

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA MAGISTRALE IN
SCIENZE STATISTICHE



**Analisi dei fattori genetici e ambientali che influiscono
sul colore del piumaggio di un gruppo di uccelli selvatici.**

Relatore Prof. Livio Finos
Dipartimento di Scienze Statistiche

Correlatori:

Dott. Riccardo Ton
Prof. Gianmarco Altoè
Prof. Simon Griffith

Laureanda/o Francesco Frizzarin
Matricola 2062880

Anno Accademico 2023/2024

Indice

Introduzione	1
1 Dal piano sperimentale ai dati raccolti	5
1.1 Esperimento	5
1.2 Dataset	7
2 Processamento dei dati	9
2.1 Pre-processing e analisi valori mancanti	9
2.1.1 Creazione di nuove variabili	10
2.1.2 Scelta delle variabili che descrivono l'ambiente del nido	14
2.1.3 Standardizzazione delle variabili quantitative	16
2.2 Creazione del dataset per lo studio della probabilità di sopravvivere	16
2.3 Creazione dataset per l'analisi del colore	18
3 Modelli	21
3.1 Studio della probabilità di sopravvivere	21
3.1.1 Analisi esplorativa	24
3.1.2 Modello glmm per la probabilità π di sopravvivere alla prima stagione riproduttiva	28
3.2 Analisi del colore	31
3.2.1 Analisi esplorativa	33
3.2.1.1 Analisi esplorativa per <i>headpc1</i>	33
3.2.2 Modelli lmm per <i>headpc1</i>	38
3.2.3 Modelli lmm per <i>chestpc1</i>	41
3.2.4 Modelli lmm per <i>wingpc1</i>	43
3.2.5 Analisi dei residui	44
4 Modelli finali	49
5 Conclusioni	55
5.1 Conclusioni per lo studio della probabilità di sopravvivere	55
5.2 Conclusioni per l'analisi del colore	55

Introduzione

La colorazione del piumaggio di alcuni uccelli passeriformi, come le cinciarelle, è stata per anni oggetto di studio in campo Biologico (Parker et al., 2011).

Una domanda che molti ricercatori nel campo della Biologia stanno ancora cercando di rispondere è quali siano i benefici che le femmine ricevono dall'accoppiamento con maschi maggiormente ornati e se questi benefici derivino esclusivamente da fattori genetici (quindi dalla scelta del compagno) oppure se questi possano essere influenzati anche da fattori ambientali (Hadfield et al., 2006).

Con questo lavoro si è voluto invece spostare il focus dalla selezione sessuale al colore di questa specie di uccello e alla sua probabilità di sopravvivere. Sono stati quindi analizzati a questo scopo quali fossero i “driver” principali e in quale misura ognuno di questi portasse le cinciarelle ad avere un ornamento migliore oltre che a sopravvivere con maggiore probabilità alla prima stagione riproduttiva, cercando di tener conto di entrambi gli effetti: genetici e ambientali. Una sfida raramente soddisfatta infatti nelle popolazioni selvatiche è quella di separare efficacemente i benefici genetici da quelli ambientali (Hadfield et al., 2006).

La colorazione del piumaggio della cinciarella è un ornamento sessuale ampiamente studiato (Andersson et al., 1998; Hunt et al., 1998; Sheldon et al., 1999) e diversi sono i fattori ambientali che possono concorrere assieme a quelli genetici allo sviluppo del colore di questi uccelli, per esempio: l'area della foresta dove risiedeva il nido, la densità di altri nidi all'interno dell'area, la distanza del nido dove cresce il pulcino dal nido più vicino, la competizione derivata dal numero di fratelli con cui è cresciuto il pulcino, il livello di cure parentali che l'uccello ha ricevuto durante i primi 14 giorni di vita o banalmente l'annata della covata, poiché nei fattori ambientali ricadono anche tutti i fenomeni atmosferici che colpiscono l'area in quello specifico anno (Parker et al., 2011).

Tutta la ricerca sulla trasmissione genetica dei tratti animali è stata approfondita non senza difficoltà dai ricercatori durante il corso degli anni (Parker et al., 2011). Questo perché spesso soprattutto per alcune specie animali difficili da osservare dal punto di vista statistico-sperimentale come nel nostro caso gli uccelli selvatici, non si hanno a disposizione molti dati, essendo il monitoraggio degli uccelli selvatici molto difficile e costoso al lato pratico, rendendo difficile e talvolta pericoloso lo sviluppo di una solida base statistica sulla quale fondare teorie biologiche affidabili.

Domande come queste, che dal punto di vista della ricerca in campo Biologico sono essenziali, risultano dal punto di vista statistico molto complesse da risolvere, infatti necessitano di molta attenzione e cura di ogni dettaglio: dalla stesura di un corretto disegno sperimentale, fino alla corretta e scrupolosa interpretazione dei risultati, che essendo lo scopo principale del lavoro di uno statistico in questo caso non possono esimersi dalle rigide logiche del disegno sperimentale e della Biologia stessa, oltre che da come sono stati raccolti e memorizzati i dati .

Da studi precedenti in questo campo come (Parker et al., 2011; Hadfield et al., 2006; Andersson et al., 1998), sembra emergere che il colore delle cinciarelle sia una forma di selezione per la fecondità, ma che questa dipenda anche da altri fattori di origine ambientale in diversa misura.

Con questo lavoro si è cercato quindi di sfruttare al meglio la capacità informativa dello stesso dataset utilizzato in (Parker et al., 2011), per capire invece se il colore della cinciarella, sia per i maschi sia per le femmine, dipenda e in quale misura, dal colore dei genitori e/o anche dai medesimi fattori ambientali considerati da (Parker et al., 2011; Hadfield et al., 2006; Andersson et al., 1998), oltre che da un nuovo fattore che considereremo in questo lavoro di origine sperimentale che ha sempre a che fare con l'ambiente ma deriva da una manipolazione delle vite naturali di questi animali e fa parte di un esperimento di cui parleremo nel seguito. Gli obiettivi di questo lavoro quindi si divideranno in due parti: da un lato si vuole capire come e in quale misura contribuiscano all'intensità del colore i fattori dell'ambiente dove nasce e cresce il pulcino, i fattori genetici e la manipolazione sperimentale, d'altra parte si vuole studiare quali fattori contribuiscano maggiormente alla morte precoce di questa specie.

Sono emersi numerosi risultati importanti, sia per quanto riguarda l'analisi del colore sia per quanto riguarda lo studio della probabilità di sopravvivenza. In particolare per lo studio del colore è emerso che ogni parte del corpo reagisce ai diversi fattori ambientali e genetici in diverso modo e solo una delle tre parti del corpo sembra essere

influenzata dalla manipolazione prevista dall'esperimento. Lo studio della probabilità di sopravvivere ha invece condotto a risultati in linea con quelli riportati da studi simili eseguiti su altre specie di uccelli come (Griffith, 2000).



FIGURA 1: Cinciarella selvatica (*Cyanistes caeruleus*).

fonte: Wikipedia, autore: Sławek Staszczuk.

Capitolo 1

Dal piano sperimentale ai dati raccolti

1.1 Esperimento

Per far fronte alle domande di ricerca, l'esperimento su cui si basa il dataset in questione prevedeva la costruzione di 1100 nidi artificiali oltre a quelli già presenti nel sito e il monitoraggio di tutti i relativi uccelli, su cui sono stati posti dei braccialetti sulle zampe per il loro riconoscimento nelle diverse fasi di vita, prima e dopo una possibile manipolazione delle loro vite naturali (Parker et al., 2011).

Per capire se ci sia un effetto del grado di cure parentali sul colore del piumaggio di questi uccelli, alcuni di questi (circa il 30%) sono stati manipolati da piccoli (attorno al secondo giorno a partire dalla schiusa delle uova), ossia sono stati spostati su un nido diverso da quello natale che poteva essere di dimensione maggiore, uguale o minore (Hadfield et al., 2006). Per altri invece la manipolazione è avvenuta esclusivamente una volta diventati adulti, ossia sono stati spostati dal proprio nido i relativi pulcini appena procreati. Spesso però gli uccelli essendo catturati per la rilevazione del colore in modo casuale dalla foresta più volte nel corso della loro vita hanno subito entrambi i tipi di manipolazione.

L'esperimento inizialmente prevedeva inoltre altri tipi di manipolazione (Hadfield et al., 2006) ma in questo lavoro ci concentreremo esclusivamente sull'effetto della manipolazione dei pulcini nella loro fase iniziale di vita come in Parker et al. (2011), e di come questo effetto abbia o meno poi delle conseguenze a parità di altre variabili sul loro

colore nel corso dell'anno successivo. Escluderemo quindi dall'analisi i soggetti adulti e di conseguenza l'effetto della manipolazione dei nidi sul colore degli adulti.

Per rispondere a domande come: la manipolazione dei nidi influisce anche sul colore dei genitori oltre che alla colorazione dei pulcini una volta cresciuti? Si sarebbe dovuto avere a disposizione un numero maggiore di rilevazioni sui soggetti adulti. Purtroppo a causa di una quantità limitata di dati sui pulcini sopravvissuti a più di una stagione riproduttiva si è ritenuto necessario concentrarsi esclusivamente sullo studio del piumaggio dei pulcini durante il primo anno di vita e della probabilità di sopravvivere a questo primo anno.

Per capire invece le cause principali di una possibile morte prematura, è stata rilevata per ogni uccello una variabile dicotomica che indica se l'animale è sopravvissuto o meno alla prima stagione riproduttiva ed è stata messa in relazione oltre che con alcune variabili ambientali anche con delle variabili che misurano quantitativamente l'aumento di stress causato agli uccelli da parte dalla manipolazione artificiale dei nidi.

E' possibile quindi, anche se non proprio scontato, tramite un disegno sperimentale di questo tipo riuscire a isolare i possibili effetti ambientali che descrivono le caratteristiche dell'ambiente dove sono nati e cresciuti gli uccelli, dall'effetto genetico di base dovuto alla parentela; sia per quanto riguarda l'analisi del colore sul gradiente UV (misurato tramite uno spettrometro con illuminazione al xeno seguendo la metodologia ideata e presentata da Hadfield et al. (2006) nel loro studio sulla stessa specie), sia sulla probabilità di sopravvivere alla prima stagione riproduttiva.

Tutto ciò è reso possibile grazie a diversi accorgimenti: prima di tutto le misurazioni sono avvenute in maniera randomica, ossia per ogni evento di misurazione sono stati rilevati casualmente degli uccelli dalla foresta. In secondo luogo per ognuno di essi è quasi sempre avvenuta la registrazione di un codice alfanumerico unico che identifica la covata da cui proviene la sua nascita, un codice identificativo per ogni covata in cui si è trovato a crescere nel caso sia stato spostato dal suo nido natale e infine un codice che identifica in quale covata è riuscito a procreare, nel caso fosse riuscito a sopravvivere al primo anno di vita, oltre che ad altre centinaia di variabili che descrivono ogni pulcino.

Il punto chiave dell'esperimento quindi è la manipolazione artificiale dei nidi, il tracciamento dei loro spostamenti e la randomizzazione delle rilevazioni, oltre che alla registrazione per ogni misurazione di tutte le variabili che descrivono l'uccello e l'ambiente in cui è nato, cresciuto e diventato adulto.

Domanda di ricerca e disegno sperimentale molto interessanti, ma che celano numerose problematiche a livello di analisi e modellazione statistica. Uno di questi come vedremo è la rilevazione multipla del colore per gli uccelli che sopravvivono, che spesso

avviene quindi anche in stagioni diverse, dove per rilevazione multipla si intende l'avvenuta misurazione casuale sullo stesso uccello in diverse fasi di vita. Queste rilevazioni sono state eseguite sugli stessi soggetti probabilmente per cercare di ricavare più informazioni possibili per altri scopi di ricerca, come quelli in Parker et al. (2011), ma che di fatto come vedremo non farebbero altro che creare distorsione nei risultati se inserite nel modello.

Un'altra problematicità che abbiamo affrontato e che ci ha costretto a limitare l'analisi solo ai pulcini nel primo anno di vita è stata la scarsità di informazioni sugli uccelli sopravvissuti, data l'elevatissima mortalità presente in questo esperimento. Come avremo modo di vedere dall'analisi dei valori mancanti infatti pochi sono stati i monitoraggi fatti sugli uccelli sopravvissuti che erano stati catturati e rilevati anche nel corso del primo anno di vita.

1.2 Dataset

In questo lavoro, si è analizzato un dataset concesso dal professor Simon Griffith (*Macquarie University*), un ricercatore che attualmente lavora nel campo della Biologia a Sidney e che da anni possiede e studia questi dati su cui si basa il lavoro di Parker et al. (2011).

Il dataset, costruito in diversi anni a partire da un esperimento scientifico messo in atto nella foresta Londinese *Wytham Wood*, un bosco di 380 ettari nell'Oxfordshire (UK), nel periodo che intercorre dal 2002 al 2004 (Parker et al., 2011), conteneva all'inizio della mia analisi 11430 rilevazioni del colore, ottenute tramite una spettroscopia in riflettanza (Parker et al., 2011), una tecnica complessa per analizzare l'intensità del colore attraverso uno spettrometro, ideata e presentata da Hadfield et al. (2006). Queste rilevazioni apparivano eseguite in modo non bilanciato d'estate e d'inverno, ossia prima e dopo la stagione riproduttiva, che per questi uccelli avviene in primavera e per ogni misurazione sono state registrate 162 variabili, tra cui le più importanti :

- le quattro variabili risposta : *headpc1*, *chestpc1*, *wingpc1* e *survived.to.breed.season*. Le prime tre riguardano la riflettanza del colore rispettivamente della testa, del petto e delle ali e sono rispettivamente le prime componenti principali di tre variabili che originariamente descrivevano in tre dimensioni il colore in base alla lunghezza d'onda considerata dallo spettrometro (Parker et al., 2011). L'ultima è la variabile dicotomica che indica se il pulcino è riuscito a sopravvivere o meno alla prima stagione riproduttiva.

- codici alfanumerici unici/identificativi di ogni pulcino, di ogni covata di origine e di allevamento.
- Caratteristiche fisiche dell'uccello nel momento della rilevazione, come:
 - sesso
 - peso
 - età
 - circonferenza del torace
- Caratteristiche dell'ambiente dove si trovava il nido in cui è nato il pulcino e del nido dov'è stato allevato, nel caso fosse avvenuta la manipolazione.
- Informazioni sulla dimensione del nido natale e di allevamento prima e dopo la manipolazione, se avvenuta.

Per motivi legati all'esperimento, che non prevedeva di rispondere alla stessa domanda di ricerca a cui ora stiamo cercando di rispondere, come detto nel paragrafo precedente sono state incluse anche tutte le misurazioni del colore effettuate sullo stesso uccello, rilevato quindi più volte nel corso del tempo. Il dataset quindi inizialmente conteneva molte informazioni ridondanti, valori mancanti e variabili non utili alla domanda di ricerca. Per esempio nelle rilevazioni invernali non sono state riportate tutte le informazioni dell'uccello che erano riportate nella rilevazione estiva precedente, ma solo alcune, creando così dei valori mancanti per costruzione. È stato pertanto necessario provvedere ad una forte pulizia e sistemazione dei dati, prima di procedere alla modellazione statistica, come spesso accade nel caso di dati sperimentali.

Capitolo 2

Processamento dei dati

2.1 Pre-processing e analisi valori mancanti

Sono state eliminate per prima cosa tutte le variabili non utili ai fini dell'analisi, ottenendo un dataset di dimensione 11430x77.

Successivamente, in accordo con il proprietario dei dati, si è deciso di escludere dall'analisi tutte le rilevazioni eseguite su pulcini a cui son state apportate manipolazioni diverse da quelle oggetto di studio, ossia lo spostamento nei primi due giorni di vita in un nido diverso da quello natale: più grande, più piccolo o di uguale dimensione.

Abbiamo perso in questo passaggio altre 3363 rilevazioni, il 29% del totale, arrivando ad averne 8067.

A questo punto il dataset presentava una distribuzione di valori mancanti non di certo incoraggiante, come si può notare dal grafico 2.1:

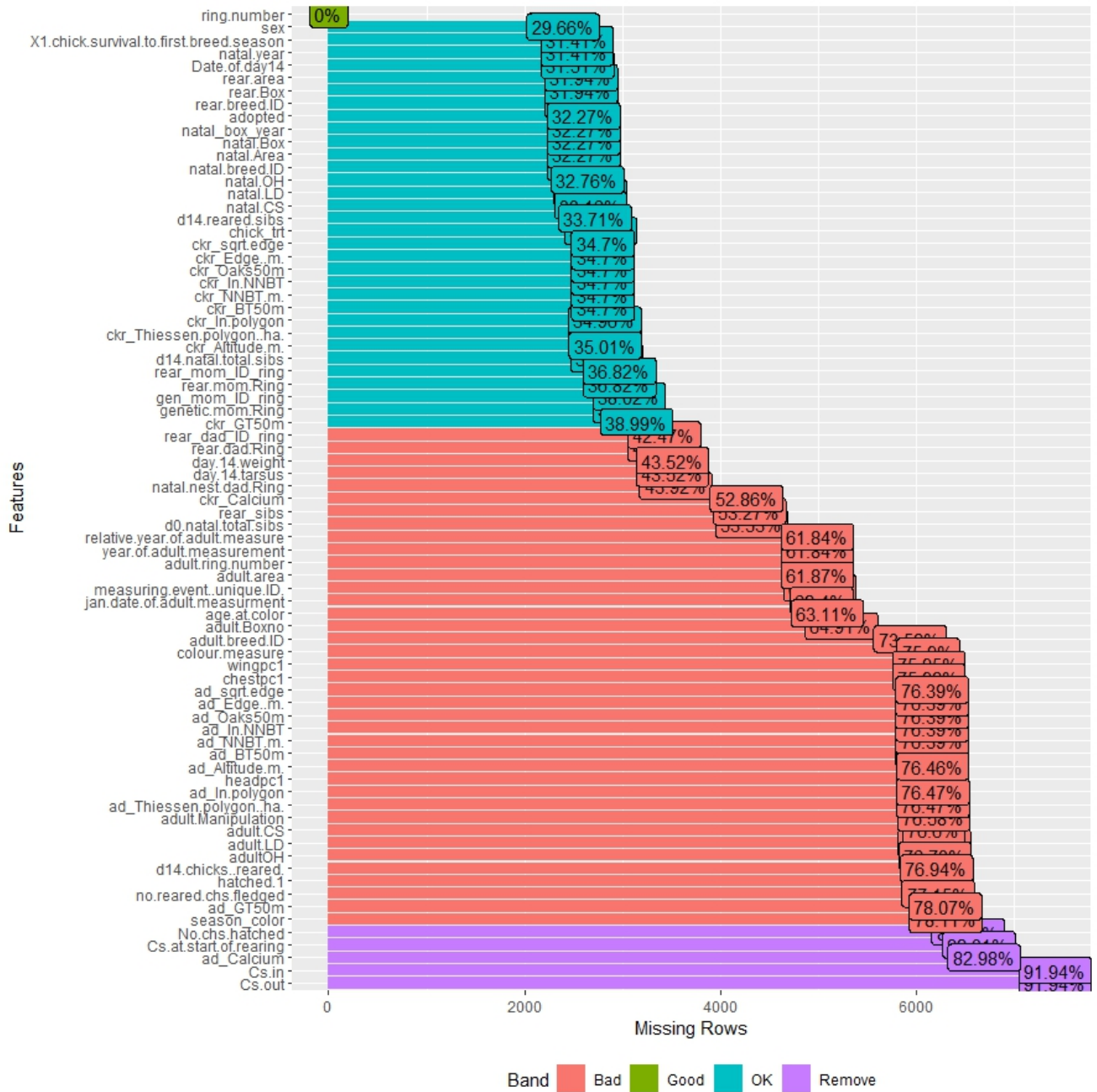


FIGURA 2.1: Distribuzione dei valori mancanti sulle variabili del dataset iniziale, dopo aver escluso le manipolazioni diverse da quelle in oggetto di studio.

2.1.1 Creazione di nuove variabili

A questo punto dopo aver pulito il dataset dalle informazioni non utili ai nostri scopi di ricerca son state create delle nuove variabili a partire da quelle esistenti oppure duplicate e modificate con ricodifiche o trasformazioni.

Allo scopo di misurare il grado di stress aggiunto o tolto dalla manipolazione e del suo effetto sulle variabili risposta si è scelto di creare una variabile chiamata *trt.comp* che misurasse quantitativamente la differenza della grandezza tra il nido dov'è cresciuto il pulcino e il nido dov'è nato, in tre modi differenti. Questa quantità vuole rappresentare la quantità di stress aggiuntiva subita dalla manipolazione dei nidi per ogni pulcino. Per i pulcini non manipolati quindi nati e cresciuti sullo stesso nido questa quantità sarà pari a zero, mentre per i pulcini spostati in nidi più grandi quindi con più fratelli rispetto al nido natale, questa quantità dovrà assumere valori positivi. Va quindi sottolineato che *trt.comp* non misura la competizione che deriva dal numero di fratelli con cui è stato allevato il pulcino, altrimenti un valore di *trt.comp* prossimo a zero indicherebbe che l'animale è stato allevato in assenza di fratelli. Essa misura invece il tasso di competizione tra fratelli aggiuntivo, derivato esclusivamente dalla manipolazione dei nidi. Valori positivi dovranno indicare uno stress maggiore, valori negativi una diminuzione di stress e valori prossimi a zero una condizione di stress simile o uguale a quella di partenza.

Un coefficiente significativamente positivo associato a questa nuova variabile in un modello per la riflettanza del colore significherebbe che aver stressato di più l'animale mettendolo in un nido con più fratelli rispetto all'originale ha portato il pulcino ad avere un ornamento migliore.

Si è scelto di costruire questa variabile attraverso il logaritmo naturale del rapporto tra la dimensione del nido di allevamento (in termini di fratelli vivi al secondo giorno dalla schiusa) e la dimensione del nido di origine (in termini di uova). Questa trasformazione di queste due importantissime quantità ha dato luogo alla variabile *trt.comp*, con valori che si distribuiscono tra -2 e 0.5 (figura: 2.2).

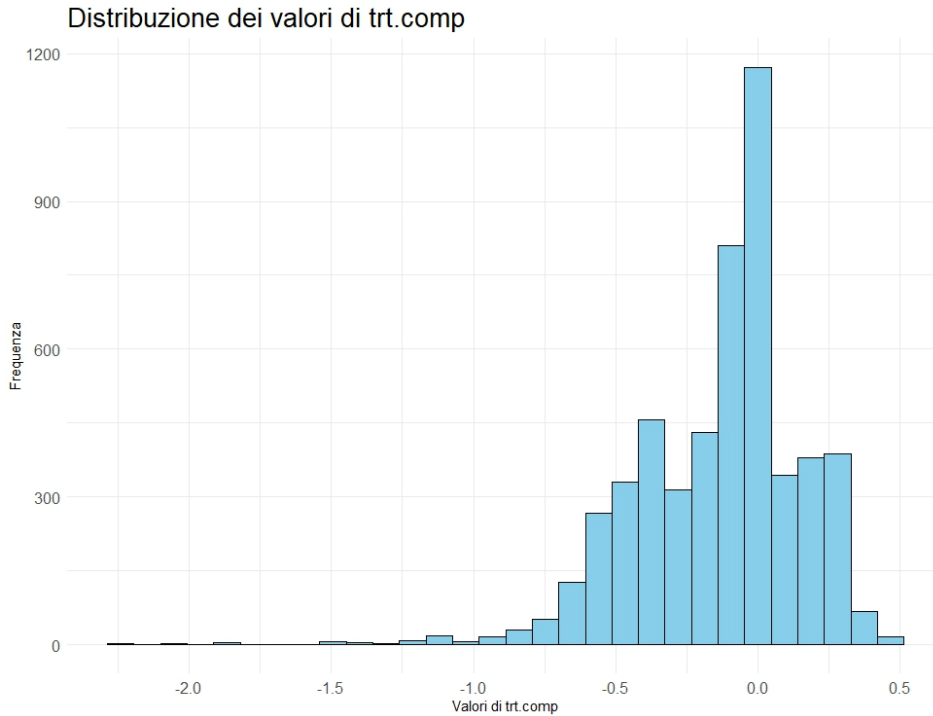
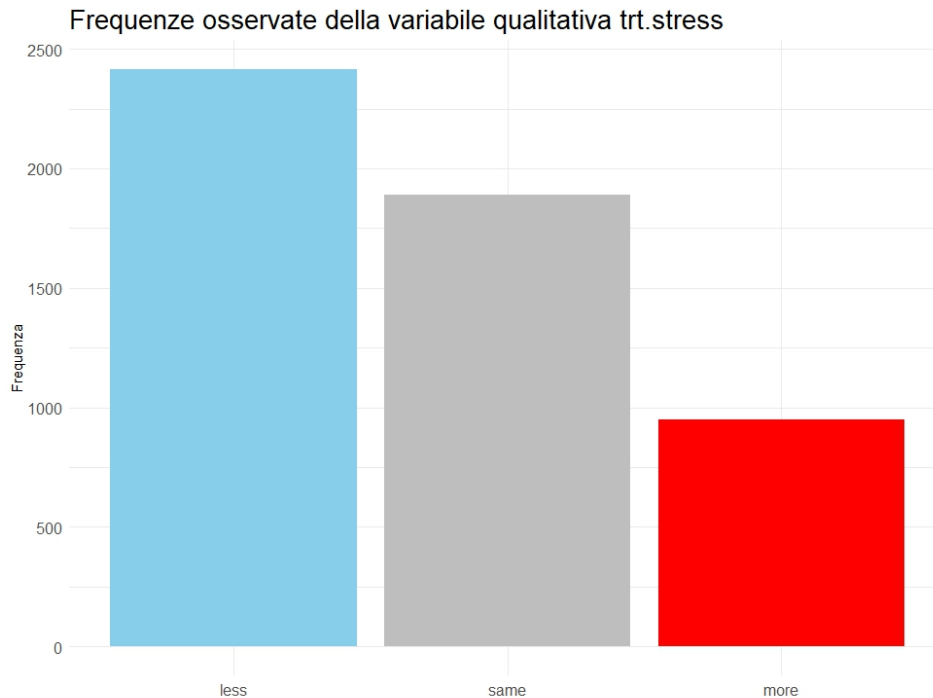


FIGURA 2.2: Distribuzione empirica dei valori di *trt.comp*.

Dei valori elevati positivi di questa variabile indicano una manipolazione che ha portato maggiore stress al pulcino, perché spostato in un nido più grande rispetto a quello natale, quindi con maggiore competizione tra fratelli. Valori bassi indicano una manipolazione che ha ridotto lo stress competitivo naturale tra fratelli, portando il pulcino a crescere in un nido con meno competizione interna.

Il secondo modo con cui abbiamo costruito questa variabile è stato trasformare *trt.comp* in variabile categoriale, suddividendola in tre intervalli di valori: $(-\text{Inf}, -0.1, 0.1, \text{Inf})$, ottenendo così una misura qualitativa chiamata *trt.stress* che misurasse in modo qualitativo il grado di stress aggiunto da parte della manipolazione artificiale dei nidi. In figura 2.3 è riportata la frequenza assoluta dei valori assunti da questa variabile. Come si può notare la maggior parte delle manipolazioni hanno portato i pulcini in un nido più piccolo rispetto a quello naturale.

FIGURA 2.3: Distribuzione empirica dei valori di *trt.stress*.

Sono state inoltre sistemate molte variabili esistenti, che erano codificate come caratteri oppure come interi anche se rappresentavano chiaramente dei fattori qualitativi, oppure che erano quantitative ma con pochi valori unici e quindi riportate a fattori.

Altre ancora invece erano già fattoriali ma per scopi di modellazione ne son state raggruppate le modalità in modo diverso e più efficiente. Una di queste, *ckr_BT50m* era una variabile che misurava il numero di altri nidi di cinciarelle nel raggio di 50m dal nido di allevamento, ossia il nido nella quale l'animale è stato cresciuto. Dal punto di vista Biologico questa variabile è molto importante perchè in quanto misura di densità di altri nidi nell'area, può essere vista sia come misura di fertilità e abbondanza di cibo del territorio sia come misura di competizione esterna, dove con esterna si intende tra i nidi e quindi tra famiglie diverse e non tra fratelli. Si è deciso di tenerla come quantitativa anche se conteneva solo 5 valori unici nel dataset: numeri interi da uno a cinque compreso. Ma per scopi di interpretazione e modellazione si è deciso di crearne altre due a partire da questa per averne a disposizione anche una versione fattoriale con livelli : "1-2", "3-4", "5" per una (*fert.area*) e "1", "2-3", "4-5" per l'altra (*fert.area2*).

È stata inoltre creata una versione categoriale della variabile quantitativa *d14.reared.sibs*, che indicava il numero di fratelli con cui è stato allevato il pulcino al quattordicesimo giorno. La nuova variabile fattoriale *fratelli* ora contiene 3 categorie: "0-5", "5-10 e "10+".

A tutte le variabili categoriali sono stati applicati i contrasti a media nulla, di modo che nel momento della valutazione della significatività dei coefficienti, in presenza di termini di interazione tra le variabili le significatività delle variabili quantitative non dipendessero dai livelli di riferimento delle qualitative.

Una volta escluse le osservazioni che non servivano ai nostri scopi e sistemato le variabili nel dataset, la processazione dei dati ha necessariamente preso due strade differenti a seconda dello scopo della modellazione: l'analisi dell'intensità del colore e lo studio della probabilità di sopravvivere alla prima stagione riproduttiva.

2.1.2 Scelta delle variabili che descrivono l'ambiente del nido

Nel nostro dataset sono presenti numerose ed interessanti variabili ambientali che descrivono il nido dove il singolo pulcino è stato allevato:

- **ckr_Thiessenn.polygon.ha.** Una misura della densità di altre specie di uccelli della famiglia *Paridae* (a cui appartengono le cinciarelle) nell'area della scatola nido di allevamento. Valori elevati indicano vicini più distanti rispetto a quelli con valori bassi.
- **ckr_ln.polygon** : Logaritmo naturale di `ckr_Thiessenn.polygon.ha.`
- **ckr_NNBT.m.** : La distanza lineare al nido di cinciarelle più vicino al nido di allevamento.
- **ckr_ln.NNBT** : Logaritmo naturale di `ckr_NNBT.m.`
- **ckr_sqrt.edge** : La radice quadrata della distanza dal nido di allevamento al limite del bosco più vicino.
- **ckr_Edge.m** : Distanza in metri dal nido di allevamento al limite del bosco più vicino.
- **ckr_Calcium** : Il livello di calcio nel terreno nell'area locale del nido di allevamento.
- **ckr_Altitude.m.** : Altitudine del nido di allevamento in metri.
- **ckr_Oaks50m** : Il numero di querce entro 50 metri dal nido di allevamento.
- **ckr_GT50m** : Il numero di nidi di cinciallegre (*Great-tits* entro 50 metri dal nido di allevamento).

- **ckr_BT50m** : Il numero di altre coppie di cinciarelle (*Blue-tits*) che nidificano entro 50 metri dal nido di allevamento.

Alcune di queste però sono semplicemente trasformazioni di altre, mentre altre ancora che descrivono misure diverse hanno però hanno un significato latente in campo biologico molto simile. Per riuscire a discriminare quali di queste variabili esprimessero un concetto biologico "unico" e quindi di conseguenza quali di queste potevano potenzialmente essere escluse dalla nostra modellazione, si son osservate le correlazioni empiriche:

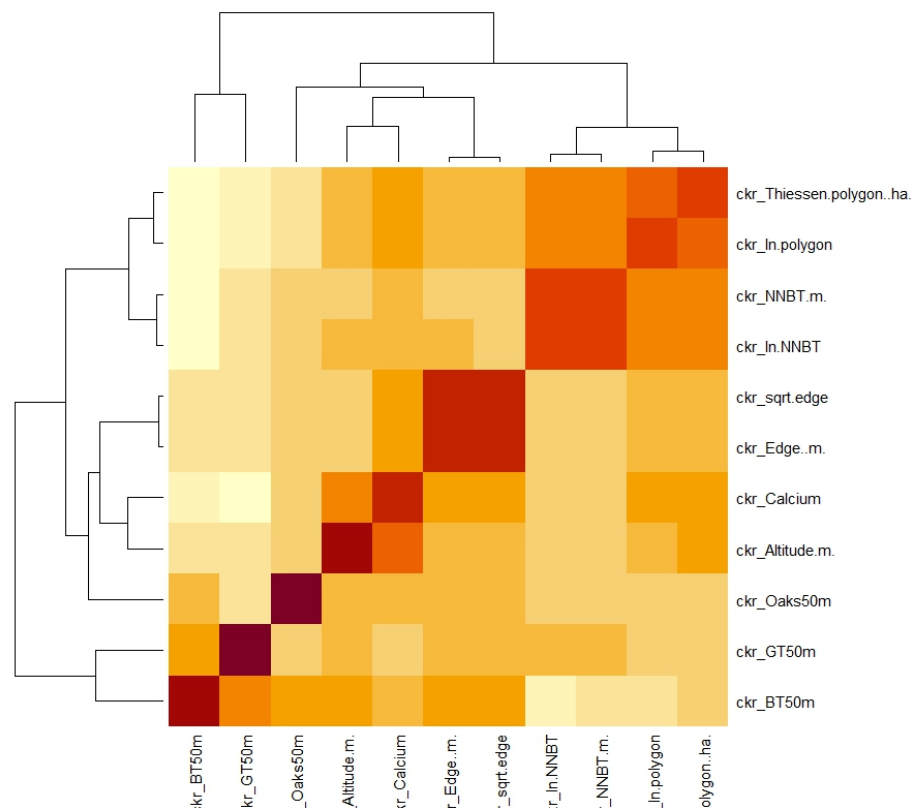


FIGURA 2.4: Correlazioni empiriche delle variabili ambientali.

Come si può notare dalla heatmap 2.4 delle correlazioni si possono in primo luogo definire tre macrogruppi di variabili poco correlate, che quindi potenzialmente possono spiegare tre concetti Biologici latenti diversi, di cui due di questi ulteriormente suddivisibili rispettivamente in tre e in due microgruppi:

- ckr_BT50m e ckr_GT50m
- ckr_Oaks50m, ckr_Altitude.m., ckr_Calcium, ckr_Edge.m, ckr_sqrt.edge.

A loro volta suddivise in tre gruppi:

- ckr_Oaks50m
- ckr_Altitude.m., ckr_Calcium
- ckr_Edge.m, ckr_sqrt.edge
- ckr_Thiessenn.polygon.ha., ckr_ln.polygon, ckr_NNBT.m., ckr_ln.NNBT.

A loro volta suddivise in due gruppi:

- ckr_Thiessenn.polygon.ha., ckr_ln.polygon
- ckr_NNBT.m., ckr_ln.NNBT

Una volta definiti questi tre macrogruppi, che ci indicano di utilizzare, tra tutte quelle considerate, come minimo tre variabili (appartenenti ognuna ad un macrogruppo diverso), si è deciso di tenere per tutti i modelli: **ckr_BT50m** e la sua corrispondente fattoriale **fert.area2**, **ckr_sqrt.edge** e **ckr_ln.polygon**, prendendo spunto anche dal lavoro fatto in Parker et al. (2011).

2.1.3 Standardizzazione delle variabili quantitative

Al fine di facilitare il processo di stima da parte dei modelli, soprattutto perchè lavoreremo con dei modelli gerarchici ad intercette casuali, si è scelto di standardizzare tutte le variabili quantitative presenti nel dataset, essendo queste su scale di misura talvolta molto differenti. È noto infatti che questo tipo di modello riesce a fornire delle stime più efficienti quando lavora con variabili sulla stessa scala di valori (Enders & Tofighi, 2007).

2.2 Creazione del dataset per lo studio della probabilità di sopravvivere

In questo caso specifico si è deciso di escludere a partire dal dataset iniziale di 8067 righe, ogni rilevazione multipla sullo stesso uccello, altrimenti avremo avuto numerose osservazioni ripetute dal punto di vista della variabile dicotomica che ci indica se l'uccello è sopravvissuto o meno. Questo perchè il colore talvolta è stato rilevato più di una volta per lo stesso soggetto e quindi includere tutte le osservazioni costituirebbe un problema dal momento che ogni singolo pulcino è sopravvissuto o meno alla prima stagione riproduttiva e questa informazione non cambia per tutte le sue rilevazioni successive, come anche tutte le informazioni su sesso, peso al quattordicesimo giorno, dimensione del nido natale ecc. In questo modo il modello binomiale lavorando sulle righe del

dataset non riuscirà ad accorgersi da solo che ci sono più righe che si riferiscono allo stesso animale, quindi prenderà ogni rilevazione come una unità statistica indipendente, creando distorsione nei risultati: per esempio se il pulcino *P804736* è sopravvissuto ed è stato rilevato 3 volte, il modello considererebbe 3 pulcini sopravvissuti con le stesse caratteristiche.

Il dataset ottenuto quindi prendendo la prima riga tra tutte quelle con lo stesso valore di *ring.number* (codice identificativo individuale) ora comprende 7031 osservazioni, il 60% del totale originale, ma il numero di osservazioni effettivamente utili saranno molte meno a causa di numerosi valori mancanti presenti in questo dataset. Solo per quanto riguarda la variabile risposta dicotomica che ci indica se il pulcino è sopravvissuto alla prima stagione riproduttiva abbiamo 1683 valori mancanti con 261 osservazioni di uccelli sopravvissuti e 5087 no. L'elevata mortalità presente in questo set di dati purtroppo costituirà un problema anche per quanto riguarda l'analisi del colore. Le osservazioni disponibili quindi escludendo gli NA sulla risposta sono in realtà 5348, senza contare quelle che perderemo perché contengono valori mancanti nelle variabili predittive. Il dataset infatti purtroppo conteneva inizialmente molti valori NA non facilmente imputabili e non solo nella variabile risposta ma anche in alcune delle variabili utili per la modellazione come quelle ambientali o che descrivono le caratteristiche dell'animale.

Questa risulta essere la distribuzione dei valori mancanti per questo dataset (figura: 2.5):

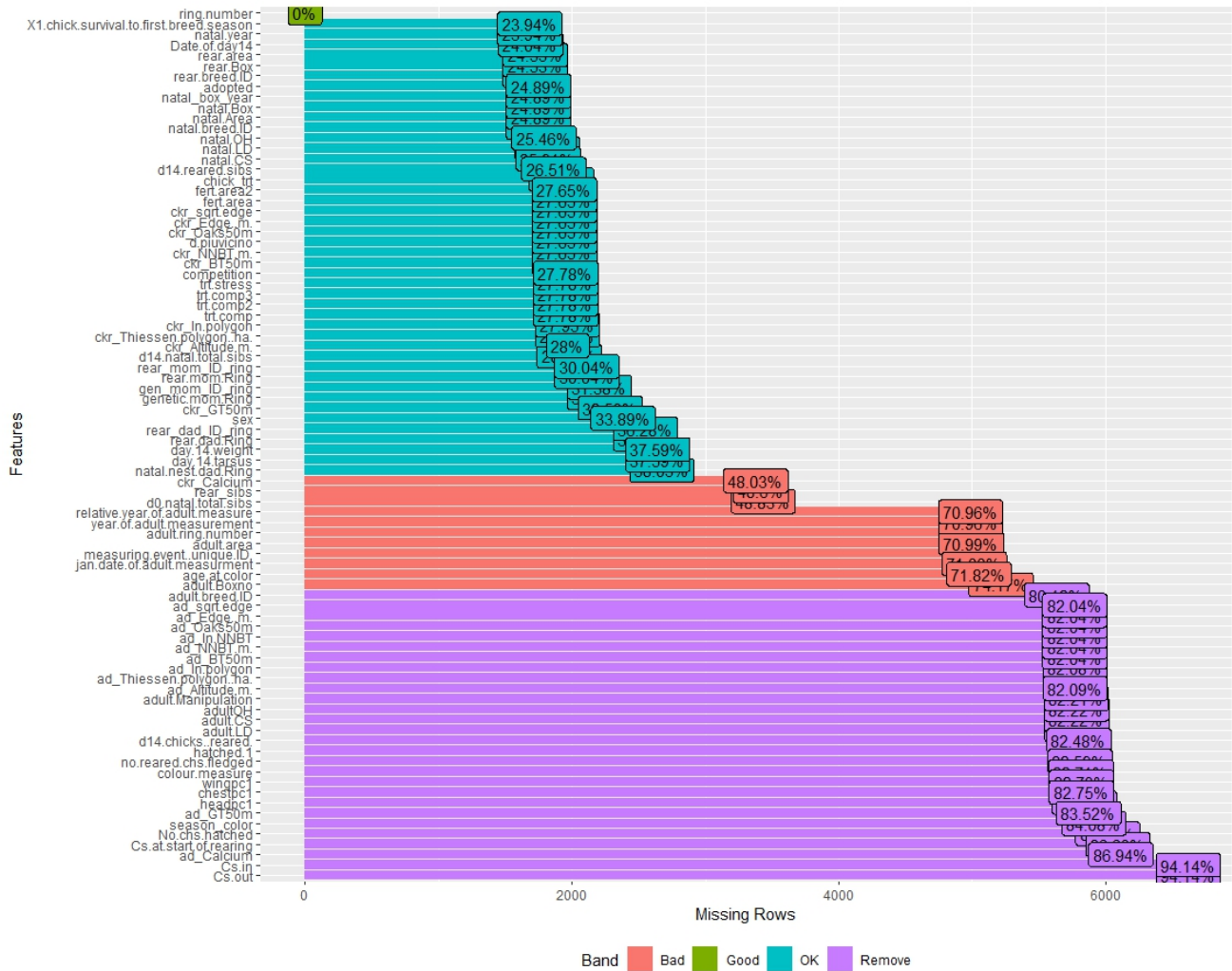


FIGURA 2.5: Distribuzione dei valori mancanti nel dataset finale per lo studio della probabilità π di sopravvivere alla prima stagione riproduttiva.

2.3 Creazione dataset per l'analisi del colore

In questo caso specifico, a partire dal dataset dove son state eliminate esclusivamente le osservazioni a cui son state applicate altre manipolazioni, con 8067 osservazioni, si è deciso di escludere per prima tutte le osservazioni con valori NA sulle tre variabili risposta: *headpc1*, *chestpc1* e *wingpc*. Con questo drop di osservazioni si son perse 6174 osservazioni, il 67% del totale originale. Nonostante questo passaggio non fosse strettamente necessario, dal momento che il modello eliminerà automaticamente tutte le osservazioni con valori mancanti sulla variabile dipendente, ci ha permesso di dare un'occhiata alla distribuzione dei valori mancanti una volta escluse queste osservazioni, per capire su quali rilevazioni verrà effettivamente stimato il modello.

La distribuzione dei dati mancanti infatti cambia ora radicalmente (figura: 2.6) a causa dell'elevata mortalità presente in questo dataset. Le variabili più vuote ora sembrano essere quelle relative ai pulcini nelle prime fasi di vita, mentre prima erano le variabili relative agli adulti. Questo strano fenomeno accade poichè una volta tolte le rilevazioni con valori mancanti sulle variabili risposta, restano praticamente quasi tutti pulcini che sono sopravvissuti alla prima stagione riproduttiva, quindi adulti che però spesso sono stati catturati solo una volta cresciuti e quindi non abbiamo a disposizione le informazioni sul loro primo anno di vita.

Problema che ci porterà purtroppo a stimare dei modelli che lavorano con poche centinaia di osservazioni, relative ad uccelli che hanno quindi sia le informazioni sul colore, sia le informazioni sul codice di covata natale e di allevamento. Infatti come si può vedere dalla figura 2.6, *natal.breed.ID* ora ha più dell'80% di valori mancanti.

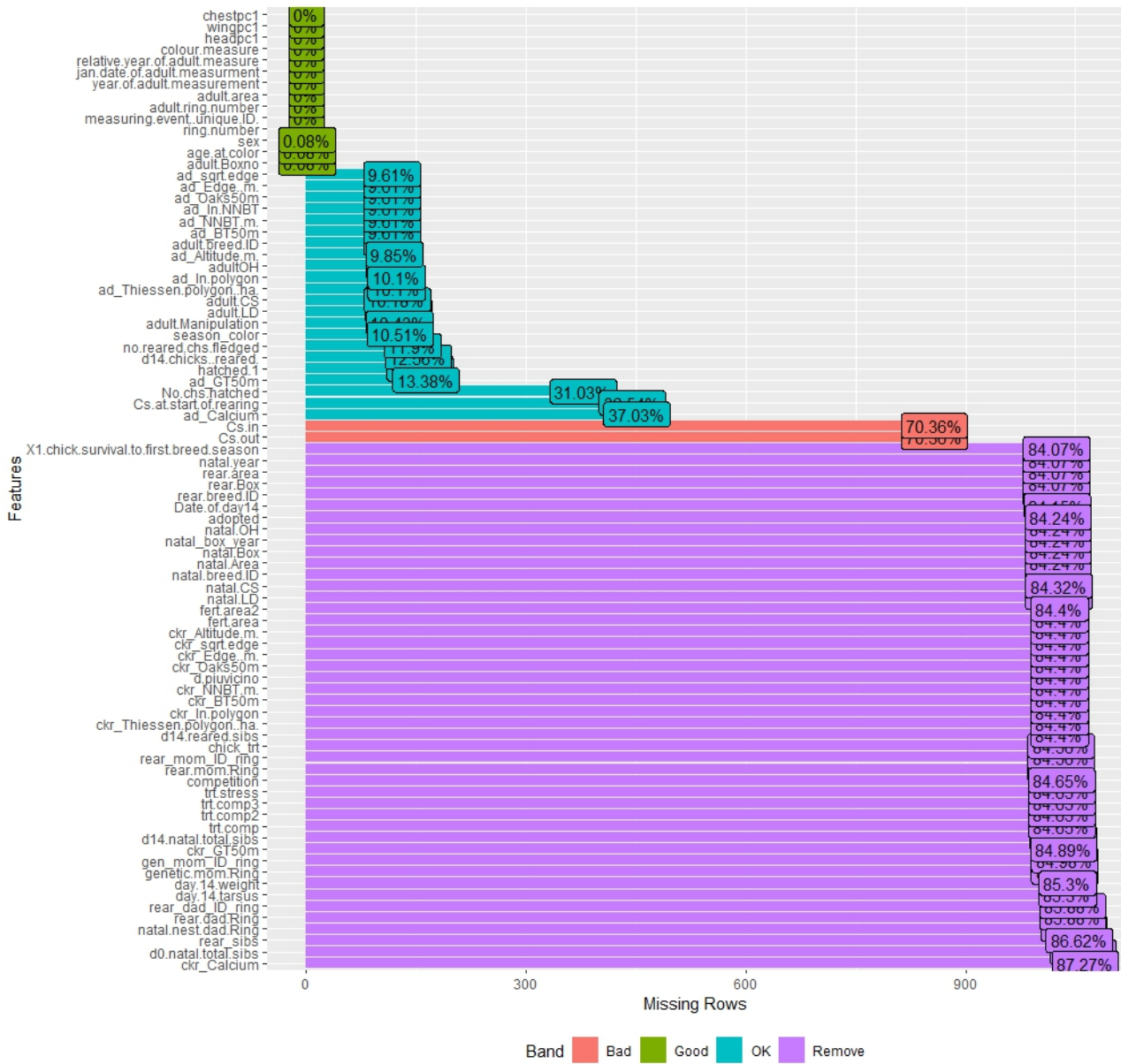


FIGURA 2.6: Distribuzione dei valori mancanti nel dataset finale per l'analisi del colore.

Sono state successivamente tolte inoltre tutte le rilevazioni invernali poiché queste contenevano moltissimi valori NA non imputabili e successivamente tutte le rilevazioni multiple sugli stessi soggetti, arrivando finalmente ad ottenere un dataset definitivo per questa analisi di dimensione 1218x83.

Capitolo 3

Modelli

3.1 Studio della probabilità di sopravvivere

Per la modellazione di π : probabilità di sopravvivere alla prima stagione riproduttiva, si è scelto di utilizzare un modello lineare gerarchico generalizzato ad effetti casuali per risposta binomiale (*GLMM*), attraverso la funzione *glmer* della libreria *lme4* su Rstudio.

Gli effetti casuali considerati saranno i codici identificativi delle covate dove è nato il pulcino e dov'è cresciuto, rispettivamente *natal.breed.ID* con 692 covate uniche in questo insieme di dati e *rear.breed.ID* con 647 covate uniche. Gli effetti delle covate ricordiamo che includono potenzialmente:

- Un effetto genetico
- un effetto dell'annata della schiusa
- un effetto della posizione precisa di quella schiusa nonché di tutte le sue possibili caratteristiche

Questo perchè l'effetto della "covata" è in una gerarchia di effetti e si trova a livello più basso.

Il modello lineare generalizzato ad intercette casuali mette in relazione la probabilità di sopravvivere, attraverso una sua trasformazione chiamata *link-function*, usata per riportare il codominio della parte sinistra dell'equazione del modello allo stesso codominio del predittore lineare, quindi \mathbb{R} , con una combinazione lineare di variabili predittive ed i loro rispettivi coefficienti fissi, che non cambiano quindi da gruppo a gruppo ma sono validi per l'intera popolazione. Ciò che varia da gruppo a gruppo, dove i gruppi son definiti dai valori unici dei codici delle covate *natal.breed.ID* e *rear.breed.ID*, sarà il valore dell'intercetta, composta da una parte fissa (livello base della popolazione) e da

una parte variabile o casuale composta dalla somma di α_i intercetta per il gruppo "i" di *natal.breed.ID* e α_j intercetta per il gruppo "j" di *rear.breed.ID*.

Le intercette vengono dette casuali perchè il loro valore stimato per il loro specifico gruppo può essere visto come una "stima" *Bayesiana*, dove a priori l'intercetta si distribuisce come una normale con media l'intercetta globale e varianza $\sigma_{\beta_0}^2$ che corrisponde alla varianza delle stime *no-pooling* ossia le stime ottenute come se *natal.breed.ID* e *rear.breed.ID* fossero entrate nel modello come effetti fissi. Il termine "stima" qui è usato in modo improprio, dato che non si può parlare di stimatore per una variabile casuale. In realtà si tratta di una media della distribuzione *a-posteriori*, che di fatto è solo un modo con cui si può vedere la classica stima frequentista dell'intercetta casuale.

Qui sta il compromesso del modello gerarchico tra le soluzioni *pooling*: dove il valore dell'intercetta casuale è uguale per tutti e quindi non ha variabilità e la soluzione *no-pooling*: dove nessun vincolo è imposto ai valori delle singole intercette che son stimate considerando esclusivamente le osservazioni interne al gruppo. Nell'approccio *multilevel* le intercette non sono tutte uguali ma appartengono alla medesima distribuzione e il valore di $\sigma_{\beta_0}^2$ determina quanto si è vicini alla soluzione *pooling* o *no-pooling*. Infatti se $\sigma_{\beta_0}^2 \rightarrow 0$ si impone che le intercette casuali siano tutte uguali, mentre se $\sigma_{\beta_0}^2 \rightarrow \infty$ di fatto ogni intercetta casuale può assumere ogni valore con uguale probabilità come nella soluzione *no-pooling*. Il valore stimato dal modello per $\sigma_{\beta_0}^2$ ci fornirà quindi una importante indicazione su quanto variano le nostre intercette casuali da gruppo a gruppo. Se questo valore fosse prossimo a zero significherebbe che non esiste una eterogeneità tra gruppi e avremo potuto considerare solo un'intercetta globale unica in stile *pooling*.

L'assenza di eterogeneità tra i gruppi di *natal.breed.ID* ha anche soprattutto un'importante interpretazione Biologica, ossia significherebbe un'assenza di un effetto genetico latente del nido natale che sottende le osservazioni. L'assenza di eterogeneità tra i gruppi di *rear.breed.ID* significherebbe invece l'assenza di un effetto latente del nido di allevamento, effetto quindi che si riferisce alla specifica covata dove il pulcino è stato allevato.

Per lo studio di questa probabilità si è scelto di inserire nel modello le seguenti variabili predittive, in accordo anche a quelle variabili selezionate da altri ricercatori in studi simili come Parker et al. (2011):

- **sex** sesso dell'animale
- **day.14.weight** ossia il peso al quattordicesimo giorno dalla nascita
- **trt.comp** ossia la variabile che abbiamo costruito che misura la quantità di stress indotta dalla manipolazione dei nidi

- **d14.reared.sibs** ossia il numero di fratelli con cui è stato allevato il pulcino
- **adopted** che ci indica se il pulcino ha subito la manipolazione e quindi è stato adottato
- **3 variabili ambientali**, una per ogni macrogruppo:
 - *ckr_BT50m*
 - *ckr_ln.polygon*
 - *ckr_sqrt.edge*
- **interazione tra day.14.weight e trt.stress** (versione fattoriale di trt.comp)
- **interazione tra fert.area** (versione fattoriale di *ckr_BT50m*) e **day.14.weight**
- **interazione tra ckr_ln.polygon e fratelli** (versione fattoriale di d14.reared.sibs)

Nel nostro caso la *link-function* sarà la funzione logistica:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \quad (3.1)$$

E il modello assumerà la forma di questa equazione:

$$\text{logit}(\pi_{kij}) = \beta_0 + \beta_{0i} + \beta_{0j} + \beta_1 X_{kij} + \dots + \beta_p X_{kij} \quad (3.2)$$

dove:

- $\text{logit}(\pi_{kij})$ è la variabile dipendente per l'osservazione k nel gruppo i, j ,
- X_{kij} è la variabile indipendente per l'osservazione k nel gruppo i, j ,
- $\beta_1 \dots \beta_p$ sono i coefficienti fissi associati alle variabili indipendenti,
- β_0 è l'intercetta fissa
- β_{0i} e β_{0j} sono gli effetti casuali delle intercette rispettivamente per il gruppo i e j , dove $\beta_{0i} = \frac{\sigma_{\beta_0}^2 * \bar{y}_{..i} + \sigma_{\epsilon}^2 * \bar{y}_{..}}{\sigma_{\beta_0}^2 + \frac{\sigma_{\epsilon}^2}{n_i}}$ e $\beta_{0j} = \frac{\sigma_{\beta_0}^2 * \bar{y}_{..j} + \sigma_{\epsilon}^2 * \bar{y}_{..}}{\sigma_{\beta_0}^2 + \frac{\sigma_{\epsilon}^2}{n_j}}$ con $\bar{y}_{..i}$ e $\bar{y}_{..j}$ media della risposta o stima *no-pooling* per il gruppo i e j , $\bar{y}_{..}$ media globale o stima *pooling*, n_j e n_i numerosità dei gruppi i e j e σ_{ϵ}^2 varianza del termine d'errore.

3.1.1 Analisi esplorativa

Riportiamo qui alcuni grafici che mostrano le relazioni osservate marginalmente tra la variabile risposta π e le variabili da inserire nel modello:

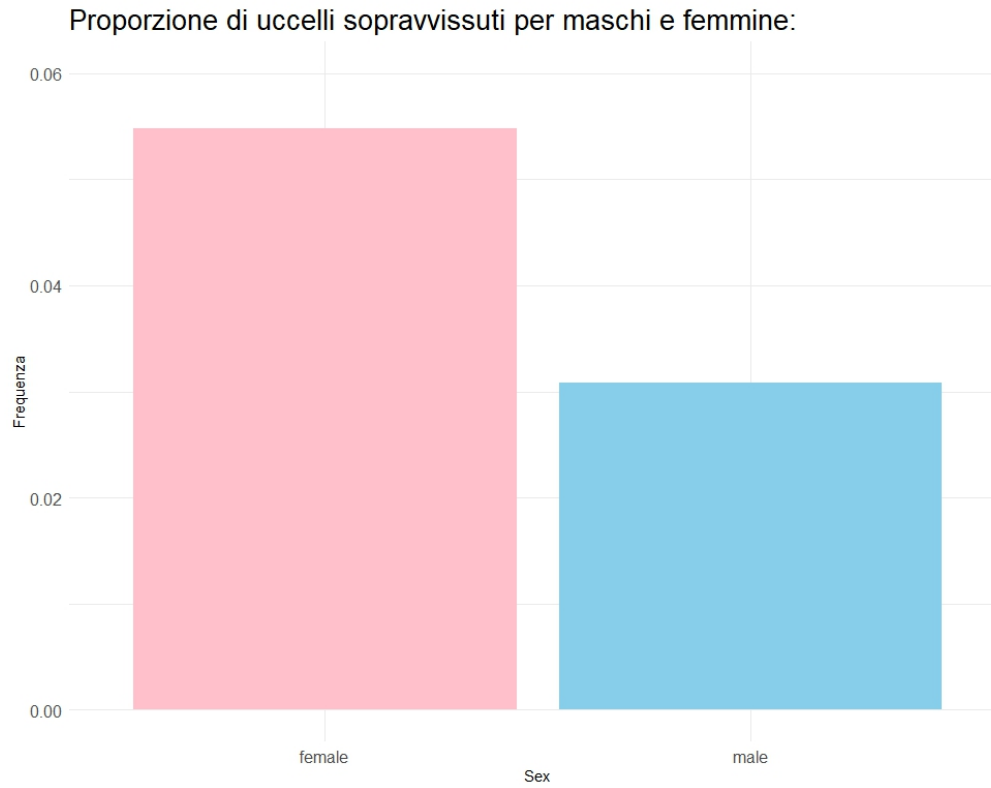


FIGURA 3.1: Proporzione di sopravvissuti per maschi e femmine.

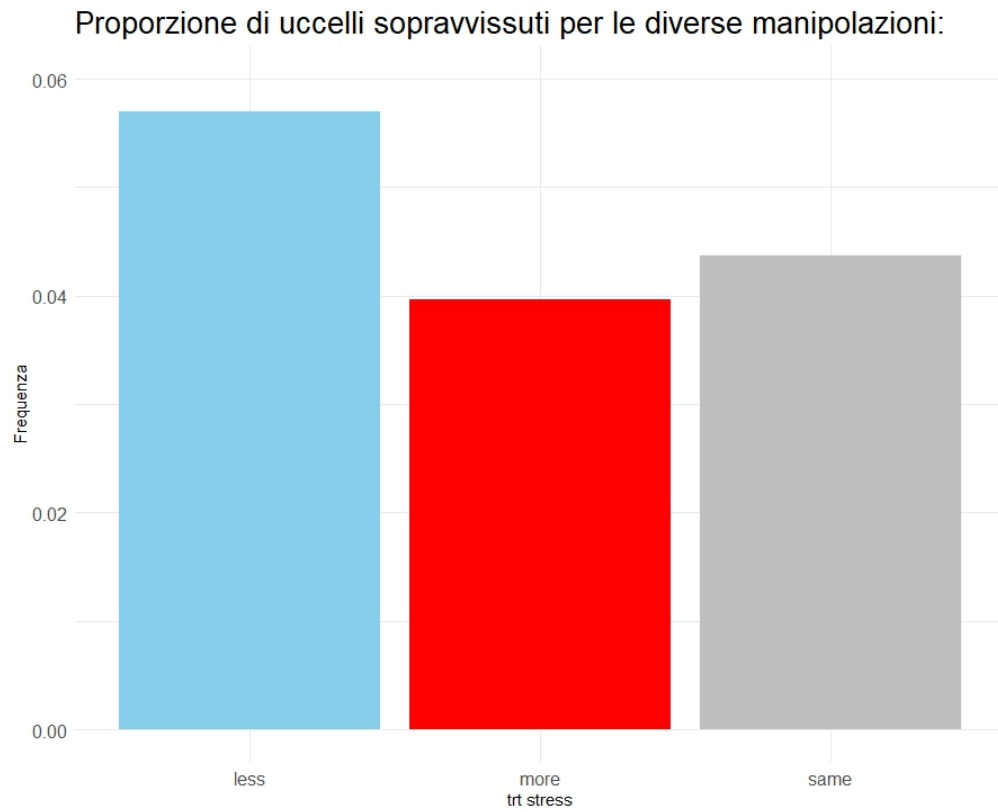


FIGURA 3.2: Proporzion e di sopravvissuti per i diversi tipi di manipolazione.

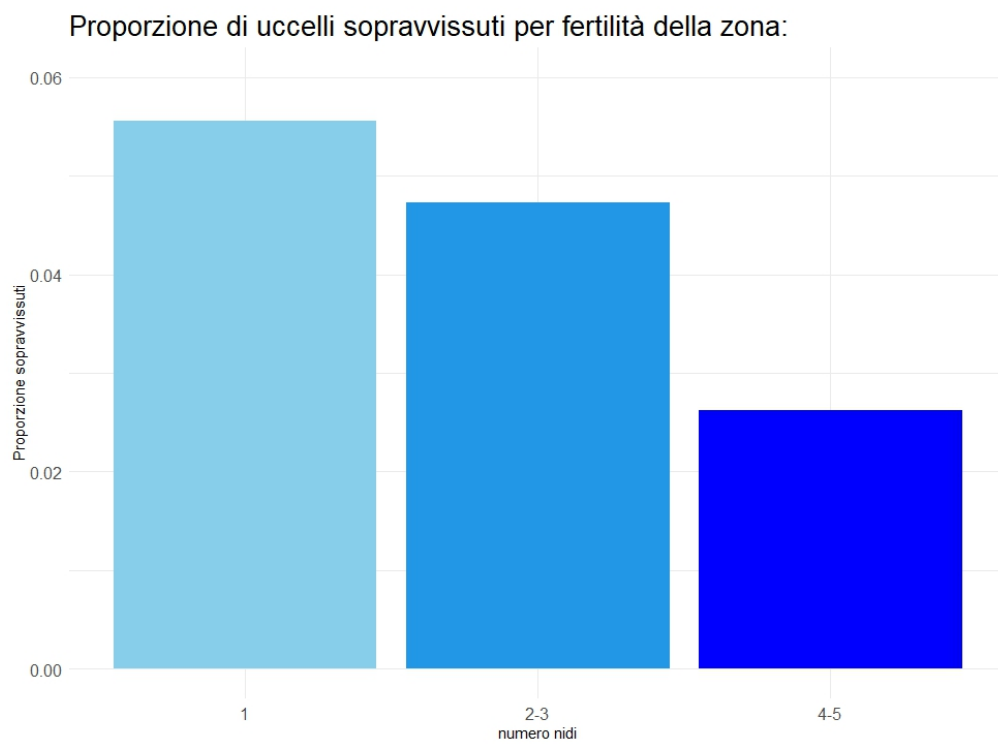


FIGURA 3.3: Proporzion e di sopravvissuti per tipo di fertilità dell'area data dal numero di altri nidi nel raggio di 50m dal nido di allevamento.



FIGURA 3.4: Distribuzione del peso per sopravvissuti e non.

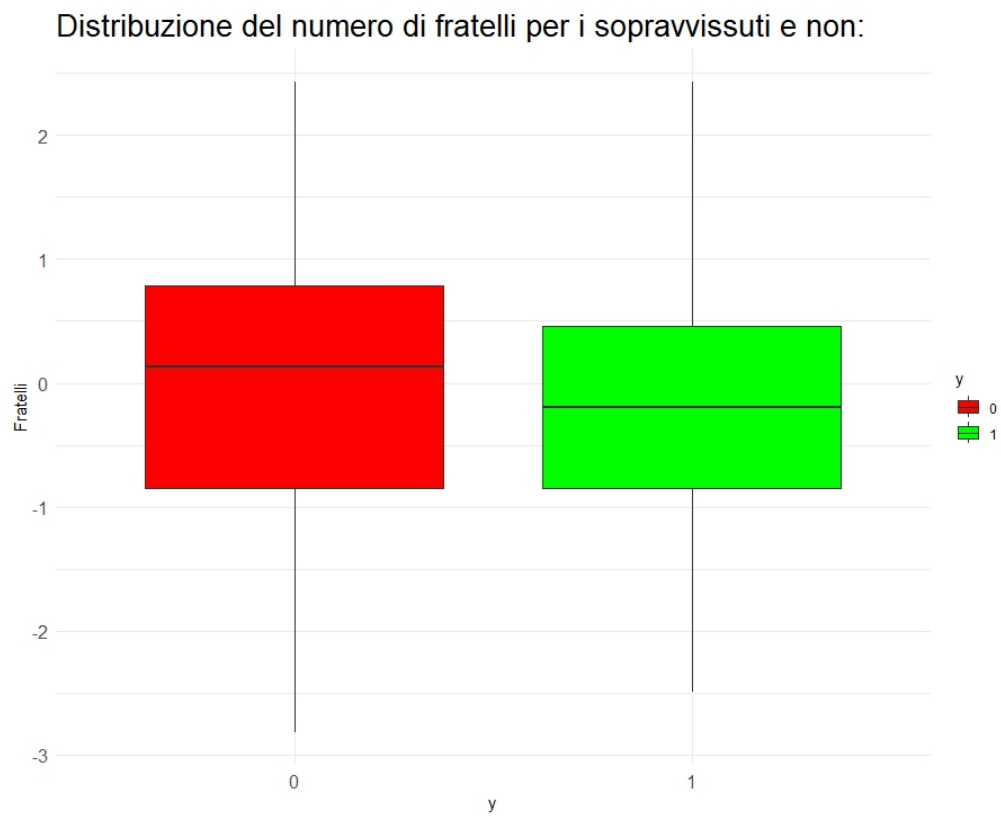


FIGURA 3.5: Distribuzione del numero di fratelli con cui il pulcino è stato allevato per sopravvissuti e non.

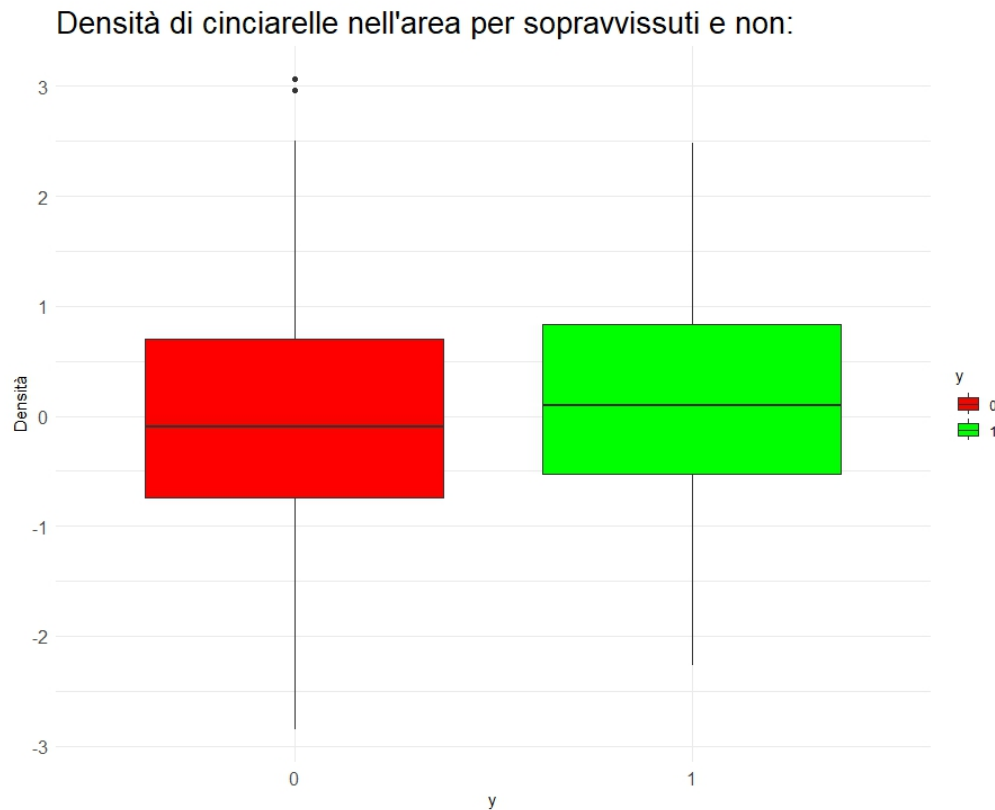


FIGURA 3.6: Distribuzione della densità di cinciarelle per sopravvissuti e non.
(Valori elevati di `ckr_ln.polygon` indicano nidi più distanti dal nido di allevamento.)

3.1.2 Modello glmm per la probabilità π di sopravvivere alla prima stagione riproduttiva

È stato stimato il modello lineare generalizzato gerarchico descritto nell'equazione 3.2 con le variabili indicate nella sezione precedente per la variabile risposta dicotomica che vale 1 se e solo se l'uccello a cui si riferisce è sopravvissuto alla prima stagione riproduttiva, che da questo momento chiameremo per semplicità y .

Qui di seguito è riportato l'output del modello:

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## y ~ sex + adopted + day.14.weight + trt.comp + ckr_BT50m + d14.reared.sibs +
## ckr_ln.polygon + ckr_sqrt.edge + day.14.weight:trt.stress +
## fert.area2:day.14.weight + ckr_ln.polygon:fratelli + (1 |
## natal.breed.ID) + (1 | rear.breed.ID)
```



```

##      Data: dati
##
##      AIC      BIC   logLik deviance df.resid
##    1294.7    1394.9   -630.4   1260.7     2660
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -0.6547 -0.2847 -0.2216 -0.1687  7.2028
##
## Random effects:
##  Groups          Name          Variance Std.Dev.
## natal.breed.ID (Intercept) 4.336e-05 0.006585
## rear.breed.ID  (Intercept) 3.112e-01 0.557810
## Number of obs: 2677, groups: natal.breed.ID, 385; rear.breed.ID, 359
##
## Fixed effects:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.8295646  0.1230448 -22.996 < 2e-16 ***
## sex1             0.1327479  0.0838015  1.584 0.113177
## adopted1        -0.0118151  0.0820354 -0.144 0.885482
## day.14.weight    0.5035374  0.1376165  3.659 0.000253 ***
## trt.comp         -0.0686058  0.1332982 -0.515 0.606777
## ckr_BT50m        -0.0241608  0.1260572 -0.192 0.848005
## d14.reared.sibs  0.0331498  0.1441593  0.230 0.818129
## ckr_ln.polygon   0.2595996  0.1457716  1.781 0.074934 .
## ckr_sqrt.edge    0.1772553  0.0991110  1.788 0.073703 .
## day.14.weight:trt.stress1 -0.0009813  0.1450061 -0.007 0.994601
## day.14.weight:trt.stress2  0.1587994  0.1672996  0.949 0.342523
## day.14.weight:fert.area21  0.3975368  0.1626011  2.445 0.014491 *
## day.14.weight:fert.area22  0.1535455  0.1565279  0.981 0.326619
## ckr_ln.polygon:fratelli1  0.2197000  0.2016464  1.090 0.275920
## ckr_ln.polygon:fratelli2 -0.1300841  0.1324940 -0.982 0.326193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Il modello è arrivato a convergenza sulla base di 2677 osservazioni ed ha prodotto numerosi risultati interessanti. Prima cosa non sembra ci sia un effetto genetico del

nido natale o comunque sia questo è veramente minimo, come si può notare dalla deviazione standard dell'intercetta casuale relativa a *natal.breed.ID*. Questo significa che i fattori che portano a sopravvivere alla prima stagione con maggiore probabilità non sono genetici ma sono prevalentemente ambientali e legati al nido di allevamento (la deviazione standard dell'intercetta casuale *rear.breed.ID* è pari a 0.55), quindi il nido di allevamento cattura una parte di variabilità non spiegata dagli effetti fissi inseriti e contribuisce quindi a spiegare la probabilità di sopravvivere.

Secondo, lo stress introdotto dalla manipolazione dei nidi e il fatto di essere adottati, quindi di crescere con dei genitori diversi rispetto a quelli "naturali", non ha un effetto significativo sulla probabilità di sopravvivere e questo ci porta a pensare che questa specie abbia una buona capacità di adattamento a diverse condizioni soprattutto nelle prime fasi di vita.

Sono significative invece a livello di significatività del 5% le variabili: **day.14.weight** con coefficiente positivo come ci aspettavamo, che ci indica che all'aumentare del peso, a parità di tutte le altre variabili, le cinciarelle hanno maggiore probabilità di sopravvivere alla prima stagione riproduttiva e il suo termine di **interazione con fert.area21**.

Ricordando che abbiamo applicato i contrasti a media nulla per tutte le variabili qualitative e che la variabile *fert.area2* presenta i seguenti contrasti per i suoi tre livelli:

```
contrasts(dati$fert.area2)
```

```
##      [,1] [,2]
## 1      1   0
## 2-3    0   1
## 4-5   -1  -1
```

Il termine di interazione *day.14.weight:fert.area21* ci indica che l'effetto del peso sulla probabilità di sopravvivere ha un effetto positivo ancora più marcato in quelle zone in cui abbiamo un solo nido nel raggio di 50m dal nido di allevamento rispetto alle zone in cui ne abbiamo 4 o 5. Questo è un risultato molto interessante perchè ci dice che il peso è sì importante "da solo" indipendentemente dal numero di nidi attorno per sopravvivere con maggiore probabilità ma è ancora più importante quando nel raggio di 50m dal nido di allevamento ho un nido solo, rispetto a quando ne ho 4 o 5. Questo fatto ci dice che il peso al quattordicesimo giorno ha in media una certa importanza nel predire la probabilità di sopravvivenza, ma è più importante quando l'uccello ha poca competizione con gli altri nidi rispetto a quando ne ha molta.

Risultano inoltre significative al 10% le variabili **ckr_ln.polygon** e **ckr_sqrt.edge** entrambe con coefficiente positivo come ci aspettavamo dall'analisi esplorativa. **Ckr_ln.polygon**

misura la densità di altri nidi di uccelli della stessa specie nell'area in termini di distanza media dagli altri nidi, quindi più distanti si trovano tra loro i nidi maggiore è la probabilità di sopravvivere. Mentre il coefficiente positivo di **ckr_sqrt.edge** ci dice che più il nido è distante dal margine della foresta più c'è probabilità che il pulcino sopravviva.

È stato inoltre stimato successivamente lo stesso modello su una versione dello stesso dataset bilanciato, ossia che contenesse la stessa proporzione di uccelli morti e sopravvissuti, per confermare la validità delle conclusioni ottenute. I risultati ottenuti sono praticamente identici sia per quanto riguarda la significatività dei parametri sia per quanto riguarda i singoli valori dei coefficienti stimati.

3.2 Analisi del colore

Come discusso nel capitolo dove spiegavamo la costruzione del dataset per l'analisi del colore, in questo caso specifico purtroppo le osservazioni sfruttabili saranno davvero poche a causa dell'elevata mortalità presente in questo dataset. Si vuole stimare un modello a scopo interpretativo, per le tre variabili: *headpc1*, *chestpc1* e *wingpc1*, per cercare di rispondere alla domanda di ricerca. A questo scopo si è scelto un modello lineare gerarchico a due intercette casuali:

- **natal.breed.ID**
- **rear.breed.ID**

con queste variabili predittive sicuramente presenti per tutte e tre le parti del corpo:

- **sex** sesso dell'animale
- **age.at.color** variabile dicotomica che indica se al momento della rilevazione l'animale ha raggiunto o meno il primo anno di vita
- **day.14.weight** peso dell'uccello al quattordicesimo giorno dalla nascita
- **trt.comp** variabile che abbiamo costruito al fine di misurare lo stress indotto dalla manipolazione dei nidi
- **adopted** variabile dicotomica che ci indica se l'animale ha subito o meno il trattamento quindi se è stato adottato
- **d14.reared.sibs** numero di fratelli con cui è stato allevato il pulcino, misura di competizione tra fratelli di base aldilà della manipolazione

- **3 variabili ambientali**, una per ogni macrogruppo:
 - *ckr_BT50m*
 - *ckr_ln.polygon*
 - *ckr_sqrt.edge*
- **interazione tra day.14.weight e trt.stress** (versione fattoriale di trt.comp)
- **interazione tra fert.area e day.14.weight**
- **interazione tra day.14.weight e fratelli** (versione fattoriale di d14.reared.sibs)

Le considerazioni fatte sulla scelta del modello gerarchico come modello per rispondere alla nostra domanda di ricerca e delle variabili da inserire come intercette casuali valgono anche qui come per quanto riguarda l'analisi precedente. Il modello lavorerà quindi con delle intercette casuali che rappresenteranno il cosiddetto "effetto del nido", che sono nel nostro caso al livello più basso della gerarchia di tutte le possibili scelte degli effetti casuali da inserire e che includeranno nella loro specifica stima anche l'effetto dei gruppi a livello gerarchico più alto come l'area e l'annata dove è stato allevato il pulcino.

Quindi ricapitolando il modello diventa:

$$Y_{kij} = \beta_0 + \beta_{0i} + \beta_{0j} + \beta_1 X_{kij} + \dots + \beta_p X_{kij} + \epsilon_{kij} \quad (3.3)$$

dove:

- Y_{kij} è la variabile dipendente per l'osservazione k nel gruppo i, j ,
- X_{kij} è la variabile indipendente per l'osservazione k nel gruppo i, j ,
- $\beta_1 \dots \beta_p$ sono i coefficienti fissi associati alle variabili indipendenti,
- β_0 è l'intercetta fissa
- β_{0i} e β_{0j} sono gli effetti casuali delle intercette rispettivamente per il gruppo i e j , dove $\beta_{0i} = \frac{\sigma_{\beta_0}^2 \bar{y}_{..i} + \sigma_{\epsilon}^2 \bar{y}_{..}}{\sigma_{\beta_0}^2 + \frac{\sigma_{\epsilon}^2}{n_i}}$ e $\beta_{0j} = \frac{\sigma_{\beta_0}^2 \bar{y}_{..j} + \sigma_{\epsilon}^2 \bar{y}_{..}}{\sigma_{\beta_0}^2 + \frac{\sigma_{\epsilon}^2}{n_j}}$ con $\bar{y}_{..i}$ e $\bar{y}_{..j}$ media della risposta o stima *no-pooling* per il gruppo i e j , $\bar{y}_{..}$ media globale o stima *pooling*, n_j e n_i numerosità dei gruppi i e j e σ_{ϵ}^2 varianza del termine d'errore.
- ϵ_{kij} è il termine di errore associato all'osservazione k nel gruppo i, j .

3.2.1 Analisi esplorativa

3.2.1.1 Analisi esplorativa per *headpc1*

Riportiamo qui alcuni grafici che mostrano le relazioni osservate marginalmente tra la variabile risposta *headpc1* e le variabili da inserire nel modello:

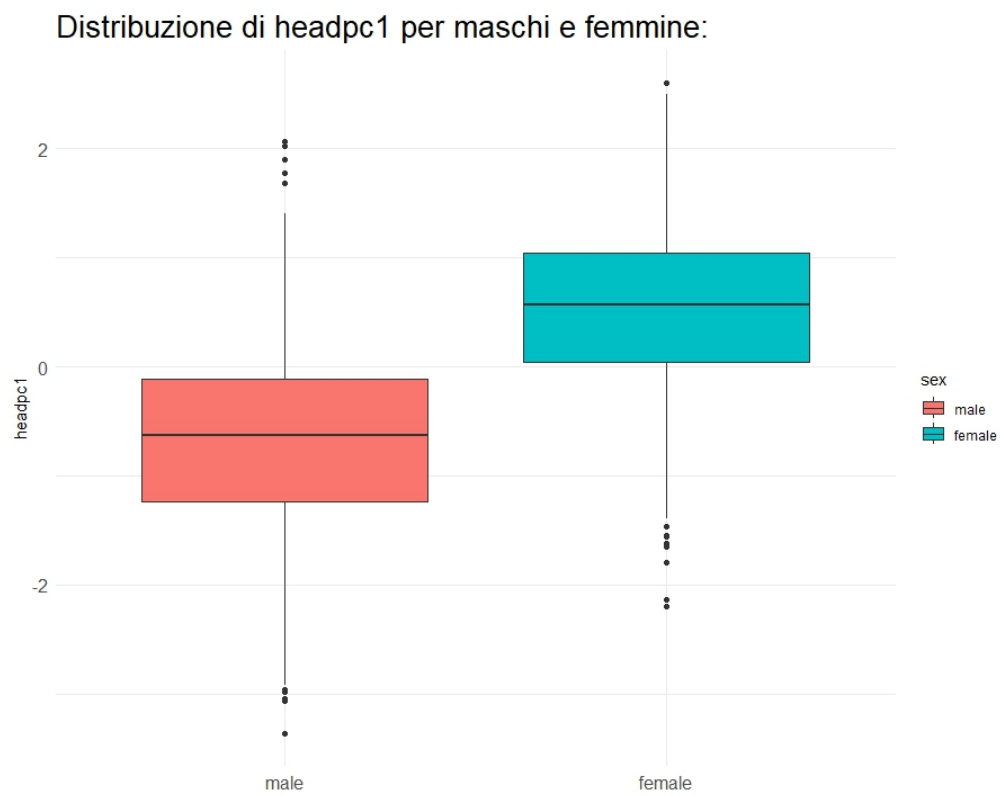


FIGURA 3.7: Relazione tra *headpc1* e sesso a livello marginale.

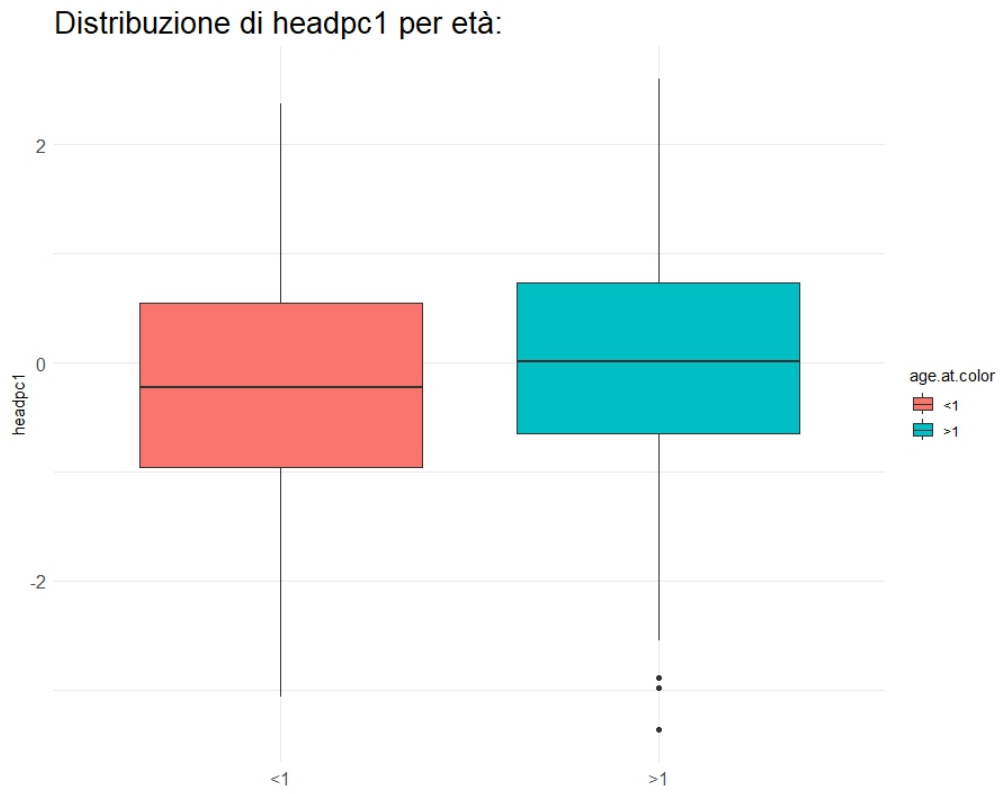


FIGURA 3.8: Relazione tra headpc1 e età a livello marginale.

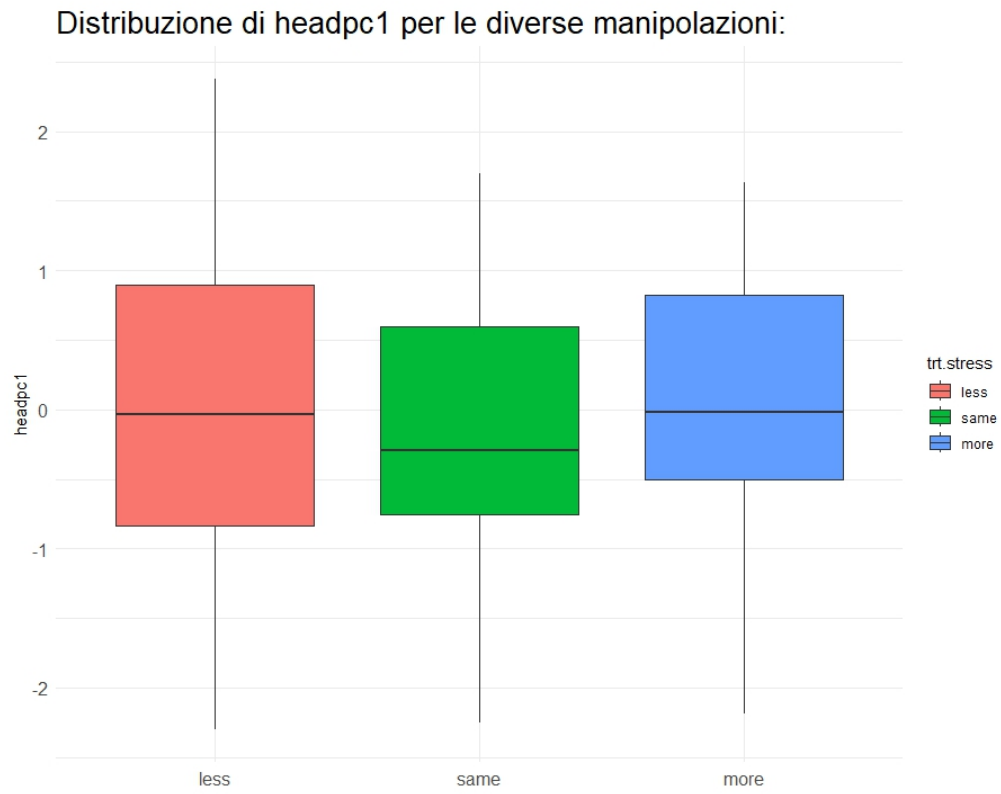


FIGURA 3.9: Relazione tra headpc1 e trt.stress a livello marginale.

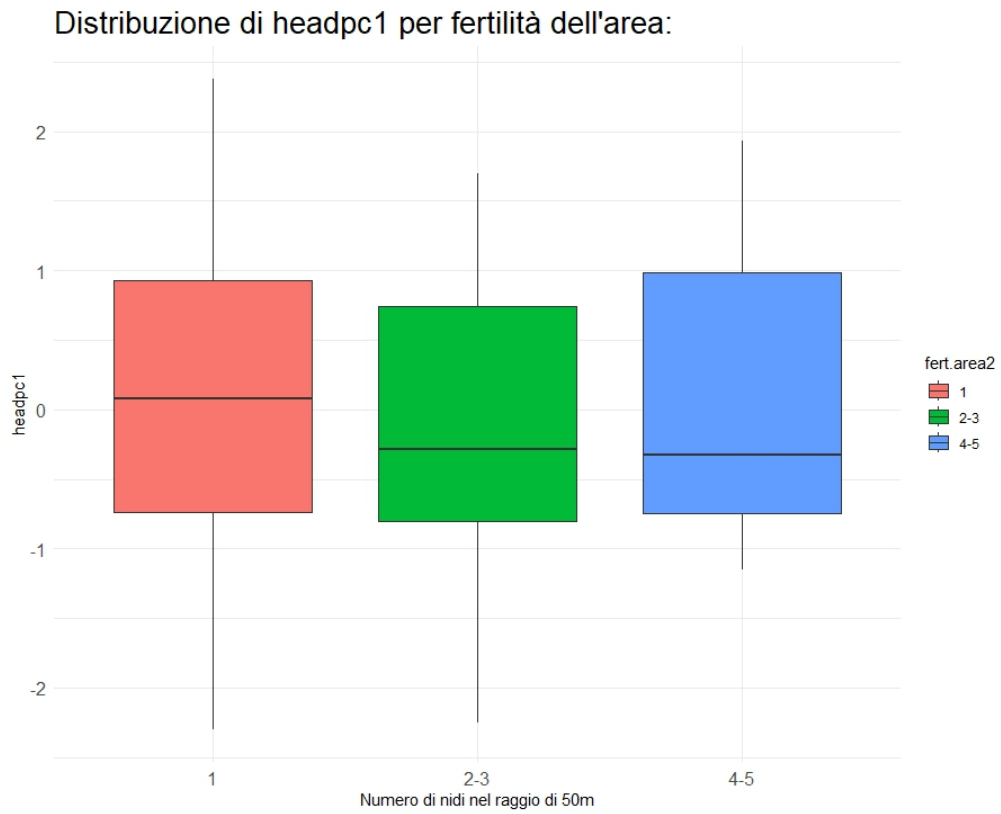


FIGURA 3.10: Relazione tra headpc1 e fert.area2 a livello marginale.

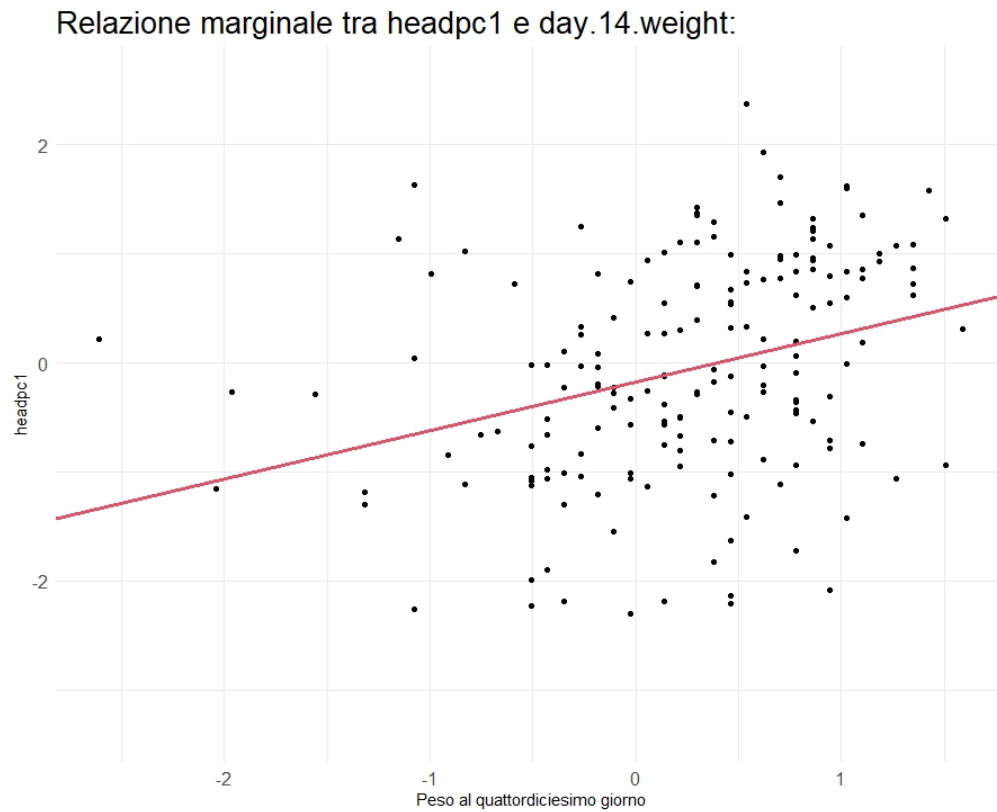


FIGURA 3.11: Relazione tra headpc1 e il peso del pulcino al quattordicesimo giorno dalla nascita a livello marginale.

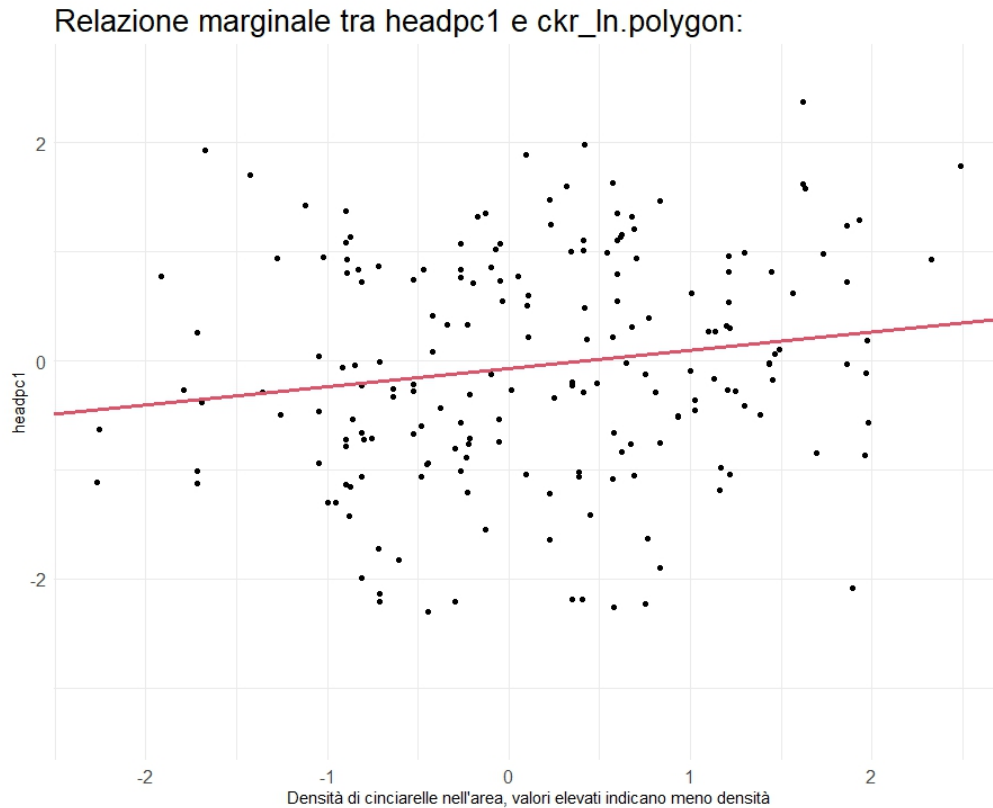


FIGURA 3.12: Relazione tra *headpc1* e la densità della popolazione di cinciarelle nell'area a livello marginale.

3.2.2 Modelli lmm per *headpc1*

Il modello descritto in 3.3 stimato per la variabile risposta *headpc1* con le variabili predittive descritte nel capitolo precedente: 3.2 ha prodotto i seguenti risultati:

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: headpc1 ~ sex + adopted + day.14.weight + trt.comp+
age.at.color +
##      ckr_BT50m + d14.reared.sibs + ckr_ln.polygon + ckr_sqrt.edge +
##      day.14.weight:trt.stress + fert.area2:day.14.weight
+day.14.weight:fratelli +
##      (1 | natal.breed.ID) + (1 | rear.breed.ID)
##      Data: dati
##
## REML criterion at convergence: 427.9
##
```

```

## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.60325 -0.62280  0.06819  0.54315  2.24313
##
## Random effects:
##   Groups          Name          Variance Std.Dev.
## natal.breed.ID (Intercept) 0.0000    0.0000
## rear.breed.ID  (Intercept) 0.1359    0.3687
## Residual                0.4677    0.6839
## Number of obs: 172, groups: natal.breed.ID, 132; rear.breed.ID, 124
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   -0.114825   0.085395 140.922533  -1.345    0.1809
## sex1          -0.599100   0.065074 152.347489  -9.206 2.48e-16 ***
## adopted1     -0.004129   0.061062 154.291243  -0.068    0.9462
## day.14.weight  0.485530   0.197780 144.242630   2.455    0.0153 *
## trt.comp      0.214080   0.117026 136.263590   1.829    0.0695 .
## age.at.color1 -0.097065   0.069465 153.145278  -1.397    0.1643
## ckr_BT50m     0.163070   0.103638 118.293493   1.573    0.1183
## d14.reared.sibs -0.101855   0.120722 126.590715  -0.844    0.4004
## ckr_ln.polygon  0.183564   0.090088 105.621388   2.038    0.0441 *
## ckr_sqrt.edge -0.034621   0.074412 113.948527  -0.465    0.6426
## day.14.weight:trt.stress1 -0.425598   0.303139 151.927540  -1.404    0.1624
## day.14.weight:trt.stress2  0.374954   0.183270 152.993291   2.046    0.0425 *
## day.14.weight:fert.area21  0.031272   0.137402 142.408813   0.228    0.8203
## day.14.weight:fert.area22 -0.042032   0.143386 139.755282  -0.293    0.7698
## day.14.weight:fratelli1    0.346993   0.235963 136.694529   1.471    0.1437
## day.14.weight:fratelli2    0.082295   0.177007 146.631097   0.465    0.6427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Il modello è stato stimato con 172 osservazioni ed è arrivato a convergenza. La deviazione standard dell'effetto casuale *natal.breed.ID* è stata stimata pari a zero mentre per l'intercetta casuale relativa a *rear.breed.ID* è stata stimata una deviazione standard di 0.3687. Questo risultato ci dice che per quanto riguarda il colore della testa sembra non esserci un "effetto genetico" derivato dal nido di nascita, sembra invece ci sia un

effetto ambientale catturato dall'intercetta casuale relativa al nido dov'è stato allevato. Sono inoltre risultate significative al 5% le seguenti variabili predittive:

- **sex**
- **day.14.weight**
- **ckr_ln.polygon**
- **day.14.weight:trt.stress2**

Per poter però interpretare i coefficienti relativi alle variabili fattoriali si deve considerare che son stati applicati i contrasti a media nulla:

```
contrasts(dati$sex)
```

```
      [,1]
male     1
female  -1
```

```
contrasts(dati$trt.stress)
```

```
      [,1] [,2]
less     1    0
same     0    1
more    -1   -1
```

```
contrasts(dati$fert.area2)
```

```
      [,1] [,2]
1         1    0
2-3      0    1
4-5     -1   -1
```

```
contrasts(dati$fratelli)
```

```
      [,1] [,2]
0-5     1    0
5-10    0    1
10+    -1   -1
```

```
contrasts(dati$age.at.color)
```

```

[,1]
<1   1
>1  -1

```

Il coefficiente relativo alla variabile *sex* è negativo, quindi come ci aspettavamo dall'analisi esplorativa i maschi tendono ad avere valori di *headpc1* più bassi a parità di tutte le altre variabili. La variabile che descrive il peso invece ha un coefficiente positivo, quindi anche in questo caso, come ci aspettavamo dall'analisi esplorativa gli uccelli che al quattordicesimo giorno dalla nascita sono più pesanti sono portati ad avere valori più alti di *headpc1* a parità degli altri fattori. Anche per la variabile *ckr_ln.polygon* come ci aspettavamo dall'analisi esplorativa, maggiore è la distanza dai nidi della stessa specie più vicini maggiore è la qualità del colore descritta da *headpc1*. Interessante infine è l'interazione significativa **day.14.weight:trt.stress2**, ossia l'effetto del peso per i non adottati (ossia i soggetti che hanno *trt.stress* pari a *same*) è ancora più forte e positivo sul colore della testa rispetto ai soggetti maggiormente stressati. Come per la probabilità di sopravvivere anche qui c'è una interazione tra il peso e la manipolazione, anche se prima era tra il peso e il numero di altri nidi attorno, sembra comunque anche qui che il peso conti meno quando c'è competizione tra i pulcini.

Interessante notare che *trt.comp* qui risulta significativo a livello 10% e con coefficiente positivo, quindi a parità delle altre variabili sembra che ci possa essere un effetto positivo sul colore della testa per gli uccelli che hanno subito una manipolazione che portava a maggiore stress, ma con un livello di significatività osservato pari a 0.0695 non è statisticamente affidabile fidarci di questo risultato.

3.2.3 Modelli lmm per *chestpc1*

Per quanto riguarda la variabile risposta *chestpc1* lo stesso modello descritto in 3.3 con le stesse variabili predittive ha prodotto i seguenti risultati:

```

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: chestpc1 ~ sex + adopted + day.14.weight + trt.comp + +age.at.color +
##      ckr_BT50m + d14.reared.sibs + ckr_ln.polygon + ckr_sqrt.edge +
##      day.14.weight:trt.stress + fert.area2:day.14.weight + day.14.weight:fratell
##      (1 | natal.breed.ID) + (1 | rear.breed.ID)
##      Data: dati
##
## REML criterion at convergence: 501.8

```

```

##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.20244 -0.51139  0.02601  0.53174  2.20007
##
## Random effects:
##   Groups          Name          Variance Std.Dev.
##  natal.breed.ID (Intercept) 0.17408  0.4172
##  rear.breed.ID  (Intercept) 0.09582  0.3096
##  Residual                0.69946  0.8363
## Number of obs: 172, groups:  natal.breed.ID, 132; rear.breed.ID, 124
##
## Fixed effects:
##
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    -0.097862   0.108066 137.629590  -0.906  0.36674
## sex1            0.015940   0.083127 149.976710   0.192  0.84819
## adopted1      -0.054755   0.078666 141.417828  -0.696  0.48755
## day.14.weight  0.630786   0.249065 142.785635   2.533  0.01240 *
## trt.comp       0.078676   0.148508 129.218707   0.530  0.59718
## age.at.color1  0.246481   0.087225 137.936164   2.826  0.00542 **
## ckr_BT50m     0.004874   0.129671 102.806942   0.038  0.97009
## d14.reared.sibs 0.083087   0.152646 119.954562   0.544  0.58724
## ckr_ln.polygon -0.093208   0.112710  91.953142  -0.827  0.41040
## ckr_sqrt.edge  0.068149   0.093542 106.703673   0.729  0.46788
## day.14.weight:trt.stress1 -0.161510  0.380998 147.062552  -0.424  0.67225
## day.14.weight:trt.stress2  0.154460  0.231445 151.046164   0.667  0.50555
## day.14.weight:fert.area21 -0.246038  0.172000 128.398470  -1.430  0.15502
## day.14.weight:fert.area22 -0.244966  0.179744 125.264974  -1.363  0.17537
## day.14.weight:fratelli1  0.406603  0.297796 134.816662   1.365  0.17441
## day.14.weight:fratelli2  0.057133  0.223579 146.530498   0.256  0.79867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Anche in questo caso il modello è stato stimato sulla base di 172 osservazioni ed è arrivato a convergenza. Le deviazioni standard delle due intercette casuali *natal.breed.ID* e *rear.breed.ID* sono state stimate rispettivamente pari a 0.4172 e 0.3096. In questo caso quindi per quanto riguarda il colore del petto sembra esserci un importante effetto

dovuto alla genetica parentale, anche più forte dell'effetto del nido in cui è stato allevato l'uccello. Le variabili risultate significative al livello di significatività 0.05 in questo caso sono:

- *day.14.weight* anche qui con coefficiente stimato positivo
- *age.at.color* con coefficiente stimato positivo, quindi uccelli rilevati entro il primo anno di vita sono generalmente più colorati nel petto rispetto a quelli rilevati successivamente, a parità di tutte le altre variabili.

3.2.4 Modelli lmm per *wingpc1*

Anche per quanto riguarda la variabile risposta *wingpc1*, ovvero l'intensità del colore delle ali abbiamo stimato lo stesso modello ottenendo i seguenti risultati:

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: wingpc1 ~ sex + adopted + day.14.weight + trt.comp + +age.at.color +
##   ckr_BT50m + d14.reared.sibs + ckr_ln.polygon + ckr_sqrt.edge +
##   day.14.weight:trt.stress + fert.area2:day.14.weight + day.14.weight:fratell
##   (1 | natal.breed.ID) + (1 | rear.breed.ID)
##   Data: dati
##
## REML criterion at convergence: 311.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.82652 -0.54280 -0.02148  0.60396  2.22529
##
## Random effects:
##   Groups          Name          Variance Std.Dev.
##  natal.breed.ID (Intercept) 0.0000    0.0000
##  rear.breed.ID  (Intercept) 0.0000    0.0000
##  Residual                0.2803    0.5294
## Number of obs: 172, groups:  natal.breed.ID, 132; rear.breed.ID, 124
##
## Fixed effects:
##                                Estimate Std. Error      df t value Pr(>|t|)
```

```

## (Intercept)          -0.084247    0.056776 156.000000   -1.484    0.13987
## sex1                 -0.810577    0.045271 156.000000  -17.905 < 2e-16 ***
## adopted1            -0.039096    0.042327 156.000000   -0.924    0.35708
## day.14.weight        0.028297    0.132793 156.000000    0.213    0.83153
## trt.comp             0.080602    0.077618 156.000000    1.038    0.30067
## age.at.color1       -0.003969    0.048288 156.000000   -0.082    0.93460
## ckr_BT50m           0.099703    0.067313 156.000000    1.481    0.14057
## d14.reared.sibs     -0.029823    0.079168 156.000000   -0.377    0.70690
## ckr_ln.polygon       0.166231    0.057579 156.000000    2.887    0.00444 **
## ckr_sqrt.edge       -0.100921    0.048190 156.000000   -2.094    0.03786 *
## day.14.weight:trt.stress1  0.171218    0.206107 156.000000    0.831    0.40740
## day.14.weight:trt.stress2  0.072528    0.124389 156.000000    0.583    0.56069
## day.14.weight:fert.area21 -0.132258    0.091641 156.000000   -1.443    0.15096
## day.14.weight:fert.area22 -0.178862    0.094938 156.000000   -1.884    0.06143 .
## day.14.weight:fratelli1  0.018925    0.156936 156.000000    0.121    0.90417
## day.14.weight:fratelli2  -0.103760    0.119304 156.000000   -0.870    0.38579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Anche in questo caso il modello è stato stimato su 172 osservazioni ed è arrivato a convergenza. Qui le deviazioni standard delle due intercette casuali *natal.breed.ID* e *rear.breed.ID* sono state stimate entrambe pari a zero. Per quanto riguarda il colore delle ali quindi, non sembra ci sia un effetto dovuto alla genetica e nemmeno un effetto dovuto al particolare nido di allevamento. Le uniche variabili che sembrano avere un effetto significativo al 5% sul colore delle ali sono il sesso (**sex**), anche qui con segno negativo (quindi le femmine sembrano essere più colorate nelle ali rispetto ai maschi a parità di tutto il resto), la misura di densità di altri nidi nell'area del nido di allevamento (**ckr_ln.polygon**) con segno positivo anche qui come per il colore della testa e la radice della distanza dal limite della foresta (**ckr_sqrt.edge**) qui però con segno negativo, quindi le ali risulteranno più ornate per gli uccelli allevati vicino ai margini.

3.2.5 Analisi dei residui

Per quanto riguarda la bontà di adattamento di questi tre modelli rispettivamente per il colore della testa, petto e ali si sono ottenuti degli ottimi risultati, come mostrano i grafici della distribuzione dei residui e i grafici *qq-plot* in figura 3.13,3.14,3.15, dove

vengono messi in relazione i quantili teorici di una distribuzione normale con i quantili empirici dei residui ottenuti.

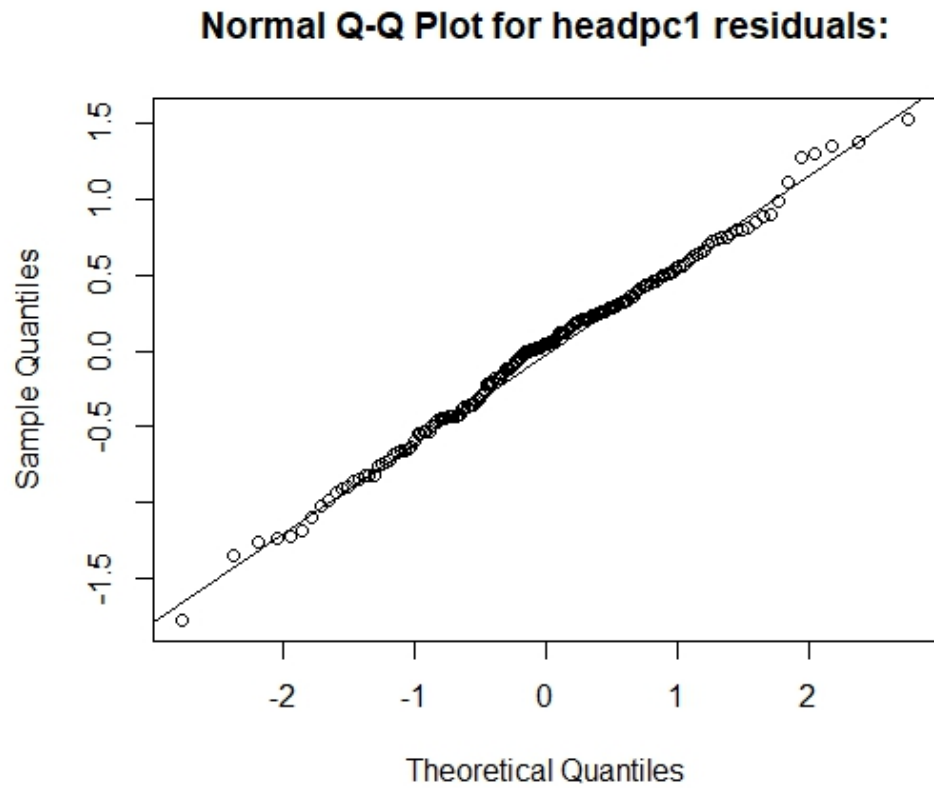


FIGURA 3.13: *qq-plot* per i residui del modello stimato per *headpc1*

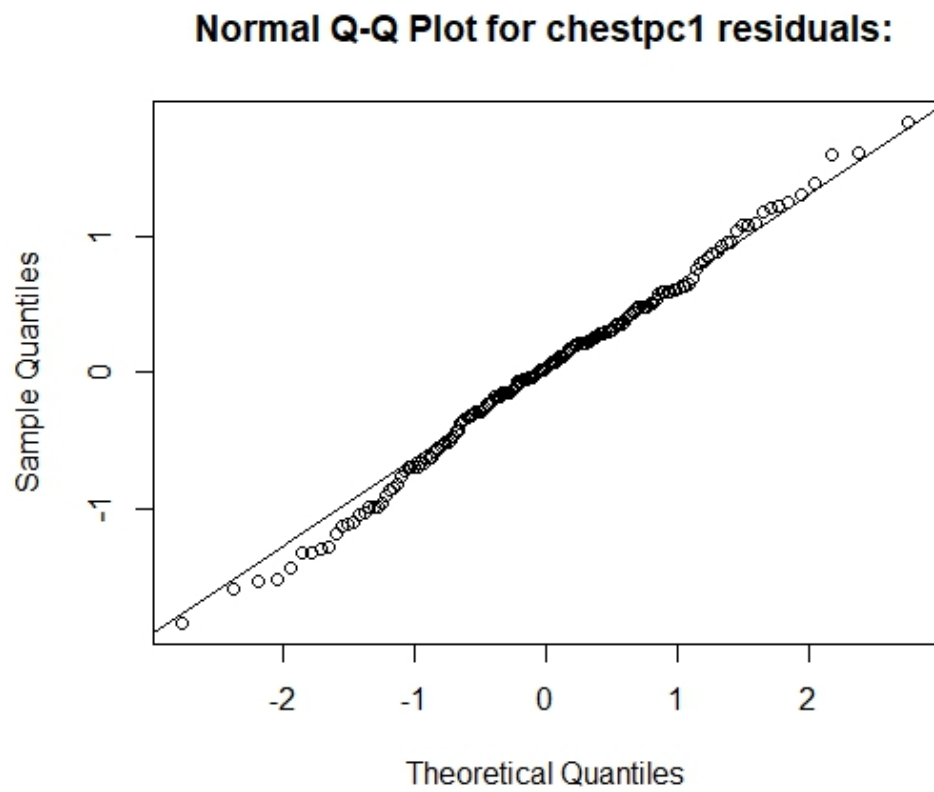
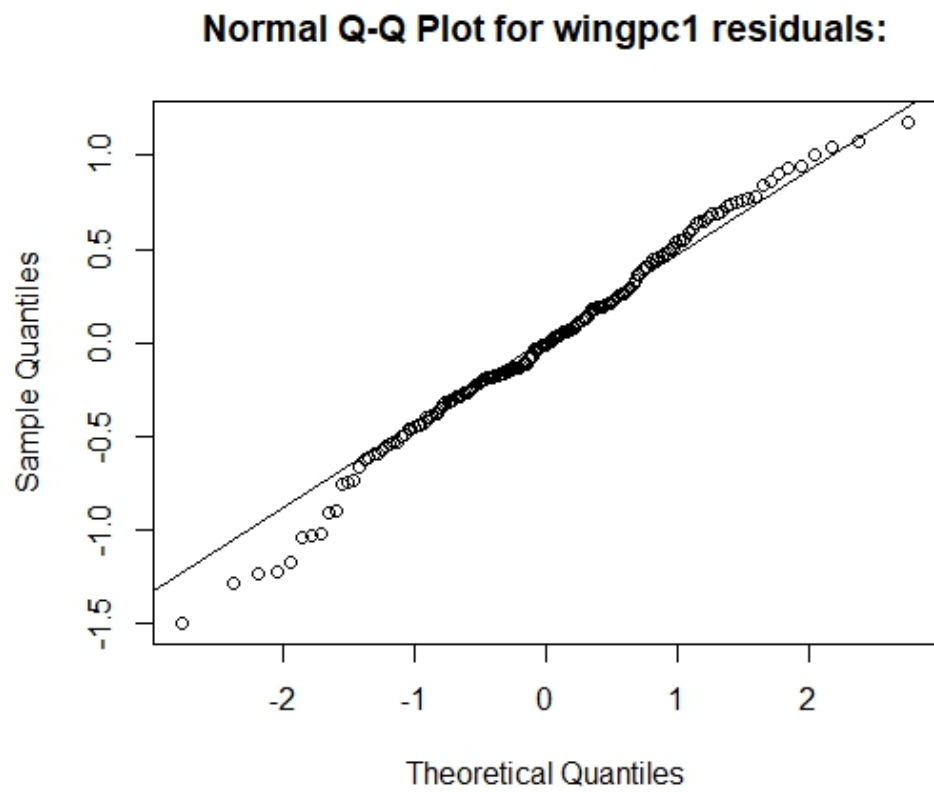


FIGURA 3.14: *qq-plot* per i residui del modello stimato per *chestpc1*

FIGURA 3.15: *qq-plot* per i residui del modello stimato per *wingpc1*

Capitolo 4

Modelli finali

Inizialmente avevamo pensato di stimare un modello gerarchico per riuscire ad interpretare e a tener conto di un eventuale effetto genetico del nido di nascita e un effetto ambientale del nido di allevamento, il problema è che per l'analisi del colore ci ritroviamo con 132 gruppi per l'intercetta casuale riferita a *natal.breed.ID*, 124 gruppi per l'intercetta casuale riferita a *rear.breed.ID* e sole 172 osservazioni totali. A causa dell'elevata mortalità presente in questo dataset, con la scelta di queste intercette casuali, le 172 osservazioni risultavano essere pericolosamente poche per ritenere i nostri risultati affidabili al 100%. Infatti avendo molti gruppi e poche osservazioni sorgeva il dubbio della precisione che potessero avere le stime degli effetti fissi, perchè per alcuni nidi potevamo trovarci ad avere una sola osservazione. Per concludere il nostro lavoro abbiamo voluto confermare la validità del nostro studio provando a considerare un solo individuo per ogni nido di nascita e per ogni nido di allevamento, in modo da far lavorare il modello per le tre variabili risposta: *headpc1*, *chestpc1* e *wingpc1* con tutte e sole osservazioni indipendenti e senza effetti casuali. In questo modo potevamo verificare se ottenevamo dei risultati coerenti con quelli ottenuti tramite i *GLMM* e quindi se potevamo convalidare le interpretazioni e conclusioni ottenute.

Questo risulta essere l'output del modello lineare per *headpc1* senza effetti casuali, stimato sul dataset di osservazioni indipendenti:

```
##  
## Call:  
## lm(formula = headpc1 ~ sex + adopted + day.14.weight + trt.comp +  
##     +age.at.color + ckr_BT50m + d14.reared.sibs + ckr_ln.polygon +  
##     ckr_sqrt.edge + day.14.weight:trt.stress + fert.area2:day.14.weight +  
##     day.14.weight:fratelli, data = datifinali)
```

```

##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.59716 -0.46124  0.03158  0.48927  1.94236
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.140592   0.103149  -1.363   0.1759
## sex1           -0.562617   0.081771  -6.880 5.28e-10 ***
## adopted1      -0.019260   0.072795  -0.265   0.7919
## day.14.weight  0.505382   0.205957   2.454   0.0159 *
## trt.comp       0.275164   0.123496   2.228   0.0281 *
## age.at.color1  0.020537   0.083103   0.247   0.8053
## ckr_BT50m     0.116658   0.113151   1.031   0.3050
## d14.reared.sibs -0.176987   0.126925  -1.394   0.1663
## ckr_ln.polygon  0.107429   0.101551   1.058   0.2927
## ckr_sqrt.edge  0.002165   0.084402   0.026   0.9796
## day.14.weight:trt.stress1 -0.413727   0.312242  -1.325   0.1882
## day.14.weight:trt.stress2  0.288246   0.209735   1.374   0.1724
## day.14.weight:fert.area21  0.013446   0.150006   0.090   0.9288
## day.14.weight:fert.area22 -0.032995   0.176282  -0.187   0.8519
## day.14.weight:fratelli1    0.307461   0.247983   1.240   0.2179
## day.14.weight:fratelli2    0.083068   0.186707   0.445   0.6573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7429 on 100 degrees of freedom
## (16 osservazioni eliminate a causa di valori mancanti)
## Multiple R-squared:  0.4817, Adjusted R-squared:  0.4039
## F-statistic: 6.195 on 15 and 100 DF,  p-value: 5.058e-09

```

I risultati di questo modello non si discostano molto dai risultati dello stesso modello con le due intercette casuali e questo ci conforta sull'affidabilità dei risultati. Interessante inoltre notare che oltre al sesso e il peso che risultavano significativi anche nel modello *glmm* ora risulta significativo anche l'effetto di *trt.comp*, il quale con coefficiente stimato positivo ci dice che maggiore è lo stress portato all'animale maggiore risulta essere la riflettanza del colore della testa a parità di tutte le altre variabili.

Questo invece è l'output del modello lineare per *chestpc1* senza intercette casuali:

```
##
## Call:
## lm(formula = chestpc1 ~ sex + adopted + day.14.weight + trt.comp +
##     +age.at.color + ckr_BT50m + d14.reared.sibs + ckr_ln.polygon +
##     ckr_sqrt.edge + day.14.weight:trt.stress + fert.area2:day.14.weight +
##     day.14.weight:fratelli, data = datifinali)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.42404 -0.56475  0.04313  0.62107  2.09030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.12279    0.13462  -0.912   0.3639
## sex1           0.01347    0.10672   0.126   0.8998
## adopted1     -0.01204    0.09500  -0.127   0.8994
## day.14.weight  0.54174    0.26879   2.015   0.0465 *
## trt.comp       0.08211    0.16117   0.509   0.6116
## age.at.color1  0.22406    0.10845   2.066   0.0414 *
## ckr_BT50m      0.04378    0.14767   0.296   0.7675
## d14.reared.sibs 0.10242    0.16565   0.618   0.5378
## ckr_ln.polygon -0.01300    0.13253  -0.098   0.9220
## ckr_sqrt.edge  0.04177    0.11015   0.379   0.7053
## day.14.weight:trt.stress1 -0.15539    0.40750  -0.381   0.7038
## day.14.weight:trt.stress2  0.15847    0.27372   0.579   0.5639
## day.14.weight:fert.area21 -0.36865    0.19577  -1.883   0.0626 .
## day.14.weight:fert.area22 -0.00936    0.23006  -0.041   0.9676
## day.14.weight:fratelli1  0.34708    0.32364   1.072   0.2861
## day.14.weight:fratelli2 -0.01840    0.24367  -0.075   0.9400
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9696 on 100 degrees of freedom
## (16 osservazioni eliminate a causa di valori mancanti)
## Multiple R-squared:  0.1134, Adjusted R-squared:  -0.01957
```

```
## F-statistic: 0.8528 on 15 and 100 DF, p-value: 0.6176
```

Anche questo modello ci rassicura sull'affidabilità dei risultati precedenti, infatti anche per il colore del petto abbiamo ottenuto gli stessi risultati: risulta ancora significativo l'effetto del peso al quattordicesimo giorno e l'età al momento della misurazione.

Infine la prova definitiva verrà dal modello per *wingpc1*, che aveva stimato pari a zero le deviazioni standard degli effetti casuali e ci preoccupava maggiormente. Questo è il risultato del modello senza effetti casuali stimato sul dataset di osservazioni indipendenti per il colore delle ali:

```
##
## Call:
## lm(formula = wingpc1 ~ sex + adopted + day.14.weight + trt.comp +
##     +age.at.color + ckr_BT50m + d14.reared.sibs + ckr_ln.polygon +
##     ckr_sqrt.edge + day.14.weight:trt.stress + fert.area2:day.14.weight +
##     day.14.weight:fratelli, data = datifinali)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30944 -0.34602  0.03021  0.32602  1.13465
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.199786   0.076324  -2.618  0.0102 *
## sex1          -0.780347   0.060505 -12.897 <2e-16 ***
## adopted1     -0.048653   0.053863  -0.903  0.3686
## day.14.weight  0.117437   0.152395   0.771  0.4428
## trt.comp      0.083751   0.091379   0.917  0.3616
## age.at.color1 0.058262   0.061491   0.947  0.3457
## ckr_BT50m    0.104298   0.083725   1.246  0.2158
## d14.reared.sibs 0.007257   0.093916   0.077  0.9386
## ckr_ln.polygon 0.163353   0.075141   2.174  0.0321 *
## ckr_sqrt.edge -0.130627   0.062452  -2.092  0.0390 *
## day.14.weight:trt.stress1 0.122596   0.231039   0.531  0.5969
## day.14.weight:trt.stress2 -0.030588   0.155190  -0.197  0.8441
## day.14.weight:fert.area21 -0.114217   0.110995  -1.029  0.3059
## day.14.weight:fert.area22 -0.034839   0.130437  -0.267  0.7899
```



```
## day.14.weight:fratelli1    0.115683    0.183491    0.630    0.5298
## day.14.weight:fratelli2   -0.063614    0.138151   -0.460    0.6462
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5497 on 100 degrees of freedom
## (16 osservazioni eliminate a causa di valori mancanti)
## Multiple R-squared:  0.7005, Adjusted R-squared:  0.6556
## F-statistic: 15.59 on 15 and 100 DF,  p-value: < 2.2e-16
```

Anche in questo caso le variabili significative sono le stesse che risultano significative nel *glmm* e anche i singoli valori dei coefficienti stimati non si discostano molto da quelli ottenuti da parte del modello gerarchico.

Capitolo 5

Conclusioni

5.1 Conclusioni per lo studio della probabilità di sopravvivere

Per quanto riguarda la nostra volontà di voler analizzare quali fattori contribuissero maggiormente alla morte prematura di questa specie animale, i nostri sforzi hanno portato a delle interessanti conclusioni e la nostra analisi si può dire abbia avuto successo per i suoi scopi. Si è scoperto che la genetica ha un ruolo praticamente irrilevante sulla probabilità di sopravvivere e tutta la variabilità dipende quindi dalle condizioni di vita che ha dovuto attraversare l'animale durante la sua crescita. Inoltre è emerso che il peso al quattordicesimo giorno dalla nascita sia il fattore più importante al di là di tutta la variabilità spiegata dall'effetto casuale del nido di allevamento, ma che questo abbia un'importanza minore quando c'è maggiore competizione tra i nidi. In conclusione quindi possiamo affermare che la probabilità di sopravvivere è influenzata soprattutto dalle condizioni ambientali del nido specifico in cui è stato allevato il pulcino oltre che dal peso dell'animale al quattordicesimo giorno. Infine, la manipolazione dei nidi sembra non aver avuto un effetto significativo sulla probabilità di sopravvivere.

5.2 Conclusioni per l'analisi del colore

Questa parte di analisi è stata purtroppo compromessa dal fatto che una volta escluse le osservazioni con valori mancanti sulle tre variabili risposta, ottenevamo un dataset con veramente poche osservazioni che avessero entrambe le informazioni sui codici delle covate di nascita e di allevamento. Nonostante questo, abbiamo deciso comunque di

scegliere il modello gerarchico ad effetti casuali per riuscire a rispondere alla nostra domanda di ricerca, che è comunque riuscito ad arrivare a convergenza e ha fornito dei risultati consistenti per quanto riguarda tutte e tre le parti del corpo, confortandoci sull'affidabilità dei risultati, confermata poi anche dai modelli finali stimati sulle osservazioni indipendenti. In conclusione possiamo quindi affermare che le tre parti del corpo della cinciarella rispondono in maniera differente ai fattori genetici ed ambientali. In particolare solo il colore del petto sembra derivare in buona parte da un fattore genetico, mentre per le ali ed il petto la genetica non sembra giocare un ruolo importante. Inoltre solo il colore della testa sembra essere influenzato positivamente dalla manipolazione dei nidi, mentre per il petto e le ali la manipolazione sperimentale non ha riscontrato effetti significativi. Infine abbiamo visto come per ogni parte del corpo contribuiscano variabili differenti e in misura diversa. A parità di tutte le altre variabili per il colore della testa sono significativi gli effetti del peso e del sesso, oltre che l'effetto della manipolazione, per il colore del petto sono significativi il peso e l'età e per il colore delle ali il sesso, la densità di altri nidi nell'area del nido di allevamento e la distanza dal limite della foresta.

Bibliografia

- ANDERSSON, S., ORNBORG, J. & ANDERSSON, M. (1998). Ultraviolet sexual dimorphism and assortative mating in blue tits. *Proceedings: Biological Sciences* **265**, 445–450.
- ENDERS, C. & TOFIGHI, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological methods* **12**, 121–38.
- GRIFFITH, S. (2000). A trade-off between reproduction and a condition-dependent sexually selected ornament in the house sparrow *passer domesticus*. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1115–1119.
- HADFIELD, J., BURGESS, M., LORD, A., PHILLIMORE, A., CLEGG, S. & OWENS, I. (2006). Direct versus indirect sexual selection: Genetic basis of colour, size and recruitment in a wild bird. *Proceedings. Biological sciences / The Royal Society* **273**, 1347–53.
- HUNT, S., BENNETT, A. T. D., CUTHILL, I. C. & GRIFFITHS, R. (1998). Blue tits are ultraviolet tits. *Proceedings: Biological Sciences* **265**, 451–455.
- PARKER, T., WILKIN, T., BARR, I., SHELDON, B., ROWE, L. & GRIFFITH, S. (2011). Fecundity selection on ornamental plumage colour differs between ages and sexes and varies over small spatial scales. *Journal of Evolutionary Biology* **24**, 1584–1597.
- SHELDON, B., ANDERSSON, S., GRIFFITH, S., [Ouml]—RNBORG, J. & SENDECKA, J. (1999). Ultraviolet colour variation influences blue tit sex ratios. *Nature* **402**, 874–877.

