio ANARB.

Tabella 3.1: record estratti dalla base dati A.N.A.R.B.

Base dati	Records
Controlli giornalieri	17.925.234
Anagrafico	3.228.472
Vendite	891.921

Le informazioni estratte per l'analisi di sopravvivenza riguardavano:

Più di 17 milioni di record relativi ai controlli giornalieri riferiti a oltre 2 milioni di lattazioni appartenenti a più di 700.000 vacche di razza Bruna allevate in aziende iscritte al libro genealogico nazionale. Il regolamento per lo svolgimento dei controlli della produzione del latte nella specie bovina sono regolamentati dal D.M. 24-5-1967 modificato con D.M. 28-9-1981 (A.I.A., 1981). I dati registrati, suddivisi per razza, sono inviati periodicamente alle rispettive associazioni nazionali.

- Oltre 3 milioni di registrazioni anagrafiche: l'acquisizione dell'informazione è affidata alle Associazioni Provinciali o Regionali Allevatori (A.P.A. e A.R.A.) in qualità uffici periferici del libro genealogico. Periodicamente le informazioni sono inviate all'ufficio centrale che ha il compito di provvedere alla verifica e all'archiviazione delle informazioni secondo quanto riportato dai "*Disciplinari del libro genealogico dei bovini di razza bruna*" (A.N.A.R.B., 2006). L'archivio è periodicamente aggiornato con le informazioni relative alla data di eliminazione e alla causa di riforma provenienti dalla stessa fonte.
- Le registrazioni di vendita sono relative a singola variazione di allevamento e all'acquisto di vacche dall'estero; sono raccolte a livello di A.P.A. e A.R.A., inviate ad A.N.A.R.B. che provvede all'archiviazione.

Da questi archivi sono state estratte, relativamente ad ogni vacca, identificata con una matricola univoca di 16 caratteri, codificata a livello nazionale dal Regolamento (CE) n. 820/97 del 21 aprile 1997 , le seguenti informazioni:

- data di nascita e, dove presente, di riforma
- padre della vacca
- data di ogni singolo parto e relativo ordine di parto
- Allevamento, identificato con codice univoco secondo protocollo metodologico A.N.A.R.B. da un codice di 8 cifre che identificano la provincia (le prime tre cifre) e la ragione sociale (le ultime cinque cifre), in cui sono stato registrato ogni singolo controllo giornaliero
- produzione espressa in equivalente vacca matura (E.V.M.) di ogni lattazione
- data di vendita della vacca ad altro allevamento
- data dell'ultimo controllo giornaliero registrato

3.2. Editing

Su ogni record è stato applicato un filtro iniziale sulla base di informazioni non utilizzate, incongruenti o impossibili dovuto probabilmente ad errate registrazioni:

- codice razza del soggetto con codice '01' corrispondente a razza Bruna
- codice matricola identificativa della vacca e del padre presenti nell'archivio anagrafico
- data di nascita presente e antecedente a alla data di termine delle osservazioni
- date di parto, di vendita, data di ultimo controllo giornaliero registrato e di riforma successive alla data di nascita
- data di riforma, dove presente, successiva alla data di ultimo controllo giornaliero registrato.

Dalle registrazioni di vendita è stato estratto il primo record di ogni soggetto relativo alla prima vendita successiva alla data di primo parto.

Sono state considerate lattazioni contenenti tutte le informazioni utilizzate nell'analisi. La mancanza di una sola informazione ha determinato l'eliminazione dell'intera lattazione.

I valori EVM utilizzati sono compresi nell'intervallo $-3/+3$ unità di deviazioni standard (uds) rispetto alla media EVM calcolata per anno di parto. L'età al primo

parto è stata ritenuta valida se compresa tra i 18 e i 46 mesi. Si è verificato che la lunghezza dell'interparto fosse compresa tra 300 giorni e +2,5 unità di deviazioni standard oltre la media di interparto della popolazione.

Riguardo ai controlli di consistenza sono state mantenute le informazioni relative a gruppi di almeno 5 lattazioni per allevamento all'anno considerando la combinazione delle informazioni come classe “*spuria*”, procedendo a un doppio controllo: la combinazione azienda-anno con frequenza inferiore a 5 è stata inizialmente sostituita con il dato di allevamento. Una frequenza inferiore alla soglia anche dopo sostituzione ha determinato l'eliminazione delle informazioni dell'intera lattazione.

Nella verifica di queste condizioni stata mantenuta la successione degli ordini di parto a partire dal primo. È stata tenuta traccia dell'eliminazione delle lattazioni anomale secondo quanto appena descritto ed è stata associata all'ultima lattazione valida di ogni soggetto la data dell'ultimo controllo giornaliero registrato.

La numerosità delle informazioni utilizzate per l'analisi di sopravvivenza è riportata in tabella 3.2 contestualmente alle informazioni che si riferiscono alla numerosità e alla percentuale delle osservazioni *uncensored*.

Per gli 11.082 tori padri delle vacche presenti nel *dataset* è stata ottenuta la genealogia in linea paterna. Il pedigree ha raggiunto la massima profondità disponibile. Le informazioni genealogiche ottenute riguardano 11.918 tori nati tra il 1949 ed il 2004.

3.3. Survival analysis

La variabile oggetto di studio è la lunghezza della vita produttiva (LVP) espressa in giorni compresi tra la data di primo parto e, dove presente e valida, la data di riforma. Le carriere produttive che dispongono di entrambe queste informazioni definiscono le osservazioni di tipo *uncensored*. È stata considerata valida come data di riforma anche la data di ultimo controllo giornaliero registrato solo nel caso in cui la data di riforma presente in archivio si collochi oltre 180 giorni da essa e contemporaneamente sia presente l'intera carriera produttiva della vacca. È stata calcolata anche la lunghezza della vita (LV) sostituendo la data di primo parto con quella di nascita.

Le informazioni troncate sono state definite come *random censoring*. La data di *censoring* è stata definita contestualmente alla definizione della causa che ha determinato l'interruzione dell'osservazione della carriera produttiva di ciascun soggetto. La tabella 3.2 riporta le cause che hanno portato all'attribuzione della codifica *censored* e la relativa definizione della data di *censoring*.

Tabella 3.2: motivi di attribuzione della codifica *censored* e date associate di termine osservazione (*censoring-date*).

Causa di <i>censoring</i>	Data associata
eliminazione di lattazioni non sequenziali relativamente a: <ul style="list-style-type: none"> - controllo di consistenza di allevamento-anno - mancanza e non determinabilità del valore EVM - EVM estremi (± 3 uds) - lattazioni non registrate - interparti estremi (<300gg; >+ 2,5 uds) 	data dell'ultimo controllo giornaliero registrato relativo all'ultima lattazione disponibile
vendita	data di vendita
mancanza di una data di riforma valida	data dell'ultimo controllo giornaliero registrato relativo all'ultima lattazione

L'attribuzione della data di *censoring* come *end-point* relativo alla carriera produttiva delle vacche è stata fatta sulla base della prima tra le tre cause di *censoring* in ordine temporale con conseguente eliminazione delle informazioni riguardanti lattazioni a essa successive. Non sono stati considerati record left-*censored*.

Le numerosità dei set di dati ottenuti sono riportate in tabella 3.3

Tabella 3.3: numerosità del dataset analizzato e relativa percentuale di informazioni *uncensored*

Informazione	Uncensored		Totali
	numerosità	%	
Vacche	250.986	77.5	323.849
Lattazioni	711.061	75.0	947.996
Allevamenti	10.533		
Gruppi di Allevamento*anno	47.923		
Tori padri di vacche	11.082		
Tori nella matrice di parentela	11.918		

È stata ottenuta la curva di sopravvivenza della popolazione con gli estimatori di Kaplan-Meier $S_{KM}(t)$ calcolati sia per LV sia per LPL. È stata verificata l'assunzione di Weibull analizzando graficamente la linearità tra $\log\{-\log[S(t)]\}$ e $\log t$ per entrambi i caratteri. Sul carattere LPL è stata ripetuta la verifica stratificando i dati per ordine di parto.

I dati sono stati analizzati con il pacchetto “*Survival kit V3.12*” (Ducrocq, 2000) con un *sire model* parametrico assumendo per la *baseline hazard function* la distribuzione di Weibull.

$$h(t) = h_0(t) \exp^{x_1' \beta_1 + x_2'(t) \beta_2 + z'u + w'(t)q}$$

dove $h(t)$ è il rischio di eliminazione al tempo t , $h_0(t)$ è la funzione del rischio basale, β_1 è il vettore sconosciuto dei coefficienti di regressione degli effetti fissi non genetici tempo indipendenti, β_2 è il vettore sconosciuto dei coefficienti di regressione degli effetti fissi non genetici tempo dipendenti, u è il vettore sconosciuto dei coefficienti di regressione dell'effetto casuale tempo indipendente dovuto al toro, q è il vettore sconosciuto dei coefficienti di regressione dell'effetto casuale tempo dipendente dovuto alla combinazione di allevamento-anno, x_1 è il vettore identificativo degli effetti fissi non genetici tempo indipendenti, $x_2(t)$ è il vettore identificativo degli effetti fissi non genetici tempo dipendenti, z è il vettore dell'effetto casuale tempo indipendente

identificativo del toro, $w_2(t)$ è il vettore dell'effetto casuale tempo dipendente identificativo dell'interazione allevamento-anno.

Gli effetti fissi tempo indipendenti inclusi nel modello sono:

- allevamento (*Allev*): classificato in 10.533 livelli identifica l'azienda in cui è stata registrata la prima lattazione. Il cambiamento di azienda durante il periodo di osservazione ha determinato, come descritto nel paragrafo precedentemente, l'attribuzione della codifica *censored* e la perdita delle informazioni successive alla data di vendita.
- età al primo parto (*Età1OP*): suddivisa in 8 classi di cui 6 trimestrali da 24 a 41 mesi più 2 livelli estremi (18-23 mesi e 42-46 mesi)

Gli effetti fissi tempo dipendenti inclusi nel modello sono:

- ordine di parto (*OP(t)*): suddiviso in 10 livelli (da 1 a 9 e ≥ 10) con variazione in corrispondenza della data di parto;
- Interazione tra dimensione dell'azienda e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo ($nAllev(t) * \Delta nAllev_{(t_y-t_{y+4})}$): gli effetti che costituiscono l'interazione sono stati così determinati:

- o dimensione dell'azienda ($nAllev(t)$): distinto in 4 livelli (meno di 7 o piccolo, da 7 a 12 o medio-piccolo, da 12 a 20 o medio-grande e oltre 20 capi o grande, che hanno inizio lattazione nell'anno solare) con variazione annuale e cambiamento al 1° gennaio di ogni anno;
- o regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo ($\Delta nAllev_{(t_y-t_{y+4})}$): per ogni azienda è stata ottenuta la numerosità annua di capi allevati e la variazione percentuale rispetto alla consistenza aziendale dell'anno di registrazione al libro genealogico (anno 0). È stata calcolata una regressione lineare semplice sui valori di deviazione percentuale rispetto all'anno 0 in gruppi quadriennali contenenti l'anno per il quale è stato valutato il coefficiente di regressione e i 3 anni seguenti. Sulla base del segno del coefficiente di regressione (b)

sono state ottenute 3 classi: $b > 0$ (allevamento in espansione), $b < 0$ (allevamento in contrazione) e $b = 0$ (allevamento stabile). La figura 16 riporta un esempio di come è possibile rappresentare l'evoluzione di questo effetto di un allevamento: inizialmente è in espansione (nei quadrienni da 0-3 a 2-5), si stabilizza per tre anni (sulle proiezioni dei quadrienni da 3-6 a 5-8) e poi tende, anche se con diverse "intensità", a contrarsi (dal quadriennio 6-9 in poi).

L'interazione delle 4 classi di $nAllev(t)$ con le 3 tre di $\Delta nAllev_{(t, -t, y+4)}$ definisce per questo effetto 12 livelli;

- deviazione percentuale dell'EVM dalla media di allevamento-anno ($\Delta EVM(t)$): 12 livelli di cui 10 con ampiezza del 5% e due classi estreme aperte ($\leq 73\%$ e $\geq 134\%$).

L'effetto casuale tempo indipendente incluso nel modello è l'effetto genetico additivo a eredità poligenica dei tori padri delle vacche alle quali si riferiscono le lattazioni (*sire model*) per il quale si è assunta una distribuzione dei coefficienti di regressione u di tipo multivariato normale $u \sim MVN(0, A\sigma_s^2)$ nella quale A è la matrice di parentela tra tori e σ_s^2 è la varianza dell'effetto toro.

L'effetto casuale tempo dipendente, per il quale è stata assunta una distribuzione log-gamma, è l'interazione "spuria" di allevamento e anno. Dal parametro γ della distribuzione è stata ricavata la varianza dell'effetto $\sigma_{Allev(t)}^2 = trigamma(\gamma)$ (Ducrocq e Casella, 1996).

La combinazione della funzione di verosimiglianza con la distribuzione a priori normale per σ_s^2 e log-gamma per $\sigma_{Allev(t)}^2$ produce un'espressione proporzionale alla distribuzione a posteriori di ciascun effetto, mentre la stima delle componenti di varianza è stata ottenuta per approssimazione di Laplace della distribuzione marginale a posteriori (Ducrocq e Casella, 1996).

La stima dell'ereditabilità del carattere nei PHM è possibile utilizzando la formula individuata da Yazdi et al. (2002) che la rende indipendente dal parametro ρ , presente nelle formule individuate precedentemente per i modelli che utilizzano

Weibull sire model (Ducrocq, 1999). La formula proposta da Yazdi per l'approssimazione della stima dell'ereditabilità effettiva (h_{eff}^2) utilizzabile nei modelli a rischio proporzionale e derivata dall'espressione per il calcolo dell'attendibilità degli EBV in un sire model, descritta da Henderson nel 1975 è:

$$h_{eff}^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_{rnd}^2 + 1}$$

Per i modelli a regressione di Weibull l'attendibilità (R_{Wei}) della stima dei valori riproduttivi è definita dalla formula (Yazdi et al., 2002):

$$R_{Wei} = \frac{n_{uncen}}{n_{uncen} + \frac{1}{\sigma_s^2}} \text{ o l'equivalente } R_{Wei} = \frac{n\bar{p}}{n\bar{p} + \frac{1}{\sigma_s^2}}$$

nelle quali n_{uncen} è il numero di figlie *uncensored* di un toro, n e \bar{p} sono rispettivamente il numero dell'intera progenie di un toro e la proporzione di figlie riformate ($\bar{p} = n_{uncen} / n$). Da questa formula Yazdi et al. (2002) hanno derivato una formula di ereditabilità definita equivalente (h_{equ}^2):

$$h_{equ}^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_{Allev(t)}^2 + \frac{1}{\bar{p}}}$$

3.4. Bibliografia materiale e metodi

A.I.A. 1981. http://www.aia.it/subpages/controlli/reg_bov_lat.pdf

A.N.A.R.B. - 2006. Disciplinari del libro genealogico dei bovini di razza bruna. <http://www.anarb.it>

Ducrocq V., Casella G. - 1996. A Bayesian analysis of mixed serviva models. *Genet. Sel. Evol.* 28:505-529

Ducrocq V. - 2000. The Survival Kit V3.12 User's Manual. <http://www.nas.boku.ac.at>

Henderson C. R. - 1975. Rapid method for computing the inverse of a relationship matrix. *J. Dairy Sci.* 58:1727

Yazdi M. H., Visscher P. M., Ducrocq V., Thompson R. - 2002. Heritability, Reliability of genetic Evaluation and respponse to selection in proportional hazard models. *J. Dairy Sci.* 85:1563-1577

4. Primo contributo sperimentale:

ANALISI DEL RISCHIO E STIMA DEI PARAMETRI GENETICI

4.1. Introduzione

L'analisi di sopravvivenza basa la sua inferenza sulla possibilità di riuscire a descrivere con un modello matematico la dinamica della probabilità che un evento si manifesti. Dal punto di vista della genetica quantitativa, la flessibilità “dell'analisi generalizzata degli eventi storici” (Gill, 1994), sfrutta lo stretto legame esistente tra tre funzioni: quella di sopravvivenza (descrivibile con gli estimatori di Kaplan-Meier 1958) la sua derivata o funzione di densità di distribuzione e la funzione di rischio (Ducrocq, 1994).

La funzione di densità di probabilità cumulativa $F(t)$ è la complementare a 1 della funzione di sopravvivenza $S(t)$ (Kalbfleisch e Prentice, 1980):

$$S(t) = \text{Prob}(T \geq t) = 1 - \text{Prob}(T < t) = 1 - F(t) \quad \text{quindi} \quad F(t) = 1 - S(t)$$

ne consegue che la funzione di densità (density function) che è la derivata prima di $F(t)$ è anche il negativo della derivata della funzione di sopravvivenza,

$$f(t) = -\frac{dS(t)}{dt}$$

In particolare quello che ha indotto lo studio di questi legami è la possibilità di sfruttare le proprietà della funzione di densità di distribuzione $f(t)$ nell'applicazione del metodo della massima verosimiglianza per stimare il determinismo delle cause di un evento. Il legame che ha permesso di raggiungere questo obiettivo è la relazione esistente tra funzione di sopravvivenza, funzione di rischio e funzione di densità :

$$h(t) = \frac{f(t)}{S(t)}$$

La possibilità di descrivere “linearmente” la funzione di rischio ha aperto le strade all'applicazione dell'analisi della sopravvivenza in un'ampia gamma di settori, tra cui la genetica quantitativa per la stima delle componenti di varianza genetica quasi al pari dei modelli misti lineari (Ducrocq e Casella, 1986). Dopo aver definito la possibilità di:

- descrivere matematicamente la dinamica del rischio con modelli parametrici e non parametrici (Cox, 1972)

- implementare la funzione di verosimiglianza per modelli di rischio parametrici e semi-parametrici sfruttando anche le informazioni delle osservazioni *uncensored* (Cox, 1972)
- applicare l'integrazione di Laplace quale approssimazione dell'integrazione algebrica della distribuzione marginale a posteriori ottenuta da modelli *frailty* di analisi di sopravvivenza

la problematica principale attualmente in studio (Ducrocq, 2006) che rimane tuttora in fase di studio è l'applicabilità di un animal model sui dati di sopravvivenza.

4.2. Materiale e metodi

La base dati utilizzata è l'archivio dell'Associazione Nazionale degli Allevatori di razza Bruna (A.N.A.R.B.), il quale sottoposto a *data scrubbing* ha permesso di ottenere:

- carriere produttive complete per soggetti con data di riforma valida e sequenzialità in lattazioni successive per la definizione di dati *uncensored*
- sequenzialità delle lattazioni, a partire dalla prima, fino al verificarsi di un evento che determinasse la perdita di informazioni successive per i dati *censored*.

Gli eventi considerati limitanti nel considerare non più informative osservazioni aggiuntive sono:

- o animali in produzione: alle vacche vive e in produzione la data di *censoring* è la data dell'ultimo controllo giornaliero registrato.
- o vendita: tale evento non comporta necessariamente la perdita di informazioni successive. La frequenza dell'evento non giustifica la sua introduzione come effetto fisso tempo dipendente. La data di *censoring* corrisponde alla data di vendita.
- o mancanza di lattazioni intermedie: il passaggio da una lattazione alla successiva potrebbe essere determinante nella stima dei coefficienti di regressione (β) che altrimenti potrebbe risultare alterata da una prolungata (circa doppia) permanenza nella stessa classe di rischio. Quest'affermazione va opportunamente verificata in particolare in merito alla frequenza di questo evento. In questa situazione è stata attribuita come data di *censoring* la data dell'ultimo controllo giornaliero

disponibile dell'ultima lattazione in esatta sequenza e sono state eliminate le lattazioni successive a quella mancante.

Le informazioni censored assumo quindi la forma di *random censoring*, necessaria per l'applicazione di modelli a rischio proporzionale. Non sono stati considerate informazioni *left-censored*, relative, ad esempio, a capi acquistati all'estero. Sono stati definiti due caratteri, espressi in giorni: la lunghezza della vita (LV) ottenuta dalla differenza tra *failure* o *censoring date* e data di nascita e la lunghezza della vita produttiva (LPL) ottenuta per differenza di LV con l'età al primo parto (intervallo data di primo parto - *failure /censoring date*).

Survival analysis

LPL e LV sono state descritte inizialmente ottenendo la curva di sopravvivenza con gli estimatori di Kaplan-Meier (Kaplan et al. 1958). I dati sono stati analizzati con un Weibull *sire model* a rischio proporzionale utilizzando il Survival kit ver. 3.12 (Ducrocq, 2000). L'assunzione di Weibull è stata verificata graficamente per entrambe le variabili e per LPL è stata verificata anche stratificando i dati per ordine di parto.

Come descritto nel capitolo 2 il modello analizzato è:

$$h(t) = \rho t^{\rho-1} \exp[Et\grave{a}1P + Allev + OP(t) + \Delta EVM(t) + Hs(t) * \Delta Hs_{(Y \rightarrow Y+4)}(t) + toro + Allev * anno(t)]$$

dove:

$h(t)$ = rischio di riforma al tempo t

$\rho t^{\rho-1}$ = distribuzione di Weibull della *baseline hazard function* ($h_0(t)$) semplificata secondo Ducrocq e Casella (1996)

Effetti fissi tempo indipendenti:

- $Et\grave{a}1P$ = età al primo parto in classi trimestrali
- $Allev$ = allevamento (in cui è stata registrata la prima lattazione)

Effetti fissi tempo dipendenti:

- $OP(t)$ = ordine di parto (10 livelli: 1, 2, ..., 9 e ≥ 10)
- $\Delta EVM(t)$ = 12 livelli di scostamento percentuale dalla media dell'EVM aziendale annuo

- $Hs(t) * \Delta Hs_{(Y \rightarrow Y+4)}(t)$ = interazione tra dimensione annua dell'azienda e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo

Effetti casuali:

- $Allev * anno(t)$ = effetto tempo dipendente ottenuto dalla combinazione dell'azienda con l'anno solare nel quale è iniziata la lattazione. Per questo effetto casuale si è assunta una distribuzione log-gamma definita da un unico parametro γ . La varianza $\sigma_{Allev * anno}^2$ è stata ottenuta dalla trasformazione trigamma(γ).
- $toro$ = effetto casuale tempo indipendente per il quale si è assunta una distribuzione normale multivariata $MVN \sim (0, A \sigma_{toro}^2)$ nella quale A identifica la matrice di parentela tra tori e σ_{toro}^2 è la varianza dell'effetto.

I risultati sono stati analizzati per ogni effetto, verificando la normalità della distribuzione degli effetti casuali e identificando le motivazioni della distribuzione assunta dai rischi relativi ($\exp^{x'\beta}$) degli effetti fissi.

In merito all'effetto casuale tempo indipendente del toro sono stati calcolati i parametri genetici (Yazdi, 2002) ed è stato ottenuto il relativo trend genetico.

4.3. Risultati e discussione

La funzione di sopravvivenza, ottenuta con gli estimatori di Kaplan Meier, per i caratteri LPL e LV è riportata in figura 17. L'assunzione di *random censoring* distribuisce lungo la curva le informazioni censored (non evidenziate). La lunghezza massima della vita osservata è pari a 6800 giorni pari a quasi 19 anni di vita. LPL raggiunge un valore massimo di 5800 giorni pari a circa 16 anni di carriera produttiva. Traslando LPL di circa 950 giorni pari a 2 anni e 7 mesi, corrispondente alla media di età al primo parto, le due funzioni verosimilmente si sovrappongono. I dati analizzati contengono il 25 % di informazioni *censored*. La probabilità di sopravvivenza rispecchia una dinamica di sopravvivenza di una popolazione bovina (Terrés, 2005). La figura 18 riporta la funzione di rischio dei due caratteri. LPL manifesta un incremento più che proporzionale della probabilità che le vacche vengano riformate lungo la loro

carriera produttiva, mentre per LV si assiste a un punto di flesso attorno ai 3500 giorni (quasi 10 anni di vita).

La tabella 4.1 riporta una statistica descrittiva semplificata dei dati analizzati ponendo come end-time l'ordine di parto nel quale è stato osservato l'evento riforma.

Tabella 4.1: medie e d.s. di LPL e LV suddivise per ordine di parto di failure.

Ordine di parto di failure	LPL (gg)		LV (gg)	
	Media	d.s.	Media	d.s.
1	261	109	1'217	169
2	690	142	1'640	193
3	1'110	173	2'058	218
4	1'517	198	2'464	237
5	1'920	222	2'864	258
6	2'314	239	3'257	271
7	2'695	254	3'634	283
8	3'077	272	4'017	299
9	3'452	285	4'392	308
>=10	3'858	280	4'787	304

Applicando la verifica grafica (Ducrocq et al, 1988) è stato possibile verificare che per LV non è assumibile, quale baseline hazard function, la distribuzione di Weibull. Per LPL invece l'approssimazione ad una retta è del 99% permettendo tale assunzione (figura 19). La scelta della lunghezza della vita produttiva (LPL) rispetto alla lunghezza della vita (LV) è fortemente influenzata anche dal peso del tempo improduttivo che intercorre tra la data di nascita ed il primo parto. Considerando un'età media attorno ai 34 mesi di età al primo parto e un interparto ideale di 12 mesi, all'inizio della seconda lattazione il "peso" della durata della prima influisce solo per un 20% sul tempo totale mentre raggiunta la metà della terza lattazione si ha un peso relativo del periodo produttivo del 50% sulla vita totale. Considerando inoltre che non sono state considerate le cause di eliminazione del periodo improduttivo risulta più esplicativa un'analisi della lunghezza della vita produttiva. (LPL).

La verifica successiva, stratificando i dati per ordine di parto non ha dato lo stesso risultato (figura 20). L'assunzione di Weibull non è verificata per questa modalità di distribuzione dei dati. L'analisi con modello a rischio proporzionale con distribuzione di Weibull ha portato alla stima dei coefficienti di regressione β e quindi

degli *Hazard Ratio* ($HR = \exp^\beta$) degli effetti considerati. La stima del parametro ρ della distribuzione di Weibull del rischio basale ha determinato un valore pari a 1,86. Le tabelle che seguono riportano per ogni singolo effetto, i suoi livelli, la numerosità e la percentuale entro ogni livello di informazioni *uncensored* e il rischio di riforma relativo alla classe con la più elevata frequenza di osservazioni *uncensored* (evidenziata in grassetto per ogni effetto). Questo tipo di test di confronto è stato scelto per rapportare gli altri livelli a quello che manifesta maggior solidità (Ducrocq, 2000).

4.3.1. Effetti non genetici

In figura 21 sono riportati i rischi relativi delle classi dell'effetto età al primo parto con gli intervalli di confidenza al 95%. Ponendo rischio nullo alla classe 33-35 mesi di età al primo parto si delinea per le primipare “precoci” un rischio relativo inferiore e per le primipare “mature” un H.R. superiore a 1. Una manna che raggiunge il primo parto in età più matura, nonostante la sua conformazione fisica possa risultare più idonea a portare a termine una gestazione e la mammella mediamente sia più sviluppata, in termini di riforma volontaria ha una minore probabilità di sopravvivenza. L'incremento medio del rischio di riforma delle primipare “mature”, passando da una classe alla successiva è maggiore della riduzione del rischio acquisita dalla “precocità” di parto. Questo è dovuto probabilmente al fatto che l'allevatore attribuisce un ritardo nel primo concepimento (e quindi nel primo parto) come problema di fertilità della vacca, “predisponendola” tra le vacche potenzialmente riformabili per scelta volontaria, ma “condizionata” dalle scadenti performance riproduttive.

Tabella 4.1: livelli, numerosità e percentuale entro livello di informazioni *uncensored* e rischio relativo dell'effetto fisso *time-independent* età al primo parto.

Effetto	Mesi	N° <i>uncensored</i>	% <i>uncensored</i>	<i>Hazard ratio</i>
età al primo parto	18-23	2'652	60.4	0.84
	24-26	19'712	57.6	0.88
	27-29	39'905	58.0	0.92
	30-32	43'118	59.6	0.95
	33-35	43'837	62.6	1.00
	36-38	30'129	62.5	1.05
	39-41	8'424	62.3	1.14
	42-46	6'503	64.5	1.25

In figura 22 sono riportati gli *hazard ratio* ed i relativi intervalli di confidenza al 95% dell'effetto ordine di parto. In figura 23 è riportata la probabilità di sopravvivenza ottenuta dalla numerosità delle informazioni *uncensored* di ogni ordine di parto e la relativa variazione percentuale tra 2 livelli consecutivi. E' evidente come ci sia una riduzione del 50% tra le vacche che hanno *failure-time* entro la fine della prima lattazione (39%) e quelle che sono riformate in seconda lattazione (19%). Questo giustifica il valore di rischio relativo pari a 0,51 del secondo ordine di parto. A questo seguono delle variazioni percentuali relativamente omogenee tra ordini di parto successivi attorno al 30%. Riduzioni più consistenti, ma comunque inferiori al 50% tra 7° e 8° e tra 8° e 9° ordine di parto, data la ridotta numerosità degli eventi incide proporzionalmente meno sulla stima dei coefficienti di regressione e sui relativi *hazard-ratio*.

L'ultima classe (≥ 10) manifesta un incremento del rischio a 0,76 dovuto probabilmente all'invecchiamento delle vacche. Più è vecchio l'animale e meno sono gli anni che probabilmente rimarrà in azienda. Rientrano in questa classe vacche che hanno raggiunto e superato i dieci ordini di parto. Tra esse sono stati considerati animali che hanno raggiunto 16 lattazioni prima di essere riformati. Le variazioni di rischio relativo vanno dall'1% al 4% per ordini di parto compresi tra il 3° ed il 9°. L'entità degli errori standard e dei relativi intervalli di confidenza al 95% permette comunque di affermare che è possibile mantenere tale suddivisione e che livelli successivi sono significativamente diversi.

Tabella 4.2 livelli, numerosità e percentuale entro livello di informazioni uncensored e rischio relativo dell'effetto fisso *time-dependent* ordine di parto

Effetto	O.P.	N° <i>uncensored</i>	% <i>uncensored</i>	<i>Hazard ratio</i>
ordine di parto	1	75'039	23.3	1.00
	2	37'066	17.3	0.51
	3	27'319	18.4	0.38
	4	20'085	20.4	0.34
	5	13'951	22.5	0.33
	6	9'192	25.1	0.34
	7	5'671	28.0	0.36
	8	3'129	30.7	0.38
	9	1'652	35.0	0.34
	≥ 10	1'176	59.1	0.76

La scelta dell'impiego del dato EVM e non dei valori fenotipici registrati al termine della lattazione discende da una serie di considerazioni volte alla ricerca delle cause di riforma volontaria: il dato EVM equipara tutte le lattazioni ad una lunghezza di 305 giorni, non dipende dall'ordine di parto dell'animale stesso quindi un valore relativo ad una prima lattazione può essere direttamente confrontato con un valore di un qualsiasi altro ordine di parto. Cause che accidentalmente hanno ridotto la lunghezza di una lattazione (e la relativa produzione effettiva), ma che non hanno inciso sulla scelta di riforma di una vacca, proprio per questo motivo non vanno considerate nel modello d'analisi. Effetti legati all'ordine di parto sono già considerati nel relativo effetto. Utilizzare il dato fenotipico comporterebbe la sua introduzione nel modello di analisi in interazione con l'ordine di parto creando possibili ridondanze e alterazioni delle stime con quest'ultimo.

In figura 24 sono riportati gli *hazard ratio* ed i relativi intervalli di confidenza al 95% dell'effetto scostamento percentuale dalla media EVM di allevamento/anno. Il rischio di eliminazione si concentra sulle vacche poco produttive che risultano essere fino a oltre 4 volte più soggette a rischio di riforma rispetto agli animali in media di allevamento anno. Tale rischio rimane superiore al rischio di riferimento per le classi a percentuali inferiori alla media, per poi scendere a valori inferiori a 1 qualora le vacche vi si scostino positivamente. Questa variazione però non risulta essere della stessa intensità rilevata sulle vacche mediamente meno produttive. Ciò probabilmente può essere riconducibile al riconosciuto legame negativo esistente tra livello produttivo e "fertilità" delle vacche (Druet et al., 1999). Elevati livelli produttivi corrispondono mediamente a tassi di ritorno in calore più elevati, attribuendo alle cause di riforma volontaria una valenza diversa. Per questo motivo soggetti appartenenti alla classe $\geq 133\%$ presentano un rischio relativo di riforma maggiore rispetto alle vacche delle due classi precedenti che, visti gli intervalli di confidenza, manifestano lo stesso *H.R.*

Tabella 4.3 livelli, numerosità e percentuale entro livello di informazioni uncensored e rischio relativo dell'effetto fisso *time-dependent* variazione percentuale dalla media EVM di allevamento anno

Effetto	Livello	N° <i>uncensored</i>	% <i>uncensored</i>	<i>Hazard ratio</i>
variazione percentuale dalla media EVM di allevamento anno	<=73%	17'218	46.3	4.34
	74-79%	11'547	35.2	2.62
	80-85%	16'728	30.2	1.94
	86-91%	22'620	26.3	1.49
	92-97%	27'577	23.5	1.19
	98-103%	32'727	23.0	1.00
	104-109%	24'733	20.3	0.89
	110-115%	17'654	19.1	0.83
	116-121%	11'099	18.6	0.78
	122-127%	6'201	18.5	0.75
128-133%	3'208	18.8	0.74	
>=134%	2'968	22.6	0.77	

In figura 24 sono riportati gli *hazard ratio* ed i relativi intervalli di confidenza al 95% dell'effetto interazione tra dimensione annua dell'azienda e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo. Il maggior numero di informazioni *uncensored* proviene da allevamenti di grande dimensione, in termini assoluti una quota di riforma uguale tra allevamenti ha numerosità più elevate in allevamenti con più animali. Così come è verosimile combinando l'effetto col fatto che l'azienda sia contrazione: un allevamento definito grande che ha intenzione di ridurre o cessare la propria attività porterà a riforma un elevato numero di vacche incrementando notevolmente la quantità di informazioni uncensored relative a questa classe. L'effetto non è facilmente definibile, in quanto allevamenti posti al limite tra due classi possono cambiare livello da un anno all'altro. Lo stesso vale per il coefficiente di regressione stimato sulla variazione percentuale rispetto alla dimensione iniziale dell'azienda nel quadriennio che inizia con l'anno analizzato. In generale è possibile affermare che le vacche appartenenti a allevamenti stabili presentano il rischio relativo di riforma maggiore. Prossimi al valore nullo quegli animali presenti in aziende in contrazione e relativamente "protette" le vacche appartenenti a allevamenti in espansione. Il motivo di un valore di H.R. maggiore relativo agli allevamenti stabili è dovuto probabilmente ad una diversa strategia

di riforma. L'allevatore che intende contrarre la propria attività lascia in azienda gli animali fino a quando non sono più economicamente convenienti e non li rimpiazza. In un'azienda la cui numerosità resta stabile nel tempo invece la gestione della riforma è oculata e rivolta ai soggetti ritenuti problematici che sono sostituiti da manze provenienti da rimonta interna o, con maggior probabilità, da altri animali acquistati altrove.

Tabella 4.4: livelli, numerosità e percentuale entro livello di informazioni uncensored e rischio relativo dell'effetto fisso *time-dependent* interazione tra dimensione annua dell'azienda e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo.

Effetto	Livello	N° <i>uncensored</i>	% <i>uncensored</i>	<i>Hazard ratio</i>	
interazione tra dimensione annua dell'azienda e regressione per la dimensione "dinamica" dell'allevamento nel quadriennio successivo	Piccolo	contrazione	10'499	38.0	1.00
		stabile	6'053	26.6	1.36
		espansione	9'061	31.5	0.76
	medio- piccolo	contrazione	19'597	33.9	1.04
		stabile	4'779	18.4	1.07
		espansione	12'009	28.0	0.86
	medio grande	contrazione	29'758	33.2	1.01
		stabile	5'702	17.8	1.04
		espansione	17'139	27.9	0.91
	Grande	contrazione	46'610	37.6	1.00
		stabile	8'887	20.5	0.99
		espansione	24'186	32.2	0.91

La figura 25 riporta la distribuzione di frequenza dei coefficienti di regressione stimati per i 10'313 livelli, tolti i 185 allevamenti per i quali sono disponibili unicamente informazioni di tipo censored e per i quali non è stato possibile stimare un coefficiente β . La stessa figura riporta la distribuzione di frequenza dei relativi H.R.. Come si può vedere l'elevata variabilità dei β determina una lunga coda verso destra della distribuzione dei rischi relativi. La stima si distribuisce attorno a 0 riferendoci ai coefficienti di regressione, attorno a 1 relativamente agli HR.

4.3.2. Componenti di varianza e parametri genetici

L'effetto allevamento anno è stato assunto come effetto casuale tempo dipendente a distribuzione log-gamma. Il valore di γ per la descrizione della

distribuzione dell'effetto è 5,24. Tramite funzione trigamma, (Roxström et al., 2003) è stato ricavato il valore di varianza dell'effetto pari a 0,21. La figura 27 riporta la distribuzione di frequenza dei coefficienti di regressione q di questo effetto e la relativa distribuzione degli Hazard Ratio (\exp^q). Anche per questo effetto si nota una distribuzione attorno allo zero per i coefficienti di regressione e ovviamente attorno a 1 per i rischi relativi.

Per l'effetto legato al toro, e quindi alla componente genetico-additiva della popolazione, considerando l'introduzione della matrice di parentela (A) nel processo di stima e l'assunzione di una distribuzione a priori di tipo multivariata normale $u \sim \text{MVN}(0, A\sigma_s^2)$, è stato determinato una moda distribuzione marginale a posteriori $\sigma_s^2 = 0,04$.

Da questi valori tramite la formula proposta da Yazdi, 2002 è stato possibile calcolare l'ereditabilità effettiva del carattere

$$h_{eff}^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_{allev*anno}^2 + 1} = 12,8\%$$

Considerando che il 25% delle informazioni introdotte nell'analisi sono osservazioni censored il valore di ereditabilità si ridimensiona in funzione del parametro \bar{p} , ovvero della percentuale di osservazioni uncensored utilizzate nell'analisi:

$$h_{equ}^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_{allev*anno}^2 + \frac{1}{\bar{p}}} = 10,1\%$$

In figura 28 sono riportate la distribuzione di u e quella del rischio relativo associato (\exp^u). Gli *hazard ratio* hanno distribuzione normale con media 1 e deviazione standard 0,1. Dei risultati ottenuti è stata calcolata la media per anno di nascita degli *hazard ratio* dei tori e sono stati esclusi i valori antecedenti il 1983 e successive al 2001 che presentano una frequenza inferiore a 100 tori annui. Il trend genetico mostra una progressiva crescita del rischio di eliminazione da 0,90 a 1,05 in 13 anni, per subire una leggera contrazione di 0,07 punti di rischio nei successivi 7 anni. Il 1995, anno di cambio della tendenza del trend genetico è l'anno in cui il

carattere longevità indiretto è stato introdotto tra gli obiettivi di selezione. L'andamento antecedente al 1995 può essere giustificato dalla potenziale selezione indiretta risultate dal processo selettivo verso altri caratteri correlati negativamente con il carattere longevità.

Sempre Yazdi et al. (2002) elaborano una formula per il calcolo dell'attendibilità della stima nel caso sia utilizzato un modello a regressione di Weibull:

$$R_{Wei} = \frac{n\bar{p}}{n\bar{p} + \frac{1}{\sigma_s^2}}$$

È stato calcolato ad ogni toro il valore dell'attendibilità della stima e la distribuzione di frequenza è riportata in figura 30. In figura 31 viene riportato il trend dell'attendibilità contestualmente al valore registrato per quell'anno. Il valore medio è fortemente influenzato dalla presenza di tori con poche figlie. Facendo un raffronto tra attendibilità dei tori con più di 20 figlie e i tori con più di 20 figlie uncensored si ottiene una dinamica dell'attendibilità come riportata in figura 32. Con questa soglia è possibile notare come la numerosità dei tori scenda notevolmente con un rapporto di 1:3 se ci si raffronta sul numero totale di figlie e con rapporto di 1:5 se ci si riferisce alle figlie *uncensored*. Volendo descrivere una dinamica generale dell'attendibilità della popolazione sono state ipotizzate classi crescenti di numerosità di figlie da 20 a 300, parallelamente ad ogni classe sono state attribuite 10 classi di percentuale crescente di informazioni ed è stata determinata l'attendibilità di ogni possibile combinazione. I valori ottenuti sono riportati in figura 33.

Sta assumendo sempre maggior importanza, anche dal punto di vista dei confronti internazionali, la suddivisione dell'effetto lattazione in “*stage*”, in particolar modo per discriminare le differenze tra il rischio di riforma nella fase produttiva da quello in asciutta. Un approccio di questo tipo non è attualmente praticabile sull'intera popolazione di razza Bruna a causa della dimensione che assumerebbe la matrice di varianza covarianza per la stima delle soluzioni. da recenti sviluppi della presente analisi e in seguito a comunicazioni intercorse con le organizzazioni europee che si occupano di analisi genetica a livello europeo (Schneider, 2007) risulterebbe

avere un impatto di maggior interesse suddividere in classi annue lo scostamento percentuale dalla media EVM di allevamento anno. Ciò è verificabile qualora venisse valutata una differente variabilità dell'effetto nell'arco del tempo.

Questo approccio alla stima delle componenti di varianza del carattere longevità diretta è tuttora ampiamente studiato. A livello nazionale italiano per la razza Bruna è previsto di sviluppare il modello in un *sire maternal-grandsire model* per verificare l'effetto dell'introduzione nella matrice di parentela dei tori derivanti dalla linea femminile. A livello internazionale si sta lavorando al perfezionamento della definizione del modello di analisi. L'utilizzo della *survival analysis* è in pieno fermento, con molte controversie nate in merito a molti argomenti, non per ultimo la stima dei valori riproduttivi delle vacche. Questo approccio non permette infatti una stima accurata nel caso di *animal model* e soprattutto necessita di approfondimenti nella costruzione di matrici di varianza covarianza tra caratteri analizzati con altri approcci quali il modello lineare. Questa caratteristica verrà affrontata nel prossimo contributo sperimentale.

4.4. Bibliografia primo contributo sperimentale

Druet T., Sölkner J., Groen Ab F., Gengler N. – 1999. Improved genetic evaluation of survival using MACE to combine direct and correlated information from yield and functional traits. Proceedings International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT) - Longevity, Jouy-en-Josas (F), May 1999, Interbull Bulletin 21: 122-127.

Ducrocq V., Quaas, B., Pollak G., Casella G. – 1998. Length of productive life of dairy cows. Variance component estimation and sire evaluation. J. Dairy Sci. 71:3071-3079.

Ducrocq V. - 2000. The Survival Kit V3.12 User's Manual. <http://www.nas.boku.ac.at>

Ducrocq V., Delaunay I., Boichard D., Mattalia S. – 2003. A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations. Interbull Bulletin 30. Proceedings of The Interbull Technical Workshop BELTSVILLE, MD, USA March 2-3, 2003, 101 – 111

Fisher R. 1930. The genetical theory of natural selection. Oxford Clarendon Press

Falconer, D. S. – 1960. Introduction to Quantitative Genetics, Ed 4. Longmans Green, Harlow, Essex, UK.

Gill R. D. – 1994. Lectures on survival analysis. Springer Berlin / Heidelberg ISSN: 0075-8434 (Print) 1617-9692 (Online)

Henderson C.R. – 1953. "Estimation of variance and covariance components." Biometrics 9: 226-252.

Roxström A, Ducrocq V., Strandberg E. – 2003. Survival analysis in dairy cattle on a lactation basis. Genet. Sel. Evol. 35: 305-318

Terrés J. – 2005. Estudi de la longevitat i la supervivència fins al deslletament en bovins de muntanya del la raça buna dels pirineus. Doctoral thesis.

Thiele T. N. – 1889. Almindelig Iagttagelselar: Sandsynlighedsregning og mindste Kvadrates Methode (The general theory of Observations: Probability calculus and the method of least squares.). Reitzel, Kobenhavn (DK).

Schneider U. – 2007. Incontro di aggiornamento tra i genetisti della razza bruna dei paesi europei aderenti a Interbull. Comunicazione personale. Zug (CH), 12 December, 2007.

Yazdi M. H., Visscher P. M., Ducrocq V., Thompson R. - 2002. Heritability, Reliability of genetic Evaluation and respponse to selection in proportional hazard models. J. Dairy Sci. 85:1563-1577

5. Secondo contributo sperimentale:

L'INDICE GENETICO LUNGHEZZA DELLA VITA PRODUTTIVA

E LE SUE CORRELAZIONI GENETICHE CON GLI ALTRI CARATTERI

5.1. Introduzione

Nel capitolo introduttivo è già stata evidenziata l'importanza economica di effettuare selezione genetica sul carattere longevità. Nel primo contributo sperimentale in accordo con tutti gli studi che utilizzano la survival analysis per ottenere stime dei valori riproduttivi è emersa la tarda disponibilità di stime attendibili che associata ad una medio-bassa ereditabilità (Druét et al 1999)

La stima delle correlazioni genetiche è stata ampiamente studiata con lo sviluppo dei modelli lineari. Alcuni software utilizzati in ambito genetico-zootecnico permettono l'analisi contemporanea di più caratteri e sono in grado di stimare congiuntamente le componenti di varianza - covarianza, e quindi le correlazioni genetiche ed i valori riproduttivi (Groeneveld 1990 e 1993) di più caratteri contemporaneamente. Questo non è ancora operativamente possibile utilizzando la Survival Analysis. Allo stesso modo non è disponibile alcuna soluzione per la stima contemporanea delle componenti di varianza – covarianza per più caratteri analizzati con i due diversi approcci. In questo contributo sperimentale si propone di adattare la procedura MACE (Schäffer 1993 – 1994) già utilizzata in altri studi per scopi analoghi (Ducrocq et al., 2001).

Relativamente alla razza Bruna Italiana le uniche correlazioni attualmente a disposizione sono state ottenute da Cassandro (1999) e successivamente ristimate da A.N.A.R.B. nel 2005 e riportate in tabella 5.1.

Tabella 5.1: correlazioni genetiche tra longevità funzionale e altri caratteri

Carattere	rFPL*	rFPL*
	1999	2005
Punteggio finale	0,04	-0,08
Statura	-0,13	-0,22
Forza vigore	0,01	-0,07
Profondità	-0,03	-0,10
Angolosità	-0,17	-0,17
Linea dorsale	-0,02	-0,10
Angolo groppa	-0,01	0,21
Arti visti di lato	-0,29	-0,56
Pastoie	0,02	0,15
Attacco anteriore della mammella	0,15	0,10
Larghezza attacco posteriore mammella	-0,09	-0,08
Altezza attacco posteriore mammella	-0,04	-0,10
Legamento centrale mammella	0,12	0,08
Profondità mammella	0,43	0,42
Lunghezza dei capezzoli	-0,42	-0,28

*FPL: functional productive life.

5.2. Materiale e metodi

È stata analizzata la correlazione esistente tra gli indici genetici di longevità diretta e longevità indiretta, suddividendo i dati in classi di numerosità di figlie. Nell'analisi successiva, oltre ai valori riproduttivi stimati per il carattere LPL, sono stati utilizzati gli EBV ottenuti nel giro indici di routine Interbull (Interbull, 2007) relativo ad agosto 2007 per tutti i caratteri riportati nella tabella 5.2 nella quale sono riportati anche media e alla deviazione standard sulla base delle quali sono espressi i valori riproduttivi stimati. Per tutti i caratteri riportati in tabella (ad eccezione di LPL) viene utilizzato un single-trait REML BLUP animal model. Nel caso dei caratteri produttivi le stime vengono ottenute con un modello ad informazioni ripetute, mentre per i restanti caratteri il modello è ad osservazioni singole (A.N.A.R.B., 2007a). A fini esplorativi viene anche introdotto il carattere longevità indiretta come attualmente calcolato a partire da 5 caratteri morfologici, come già descritto nel primo capitolo del presente lavoro. LPL è espresso come hazard ratio, esprimendo quindi un valore inverso alla reale longevità.

Tabella 5.2 Caratteri inseriti nell'analisi MACE

Carattere	Record	Ereditabilità	Media	Deviazione standard genetica
Caratteri funzionali				
Longevità indiretta	1'163	21%	100,000	12,000
Longevità diretta	1'594	13%	0,959	0,113
Body Condition Score	1'163	19%	-0,101	0,088
Somatic Cell Score	1'086	8%	-0,016	0,479
Velocità di mungitura	886	20%	0,031	0,210
Caratteri morfologici				
<i>Punteggio finale</i>		24%	1,413	1,247
Statura		32%	2,030	3,276
Forza vigore		17%	-0,129	1,955
Profondità toracica		18%	0,717	2,116
Angolo della groppa		21%	-0,190	1,723
Angolosità		18%	3,944	2,778
<i>Arti e Piedi</i>		11%	0,282	0,827
Arti visti di lato		11%	0,191	1,410
Altezza dei talloni	1'163	7%	-0,137	1,090
Forza delle pastoie		9%	-0,044	1,530
<i>Indice complessivo mammella</i>		16%	100,000	12,000
Forza dell'att. ant. mamm.		14%	2,231	2,238
Larghezza att. post. mamm.		16%	2,918	2,617
Altezza att post mamm.		16%	2,482	2,152
Forza del legamento centrale		13%	0,765	1,936
Profondità della mammella		22%	0,115	2,235
Lunghezza dei capezzoli		30%	-0,393	2,684
Direzione capezzoli		14%	0,290	1,663

Successivamente alla stima delle componenti di covarianza genetica tra i caratteri sono state eseguite prove separate volte a ottenere nuove stime di una longevità combinata e di attendibilità della nuova stima che sono state confrontate con i valori di input determinando le correlazioni medie tra gli indici genetici e poi suddividendo i dati per anno di nascita dei tori.

Le prove condotte separatamente hanno riguardato:

- A) i caratteri funzionali (B.C.S., S.C.S., Velocità di Mungitura).
- B) i caratteri legati alla struttura (statura, forza vigore, profondità toracica, angolo della groppa, angolosità)

- C) i caratteri legati ad arti e piedi (complessivo arti e piedi, altezza talloni, arti posteriori visti di lato, forza delle pastoie).
- D) i caratteri legati alla morfologia della mammella (indice complessivo mammella, forza dell'attacco anteriore, legamento centrale, larghezza dell'attacco posteriore, altezza dell'attacco posteriore, profondità della mammella, lunghezza dei capezzoli, direzione deicafezzoli)
- E) i punteggi complessivi (arti e piedi e complessivo mammella) con il punteggio finale e i kg di latte.
- F) i caratteri utilizzati attualmente dall'Associazione Nazionale di razza per il calcolo dell'indice longevità indiretta (statura, angolo della groppa, arti visti di lato, forza dell'attacco anteriore della mammella, profondità della mammella).
- G) i 10 caratteri che dalla stima delle correlazioni genetiche tra tutti i caratteri sono risultati essere i più correlati (B.C.S., statura, forza vigore, profondità addominale, arti visti di lato, forza del legamento centrale, profondità della mammella, lunghezza dei capezzoli, angolosità e S.C.S.).

5.3. Risultati e Discussione

Da un calcolo della correlazione di LPL con l'indice longevità indiretta è stata ottenuta una correlazione di -58%. L'intensità della correlazione è giustificata osservando la figura 34 che riporta le stesse correlazioni calcolate suddividendo i tori in classi di numerosità di figlie. Il passo successivo ha visto la stima contemporanea delle correlazioni genetiche tra tutti e 26 i caratteri oggetto di studio. Le correlazioni di LPL con gli altri caratteri sono riportate nella tabella 5.11 e rispecchiamo quanto descritto nei prossimi paragrafi relativi alle stime condotte sui gruppi specifici di caratteri descritti in Materiale e Metodi di questo capitolo. Le tabelle relative alle correlazioni stimate sono riportate come appendice successivamente al capitolo conclusivo.

Caratteri funzionali: le correlazioni ottenute in questa prova sono riportate in tabella 5.3. Per il carattere S.C.S. sono desiderabili valori alti. Relativamente ai caratteri funzionali la correlazione con l'indice S.C.S. supera di quasi 15 punti il valore massimo di correlazione attesa compresa tra 0,06 e 0,30 (Rogers et al, 1996). La

correlazione ottenuta rispecchia invece il valore ottenuto da van del Linde et al. (2007).

Caratteri legati alla struttura : le correlazioni ottenute in questa prova sono riportate in tabella 5.4. le correlazione ottenute hanno intensità superiori rispetto allo studio di Cassandro et al. (1999) riportate in tabella 5.1. Si conferma una relativa indipendenza genetica col carattere angolo della groppa. La correlazione con l'angolosità raddoppia passando da -0,17 a 0,35 (da ricordare l'inversione dell'indice) e triplica l'intensità della correlazione con la statura, da -0,13 a 0,34. Tra i caratteri lineari si conferma l'indipendenza genetica, considerata anche nell'armonizzazione nelle valutazioni morfologiche europee , del carattere angolo della groppa.

Caratteri legati ad arti e piedi : le correlazioni ottenute in questa prova sono riportate in tabella 5.5. Non si evidenziano correlazioni con valori elevati. Risultano ovviamente molto intensi i legami tra i caratteri relativi ad una categoria omogenea. Si riduce drasticamente la correlazione con il carattere arti visti di lato che passa da un -0,56 a un 0,15 (41 punti percentuali). Per il carattere forza delle pastoie si conferma la correlazione ottenuta da Cassandro et al. (1999) pari a 0,02.

Caratteri legati alla morfologia della mammella : le correlazioni ottenute in questa prova sono riportate in tabella 5.6. i caratteri forza del legamento centrale, lunghezza dei capezzoli e larghezza dell'attacco posteriore della mammella confermano i legami genetici che avevano con rFPL 2005. Il carattere altezza dell'attacco della mammella conferma e accentua la sua indipendenza genetica con LPL passando (da -0,10 a 0,02 $\Delta r=0,08$). Il legame con la profondità della mammella è risultato inferiore rispetto a quello stimato precedentemente.

Indici complessivi : le correlazioni ottenute in questa prova sono riportate in tabella 5.7. non sono disponibili raffronti con studi simili eseguiti sulla razza Bruna in Italia, fatta eccezione per il punteggio finale che mantiene un legame modesto con LPL. La correlazione maggiore è con il carattere indice complessivo mammella e pari a 0,22. Come ci si può attendere si ritrovano correlazioni molto strette tra Punteggio finale, i. complessivo mammella e arti e piedi, superiori a 0,70.

Caratteri utilizzati attualmente dall'Associazione Nazionale di razza per il calcolo dell'indice longevità indiretta : le correlazioni ottenute in questa prova sono

riportate in tabella 5.8. analogamente a quanto descritto nei paragrafi precedenti, le correlazioni assumono le stesse intensità riportate in precedenza, con variazioni minime dovute probabilmente ad approssimazioni o alla diversa struttura della matrice che è stata sviluppata nel processo di calcolo. Perderebbero sicuramente interesse caratteri quali angolo della groppa e forza dell'attacco anteriore della mammella. Arti visti di lato riduce la sua importanza nel legame con LPL, ma risulta ancora essere, come descritto nel paragrafo successivo, tra i 10 caratteri maggiormente correlati.

I 10 caratteri che dalla stima delle correlazioni genetiche tra tutti i caratteri sono risultati essere i più correlati : le correlazioni ottenute in questa prova sono riportate in tabella 5.9. le correlazioni con intensità maggiore, indipendentemente dal segno della correlazione sono: B.C.S., statura, forza vigore, profondità addominale, arti visti di lato, forza del legamento centrale, profondità della mammella, lunghezza dei capezzoli, angolosità e S.C.S.. Si evidenzia subito che angolosità e B.C.S. hanno una correlazione reciproca pari a -0,76 e che come espresso nel corrispondente paragrafo, anche i caratteri relativi alla struttura hanno legami genetici molto alti. Questi legami sono evidenziati nella tabella con bordatura nera marcata. Si conferma il valore -0,43 il legame con il carattere Somatic Cell Score.

Parallelamente alla matrice di correlazione genetica, per tutte le analisi condotte, sono stati stimati i BV per ciascun riproduttore. La colonna riportante rLPL input della tabella 5.10 riporta le correlazioni tra indici genetici relativi a LPL e LPL_combinata con i vari gruppi di caratteri. Il livello di correlazione identifica grossomodo l'entità del cambiamento. Maggiore è la correlazione media tra i caratteri con LPL e maggiore sarà l'entità della variazione dell'EBV, come si può evidenzia con i caratteri più correlati. Viceversa la mancanza di legame genetico tra un carattere e LPL non determina sostanziali cambiamenti nella stima dei "nuovi" valori riproduttivi. Si evidenzia anche una relazione strettamente positiva, fatta eccezione per l'analisi condotta sui caratteri complessivi tra la correlazione degli indici in input e in output e la media di incremento dell'attendibilità. Sempre in tabella 5.10 si può osservare che la percentuale di tori per i quali le informazioni aggiuntiva migliorano l'attendibilità del carattere oscilla tra il 63% e il 94%, con incrementi che mediamente possono anche raddoppiare. Andando ad indagare la suddivisione per anno di nascita

l'incremento di attendibilità della stima di LPL di input e output nelle diverse analisi, si osserva, prima in tabella 5.12 e successivamente in figura 34 limitatamente ai tori nati dopo il 1995 che l'incremento è costante sugli animali più vecchi, mentre tende a ridursi a partire dalla metà degli anni '90 assumendo, nel 1997 per l'analisi con i caratteri più correlati e nel 2000 per quasi tutte le altre, un valore negativo. L'anno 2002 non è valorizzato per una ridotta consistenza del campione.

In tabella 5.11 e successivamente in figura 35 per i dati successivi al 1995 sono invece riportate le correlazioni tra gli indici genetici relativi a LPL dei set di dati in input e quelli dei set di dati in output suddivise per anno di nascita dei tori. Prima del 1995 le stime ottenute sono sovrapponibili al 99%. Come confermato da Druét et al (1999), la correlazione tra i due valori tende a ridursi facendo riferimento ad animali più giovani, determinando, con la combinazione, nuove stime sensibilmente diverse.

5.4. Bibliografia secondo contributo sperimentale

A.N.A.R.B. – 2007a. www.anarb.it

A.N.A.R.B. – 2007b. Incontro europeo dei punteggiatori per l'armonizzazione delle valutazioni morfologiche. Zug (CH)

Cassandro M., Ghiroldi S., Bagnato A., Gallo L., Carnier P., Santus E. – 1998. Genetic evaluation of longevity in Italian brown cattle breed. Proceedings international workshop on genetic improvement of functional traits in cattle (GIFT). Jouy-en-Josas (F) May 1999. Interbull Bulletin 21: 73-81

Druet T., Sölkner J., Groen Ab F., Gengler N. – 1999. Improved genetic evaluation of survival using MACE to combine direct and correlated information from yield and functional traits. Proceedings International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT) - Longevity, Jouy-en-Josas (F), May 1999, Interbull Bulletin 21: 122-127.

Ducrocq V. Boichard D., Barba A., Larroque H. – 2001. 52nd Annual Meeting of the European Association for Animal Production – August 26-29, 2001. Budapest, Hungary

Groeneveld E, 1990: PEST User's Manual. Institut fuer Tierzucht und Tierverhalten Mariensee Neustadt (D).

Groeneveld E, 1993: VCE User's Manual. Institut fuer Tierzucht und Tierverhalten Mariensee Neustadt (D).

Interbull – 2007. <http://www-interbull.slu.se/eval/framesida-genev.htm>

Rogers g. W., Banos G., Sande-Nielsen U., Philipsson J. – 1996. Genetic correlation among somatic cell score, productive life and type traits from the United States and udder health measures from Denmark and Sweden. Interbull Bulletin 14:34-38.

Schäffer L. R. – 1993. Multi-trait Across Country Evaluation of Dairy Sires. Interbull Bulletin 8:

Schäffer L. R. – 1994. Multiple country comparison of dairy sires. J. Dairy Sci. 77:2671-2678

van der Linde C., Harbers A.G.F., de Jong G. – 2007. From functional longevity to productive longevity in the Netherlands. Proceedings of the Interbull Meeting, Dublin, Ireland, August 23-26, 2009, Interbull Bulletin 37: 203 – 207.

6. Conclusioni e prospettive future

Dai risultati ottenuti nel primo contributo sperimentale è possibile delineare con gli effetti fissi non genetici le cause di riforma volontaria in un allevamento: il livello produttivo (variazione percentuale dalla media EVM di allevamento anno), in parte, con l'età al primo parto, la fertilità della vacca, l'affidabilità che una vacca riesce a farsi attribuire dall'allevatore (ordine di parto) e la prospettiva dell'allevatore stesso di espandere o meno la propria attività. Tutti gli effetti sono risultati altamente significativi ($P < 0,0001$) e il modello rispecchia quelli utilizzati in altre realtà internazionali. La stima delle componenti di varianza e i relativi parametri genetici sono in accordo con quelli delle altre nazioni. Questo modello per la stima degli EBV può quindi essere applicato per l'ottenimento degli indici genetici della popolazione di tori. I punti sui quali sarà necessario indagare in seguito a questo studio riguardano:

- verificare se l'effetto variazione percentuale dalla media EVM di allevamento anno non possa essere considerato in classi annuali, assumendo, ad esempio, che in un determinato allevamento il rischio associato allo scostamento del 10% dalla media produttiva di un preciso anno, sia significativamente diverso dallo scostamento del 10% dalla media produttiva di un altro anno (Schneider, 2007).
- identificare la miglior tecnica adottabile per ottenere dei valori riproduttivi di LPL attendibili nella popolazione delle vacche (Ducrocq, 2001).
- studiare la possibilità, soprattutto in merito alla numerosità, di suddividere le lattazioni in stadi, tracciando la dinamica del rischio nei diversi momenti della lattazione (Ducrocq, 1999 – Vukasinovic et al., 2001 – Terawaki et al., 2006).

Relativamente al secondo contributo sperimentale è possibile affermare che al fine di ottenere una stima precoce degli indici di selezione per i tori per non aumentare l'intervallo di generazione la scelta della combinazione da adottare e quindi il carattere LPL combinato risultante sicuramente manifesterà valori più discordanti con l'indice diretto relativamente ai tori giovani, mantenendo inalterato sostanzialmente inalterato l'indice genetico per i tori più datati. L'aspetto più interessante risulta l'identificazione della migliore combinazione che porti a valori di LPL combinati, non completamente avulsi dalla stima originale ottenuta con la survival

analysis, ma accettabilmente correlati, Druét et al. (1999) mostrano correlazioni minime del 67,4% tra i valori di LPL diretta e quelli di LPL combinata e contemporaneamente determinino un incremento sostanziale dell'attendibilità della stima. Sempre Druét et al. (1999) mostrano incrementi, nei tori più giovani di 13 punti percentuali di attendibilità. Con le analisi proposte, escludendo le soluzioni controproducenti ottenute con i caratteri complessivi, l'incremento di attendibilità riscontrato, soprattutto per i tori giovani, consiglia di ricercare altre soluzioni per la combinazione con altri caratteri. Tra le soluzioni proposte, la scelta sarebbe vincolata alla combinazione LPL_arti e piedi con correlazioni del 99% tra indici genetici LPL_input e LPL_output e con la minor riduzione di incremento di attendibilità rispetto alle altre soluzioni proposte. Scelta la combinazione più adatta per raggiungere una buona attendibilità sufficiente anche per i tori sarà il momento di validare la procedura al fine di concedere all'Associazione Nazionale di razza, la possibilità di scambio internazionale dei dati anche per questo carattere.

6.1. Bibliografia Conclusione e prospettive future.

Ducrocq V. – 1999. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters. Proceedings International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT) - Longevity, Jouy-en-Josas, (F) May 1999. Interbull Bulletin 21: 60-67

Ducrocq V. – 2001. A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life. Proceedings of the 2001 Interbull Meeting in BUDAPEST, (H) August 30-31 2001. Interbull Bulletin 27: 147-152

Ducrocq V. – 2004. Illustration of a trend validation test for longevity evaluations. Proceedings of the 2004 Interbull meeting Sousse, Tunisia May 29-31, 2004. Interbull Bulletin 32151 - 156

Terawaki Y., Katsumi T., Ducrocq V. – 2006. Development of a survival model with piecewise Weibull Baselines for the analysis of length of productive life of holstein cows in Japan. J.Dairy Sci. 89: 4058-4065

Vukasinovic N., Moll J., Casanova L. – 2001. Implementation of a routine of genetic evaluation for longevity based on survival analysis techniques in dairy cattle population in Switzerland. J.Dairy Sci. 84: 2073-2080.