

Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE

**IL DISTURBO OSSESSIVO COMPULSIVO NEI CANI
E GATTI: UNO STUDIO PILOTA**

Relatore: Prof.ssa Laura Ventura
Dipartimento di Scienze Statistiche

Correlatore: Dott.ssa Stefania Uccheddu
Clinica Veterinaria e Laboratorio San Marco

Laureanda: Ginevra Mingatti
Matricola: 1223920

Anno Accademico 2021/2022

INDICE

INTRODUZIONE.....	5
IL DISTURBO OSSESSIVO COMPULSIVO.....	7
1.1 Il dataset.....	9
1.2 Le variabili.....	10
ANALISI ESPLORATIVE	13
2.1 Analisi univariate.....	13
2.2 Analisi bivariate.....	21
MODELLI DI REGRESSIONE LOGISTICA.....	33
3.1 La regressione logistica.....	33
3.2 La correzione di Firth.....	34
3.3 Modello logistico con correzione di Firth.....	35
3.4 Modelli asintotici di ordine superiore.....	38
3.5 Modello logistico di ordine superiore.....	39
CONCLUSIONI.....	41
BIBLIOGRAFIA.....	45

INTRODUZIONE

Il seguente elaborato discute i risultati di analisi statistiche effettuate su un campione di dimensioni ridotte fornito da una Clinica Veterinaria di Padova. Questa analisi è stata svolta al fine di eseguire un primo studio del disturbo ossessivo compulsivo nei cani e i gatti.

L'elaborato è organizzato nel seguente modo: Il Capitolo 1 introduce lo studio illustrandone lo scopo e i dati, approfondendo in modo più dettagliato in cosa consiste il disturbo. Il Capitolo 2 tratta delle analisi preliminari delle variabili prese in considerazione, ovvero analisi univariate e bivariate. Il Capitolo 3 verte su modelli di regressione specifici dataset di piccole dimensioni, che hanno come scopo quello di stimare le relazioni tra le covariate e la presenza o meno del disturbo ossessivo compulsivo. Il Capitolo 4 raccoglie tutte le valutazioni e le conclusioni.

CAPITOLO 1

IL DISTURBO OSSESSIVO COMPULSIVO

Il disturbo ossessivo compulsivo (DOC) (in inglese *Obsessive-Compulsive Disorder OCD*) in psicopatologia viene descritto come una patologia caratterizzata da compulsioni ripetitive date da comportamenti o rituali ossessivi che vengono compiuti rigidamente (Tonna *et al.*, 2022).

Si possono definire ossessioni i pensieri intrusivi, immagini o impulsi che turbano la tranquillità personale dell'individuo e danno forma a dei sentimenti di ansia e inquietudine. Differentemente dalle ossessioni, le quali avvengono senza alcuna logica razionale, le preoccupazioni emergono solo in seguito a situazioni sgradevoli e/o negative di vita quotidiana.

Le compulsioni, invece, sono azioni o rituali messi in atto da chi soffre di DOC al fine di moderare, controllare o ridurre il malessere provocato dalle ossessioni. Questi comportamenti danno un sollievo momentaneo. Il soggetto, infatti, verrà in breve tempo pervaso ancora da una nuova ossessione che scaturirà a sua volta una compulsione, esso si ritroverà perciò intrappolato in una sorta di “loop”. Ciò che differenzia il comportamento ossessivo compulsivo da un'abitudine quotidiana è la paura dell'individuo di causare situazioni sgradevoli alla mancata azione svolta e quindi l'incapacità del soggetto di non attuare il gesto compulsivo. Alcuni studi infatti sostengono che questi rituali vengano attuati per dare al soggetto la sensazione di sicurezza e la parvenza di potersi adattare a ogni circostanza e a ogni tipo di cambiamento. In paragone ad altri disturbi, il DOC si presenta maggiormente in circostanze in cui è elevata la frequenza e dall'imprevedibilità dell'evento critico, perciò il soggetto sente la necessità di prevedere o prevenire l'evento attraverso le compulsioni. Per questo motivo generalmente i rituali mutano al variare dell'ambiente circostante, perché esso essendo imprevedibile e in costante cambiamento, genera nel soggetto il desiderio di sentirsi “al sicuro”. Comunemente l'ossessione, cioè l'evento critico, risulta improbabile; infatti, un individuo non soggetto da DOC riconosce di avere una preoccupazione poco sensata e tende a ignorarla; al contrario, un individuo che soffre di questo disturbo, fissa l'evento e non raggiungerà una situazione di totale serenità finché non avrà attuato la compulsione.

Le compulsioni vengono per lo più espresse in modo ripetuto con azioni ordinarie, superflue, psicologiche o legate all'ordine e alla simmetria. È emerso da alcuni studi che queste azioni, oltre ad essere ripetute frequentemente, vengono attuate in un ordine spazio-temporale ben definito (Bara e Boringhieri, 2005).

I rituali ossessivi compulsivi presentano tre specifiche caratteristiche:

- ripetitività, nella quale i comportamenti o le espressioni ricorrenti sono ripetute;
- ridondanza, che comprende azioni superflue ripetute, non utili al raggiungimento di uno scopo ben preciso;
- rigidità, per il quale le azioni devono rispettare uno schema ben definito (che non è comune a tutti) senza variazione di alcun genere, cioè devono essere attuate sempre nel medesimo modo.

Identificare una causa unica alla base del disturbo non è possibile poiché varia da individuo a individuo in base al personale profilo psicologico. Ciò che si può affermare è che molti comportamenti ossessivi compulsivi possono essere simili tra uomo e animale. Il lavaggio compulsivo delle mani, che viene preceduto dall'ossessione di contaminazione di microbi o malattie, è uno dei DOC più comuni nell'essere umano, esso può essere messo a confronto in modo parallelo con il DOC animale nel quale il soggetto ripetutamente si lecca il corpo o le zampe col fine di pulirsi. Ad oggi ci sono pochi elementi a livello diagnostico che descrivono il disturbo ossessivo compulsivo negli animali domestici essendo un argomento ancora poco studiato dagli specialisti (Mancini, 2005).

1.1 Il dataset

I dati in seguito analizzati sono stati forniti dalla Clinica Veterinaria *San Marco* di Veggiano (Padova). Fanno riferimento a un campione composto da 26 cani e gatti. I soggetti sono stati scelti accuratamente al fine di avere all'interno del campione alcuni animali affetti da un disturbo ossessivo-compulsivo, precedentemente diagnosticato.

I dati sono stati raccolti mediante uno studio pilota, ovvero uno studio eseguito su un campione con numerosità bassa, nel quale gli specialisti hanno esaminato alcuni video forniti da padroni degli animali, pazienti della Clinica Veterinaria *San Marco*. Durante il periodo di registrazione, che è di differente durata, l'animale esegue determinate azioni che attua quotidianamente. Dopo aver visionato i filmati, sono stati tabulati i dati relativi alla frequenza dei gesti e l'intervallo di tempo trascorso tra l'azione e la ripetizione della stessa. Lo studio può essere catalogato come osservazionale; infatti, i video sono stati forniti su base volontaria, e di tipo caso-controllo, ovvero uno studio nel quale un gruppo di soggetti avente il disturbo ossessivo-compulsivo (casi) viene confrontato con un gruppo di soggetti che non è affetto da questo disturbo (controlli).

L'obiettivo dello studio è codificare i comportamenti in base alla frequenza e alla cadenza, rilevare successivamente quali di questi possono essere associati al DOC dell'animale. Precisamente si vuole rispondere a due quesiti: si osservano frequenze maggiori di attività funzionali nel gruppo dei casi rispetto a quello dei controlli? I comportamenti non funzionali sono attuati con più frequenza e in modo più ravvicinato nel gruppo dei casi rispetto a quello dei controlli?

1.2 Le variabili

Le variabili che vengono prese in considerazione sono le seguenti:

- **Type:** identifica la tipologia di animale che viene preso in esame. Si tratta di una variabile dicotomica che assume i valori:

Type	Meaning
0	Cane
1	Gatto

- **Pathology:** variabile dicotomica che fornisce l'informazione relativa al profilo psicologico del soggetto studiato, ovvero se è affetto da un disturbo ossessivo-compulsivo (DOC) oppure no. Essa assume le seguenti modalità:

Pathology	Meaning
0	No DOC
1	DOC

- **Gender:** definisce il sesso dell'unità statistica presa in esame. È una variabile dicotomica che assume valore:

Gender	Meaning
0	Femmina
1	Maschio

- **Duration:** durata complessiva del periodo di osservazione del soggetto in esame, misurata in secondi. Variabile quantitativa su scala continua.
- **Functional_n:** numero di volte in cui il soggetto ha eseguito un atto funzionale. Variabile quantitativa discreta di conteggio.
- **Functional_intervals:** media dell'intervallo di tempo intercorso tra le azioni funzionali. Variabile quantitativa continua misurata in secondi.
- **Unfunctional_n:** numero di volte in cui il soggetto ha eseguito un atto non funzionale. Variabile quantitativa discreta di conteggio.
- **Unfunctional_intervals:** media dell'intervallo di tempo intercorso tra le azioni non funzionali. Variabile quantitativa continua misurata in secondi.

- **Licking_paw_a_n:** numero di volte in cui il soggetto si è leccato una delle zampe anteriori. Variabile quantitativa discreta di conteggio.
- **Licking_paw_a_intervals:** media dell'intervallo di tempo intercorso tra le azioni: leccare la zampa anteriore. Variabile quantitativa continua misurata in secondi.
- **Licking_paw_p_n:** numero di volte in cui il soggetto si è leccato una delle zampe posteriori. Variabile quantitativa discreta di conteggio.
- **Licking_paw_p_intervals:** media dell'intervallo di tempo intercorso tra le azioni: leccare la zampa posteriore. Variabile quantitativa continua misurata in secondi.
- **Paw_on_face_n:** numero di volte in cui il soggetto pone la propria zampa sopra il muso. Variabile quantitativa discreta di conteggio.
- **Paw_on_face_intervals:** media dell'intervallo di tempo intercorso tra le azioni: porre la zampa sopra il muso. Variabile quantitativa continua misurata in secondi.
- **Up_head_n:** numero di volte in cui il soggetto alza la testa. Variabile quantitativa discreta di conteggio.
- **Up_head_intervals:** media dell'intervallo di tempo intercorso tra le azioni: alzare la testa. Variabile quantitativa continua misurata in secondi.
- **Belly_n:** numero di volte in cui il soggetto si è leccato la pancia. Variabile quantitativa discreta di conteggio.
- **Belly_intervals:** media dell'intervallo di tempo intercorso tra le azioni: leccare la pancia. Variabile quantitativa continua misurata in secondi.
- **Scratching_n:** numero di volte in cui il soggetto ha graffiato. Variabile quantitativa discreta di conteggio.
- **Scratching_intervals:** media dell'intervallo di tempo intercorso tra le azioni: graffiare. Variabile quantitativa continua misurata in secondi.
- **Empty_n:** numero di volte in cui il soggetto ha leccato il vuoto. Variabile quantitativa discreta di conteggio.
- **Empty_intervals:** media dell'intervallo di tempo intercorso tra le azioni: leccare il vuoto. Variabile quantitativa continua misurata in secondi.

Nel prossimo capitolo saranno eseguite le analisi esplorative. L'obiettivo è di individuare eventuali relazioni tra le variabili. Di particolare interesse sono quelle relative alla presenza o assenza della patologia ed i vari comportamenti eseguiti dell'animale.

CAPITOLO 2

ANALISI ESPLORATIVE

Nel presente capitolo si presentano le analisi esplorative. In primo luogo sono considerate le variabili singolarmente per descriverle in sintesi (analisi univariate); in seguito, sono prese in esame le coppie di variabili al fine di rilevare relazioni tra di esse (analisi bivariate).

Gli studi effettuati in questo lavoro sono stati svolti tramite l'utilizzo del software Rstudio (www.rstudio.com).

Si considera un livello di significatività del 5%.

2.1 Analisi univariate

Il campione è suddiviso equamente tra soggetti ai quali è stato diagnosticato il disturbo ossessivo compulsivo (13 unità) rispetto a chi non ne è affetto (13 unità) (variabile *Pathology*).

Analogamente, il campione è omogeneo rispetto alla tipologia di animale domestico (*Type*), ovvero risultano essere presenti 13 cani (50%) e 13 gatti (50%). Per quanto concerne il sesso dei soggetti (*Gender*), si hanno 15 animali di sesso maschile (57.7%) e 11 di sesso femminile (42.3%).

Per le variabili di tipo quantitativo, vengono riportate le principali statistiche di sintesi nella *tabella 2.1*. Viene valutata la normalità mediante il test di Shapiro-Wilk e il diagramma quantile contro quantile. Infine, mediante i boxplot, si possono osservare eventuali asimmetrie e/o outlier.

Variabile	Min	1Qu.	Median	Mean	3Qu.	Max	Sd
Duration	4.88	15.27	33.30	39.17	51.29	134.8	30.96
Functional_n	0.00	18.25	28.00	34.54	48.75	117.0	26.69
Functional_intervals	0.41	0.60	0.77	1.13	1.06	4.53	0.99
Unfunctional_n	0.00	0.00	3.00	8.08	12.50	49.00	12.74
Unfunctional_intervals	0.00	2.11	2.79	3.51	5.25	7.534	2.36
Licking_paw_a_n	0.00	0.00	0.00	10.73	18.50	43.00	15.36
Licking_paw_a_intervals	0.28	0.49	0.63	0.94	0.85	3.44	0.88
Licking_paw_p_n	0.00	0.00	0.00	10.38	7.50	116.0	26.25
Licking_paw_p_intervals	0.38	0.45	0.59	1.28	1.70	3.81	1.27
Paw_on_face_n	0.00	0.00	0.00	2.27	0.00	40.00	7.94
Paw_on_face_intervals	1.21	2.00	2.05	2.17	2.08	3.50	0.83
Up_head_n	0.00	0.00	0.00	0.65	0.75	5.00	1.32
Up_head_intervals	0.99	2.75	3.24	6.911	9.75	17.81	6.94
Belly_n	0.00	0.00	0.00	3.96	5.50	37.00	8.40
Belly_intervals	0.65	0.68	0.69	1.33	1.04	4.53	1.43
Scratching_n	0.00	0.00	0.00	3.85	2.50	37.00	8.82
Scratching_intervals	0.37	1.66	2.39	2.34	2.50	5.30	1.55
Empty_n	0.00	0.00	0.00	3.39	4.00	25.00	6.09
Empty_intervals	1.75	2.79	5.99	5.39	6.045	11.04	2.88

Tabella 2.1: Principali statistiche di sintesi per le variabili quantitative

Esaminando la distribuzione della variabile *Duration* si può notare la presenza di una leggera asimmetria verso destra causata probabilmente dalla presenza di valori anomali visibili nel boxplot (*figura 2.1*). Il diagramma quantile contro quantile (*figura 2.1*) mostra, inoltre, uno scostamento evidente dalla retta di riferimento sia nella coda superiore che inferiore. L'ipotesi di normalità non viene accettata (test di Shapiro-Wilk pari a 0.85, p-value = 0.001).

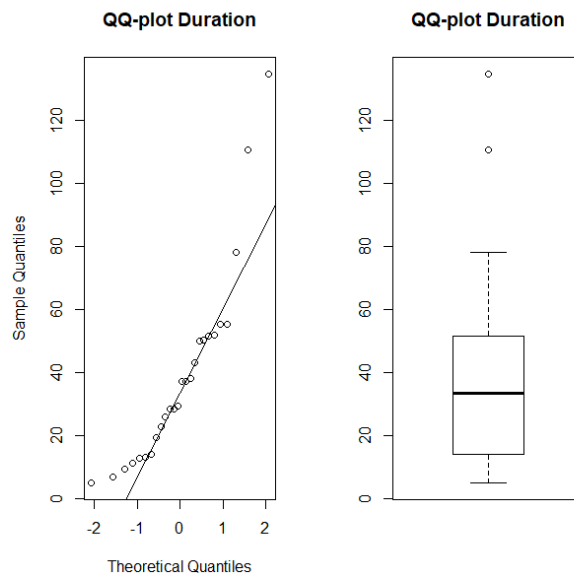


Figura 2.1: Q-Q plot e Boxplot per la variabile Duration

Le restanti variabili quantitative possono essere suddivise in due grandi categorie:

1. le variabili che descrivono la frequenza con la quale avviene una specifica azione nel periodo di osservazione;
2. le variabili che descrivono la media dell'intervallo di tempo intercorso tra le specifiche azioni.

Questo secondo gruppo di variabili non sono rilevate per ogni individuo dello studio. Esse infatti sottendono che ci sia stata un'azione eseguita in precedenza; vale a dire che ogni variabile calcola l'intervallo di tempo trascorso tra l'azione e la ripetizione della stessa, implicando che questi comportamenti siano stati eseguiti dall'unità statistica. In conclusione, ogni variabile del secondo gruppo è legata strettamente a quelle del primo, ovvero per ogni soggetto che attua un'azione con frequenza maggiore di 1 viene calcolata anche la rispettiva media dell'intervallo di tempo.

Riferendosi solo al primo raggruppamento di variabili, ossia quelle che descrivono la quantità di volte che viene ripetuta una specifica azione, possono essere costruite due variabili, ovvero *Functional_n* e *Unfunctional_n*.

La prima si riferisce a tutte le azioni funzionali, e quindi accorpa le variabili *Licking_paw_p_n*, *Licking_paw_a_n*, *Paw_on_face_n*, *Belly_n* e altri comportamenti non menzionati. La seconda rappresenta invece tutte le azioni non funzionali, quindi accorpa le variabili *Up_head_n*, *Scratching_n*, *Empty_n* e altri atti non menzionati.

Studiando la variabile *Functional_n*, dal diagramma quantile contro quantile, non si nota alcun scostamento dalla retta di riferimento. Osservando il boxplot, invece, si evidenzia una lieve asimmetria a destra con presenza di un solo outlier (*figura 2.2*). Anche in questo caso si rifiuta l'ipotesi di normalità (test di Shapiro-Wilk pari a 0.91 e p-value = 0.027).

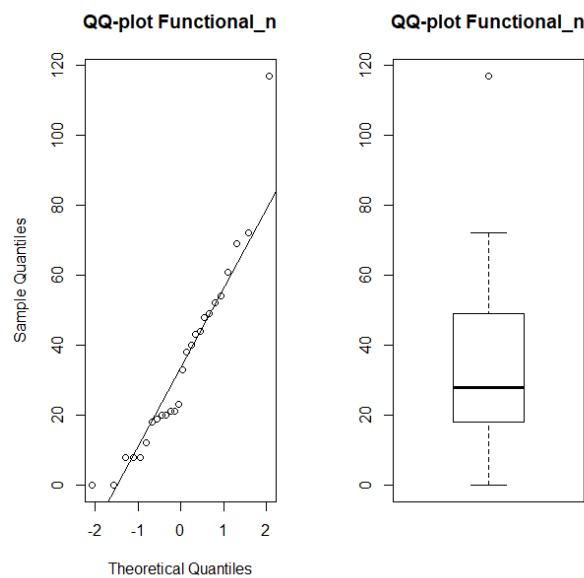


Figura 2.2: Q-Q plot e Boxplot per la variabile Functional_n

Dal diagramma quantile contro quantile per la variabile *Unfunctional_n* si nota uno scostamento dalla retta di riferimento in entrambe le code. Il boxplot mostra asimmetria caratterizzata da una coda destra lunga e presenza di valori anomali (*figura 2.3*). Il test di Shapiro-Wilk assume valore pari a 0.68 con p-value <0.001, pertanto viene rifiutata l'ipotesi di normalità.

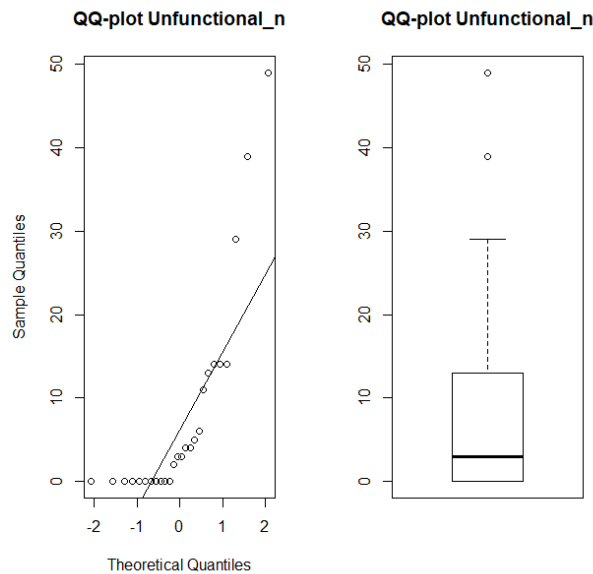


Figura 2.3: *Q-Q plot e Boxplot per la variabile Unfunctional_n*

Prendendo in esame le variabili *Licking_paw_a_n*, *Licking_paw_p_n*, *Paw_on_face_n*, *Up_head_n*, *Belly_n*, *Scratching_n* e *Empty_n* sono stati raggruppati i risultati delle analisi esplorative nelle *tabella 2.2* e *figura 2.4*. Analizzando i boxplot presenti in *figura 2.4* si può osservare la presenza di asimmetrie verso destra per ogni variabile considerata, con presenza di outlier e code lunghe a destra. Per nessuna variabile è accettata l'ipotesi di normalità risultando i vari p-value del test di Shapiro-Wilk <0.05 (*tabella 2.2*).

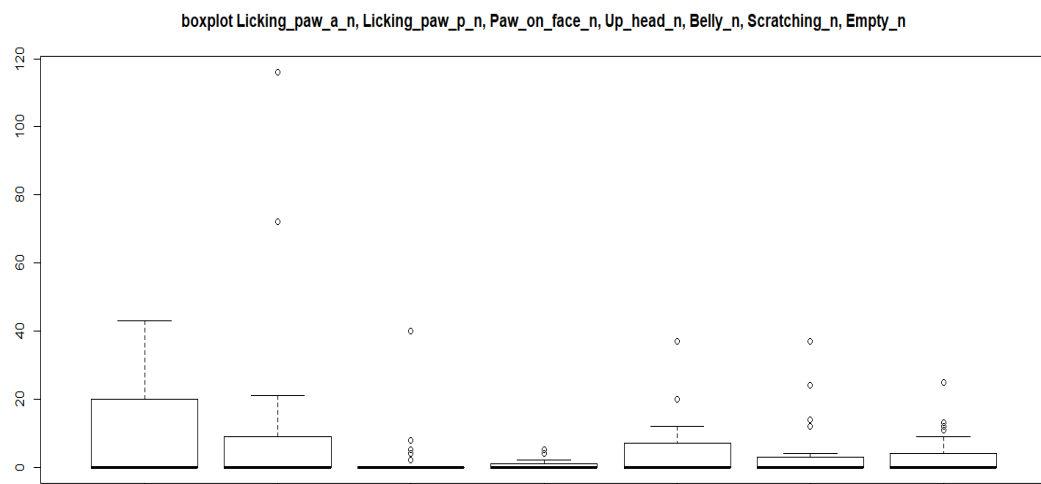


Figura 2.4: *Boxplot per le variabili Licking_paw_a_n, Licking_paw_p_n, Paw_on_face_n, Up_head_n, Belly_n, Scratching_n, Empty_n*

Variabile	Shapiro-Wilk normality test	
	SW ^{oss}	p-value
Licking_paw_a_n	0.72	<0.05
Licking_paw_p_n	0.46	<0.05
Paw_on_face_n	0.32	<0.05
Up_head_n	0.57	<0.05
Belly_n	0.55	<0.05
Scratching_n	0.52	<0.05
Empty_n	0.64	<0.05

Tabella 2.2: Valori del test di Shapiro-Wilk e rispettivi p-value per le variabili *Licking_paw_a_n*, *Licking_paw_p_n*, *Paw_on_face_n*, *Up_head_n*, *Belly_n*, *Scratching_n*, *Empty_n*

Considerando il secondo raggruppamento di variabili, ossia quelle che descrivono la media dell'intervallo di tempo intercorso tra le azioni, possono essere costruite le variabili *Functional_intervals* e *Unfunctional_intervals*.

Pertanto, *Functional_intervals* si riferisce a tutte le azioni funzionali, e quindi accorpa le variabili *Licking_paw_p_intervals*, *Licking_paw_a_intervals*, *Paw_on_face_intervals*, *Belly_intervals* e altri comportamenti non menzionati.

Unfunctional_intervals rappresenta, invece, tutte le azioni non funzionali, e quindi accorpa le variabili *Up_head_intervals*, *Scratching_intervals*, *Empty_intervals* e altri atti non menzionati.

In figura 2.5 vengono mostrati il diagramma quantile contro quantile e il boxplot relativi alla variabile *Functional_intervals*. Nel primo grafico si nota uno scostamento dalla retta di riferimento nella coda destra, mentre nel boxplot si nota una coda sinistra leggermente più lunga e un'asimmetria verso destra probabilmente causata dalla presenza di outlier. Non è accettata l'ipotesi di normalità (test di Shapiro-Wilk pari a 0.69 e p-value <0.001).

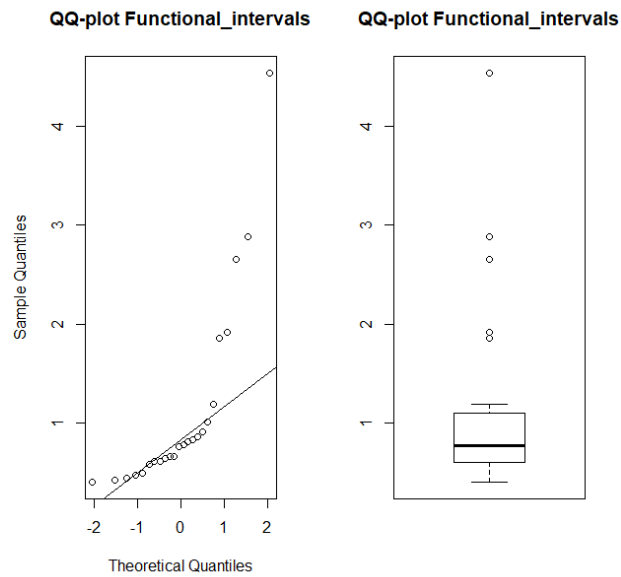


Figura 2.5: *Q-Q plot e Boxplot per la variabile `Functional_intervals`*

Studiando la variabile *Unfunctional_intervals* si nota, dal boxplot, una leggera asimmetria verso destra; in questo caso però non sono presenti valori anomali. Osservando il diagramma quantile contro quantile non sono presenti scostamenti dalla retta di riferimento (figura 2.6). Si accetta l'ipotesi di normalità, con il test di Shapiro-Wilk pari a 0.94 e $p\text{-value} = 0.443$.

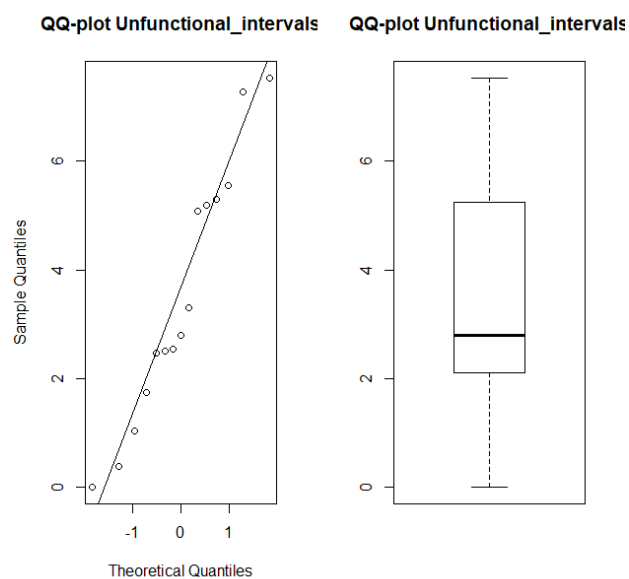


Figura 2.6: *Q-Q plot e Boxplot per la variabile `Unfunctional_intervals`*

Prendendo in considerazione le variabili *Licking_paw_a_intervals*, *Licking_paw_p_intervals*, *Paw_on_face_intervals*, *Up_head_intervals*, *Belly_intervals*, *Scratching_intervals* e *Empty_intervals* sono stati raccolti i risultati delle analisi esplorative nella *tabella 2.3* e *figura 2.7*. Nella *figura 2.7* sono rappresentati i boxplot, per ciascuna variabile, che evidenziano asimmetria a destra per le variabili *Licking_paw_p_intervals*, *Up_head_intervals* e *Belly_intervals* con presenza di code destre più lunghe; viceversa, *Scratching_intervals* e *Empty_intervals* presentano asimmetria verso sinistra e per le rimanenti due, ovvero *Licking_paw_a_intervals* e *Up_head_intervals*, non si rilevano particolari tendenze. Sono presenti valori anomali in tutte le variabili, fatta eccezione per *Licking_paw_p_intervals* e *Up_head_intervals*.

In *tabella 2.3* sono riportati i valori della statistica test di Shapiro-Wilk e i rispettivi p-value. Si può notare che l'ipotesi di normalità viene accettata per le variabili *Paw_on_face_intervals*, *Up_head_intervals*, *Scratching_intervals* e *Empty_intervals*.

boxplot Licking_paw_a_intervals, Licking_paw_p_intervals, Paw_on_face_intervals, Up_head_intervals, Belly_intervals, Scratching_intervals, Empty_intervals

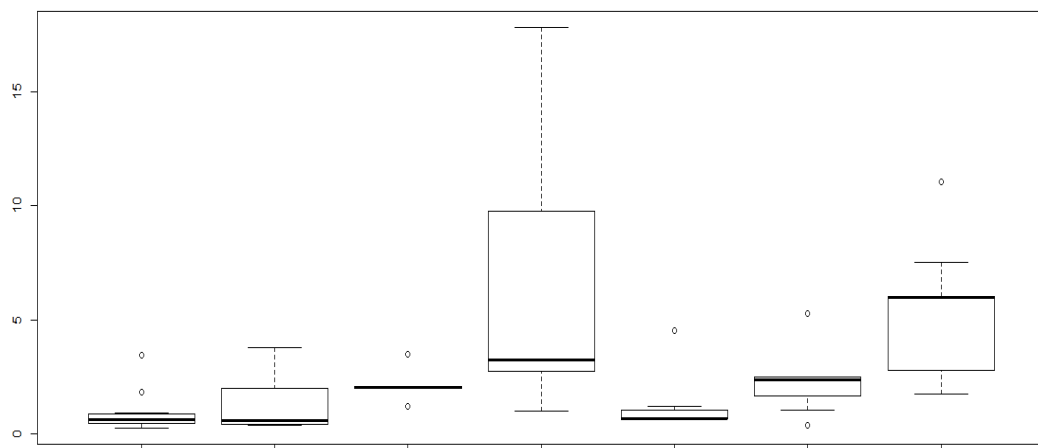


Figura 2.7: Boxplot per le variabili Licking_paw_a_intervals, Licking_paw_p_intervals, Paw_on_face_intervals, Up_head_intervals, Belly_intervals, Scratching_intervals, Empty_intervals

Variabile	Shapiro-Wilk normality test	
	SW ^{oss}	p-value
Licking_paw_a_intervals	0.66	<0.05
Licking_paw_p_intervals	0.76	0.01
Paw_on_face_intervals	0.87	0.25
Up_head_intervals	0.86	0.23
Belly_intervals	0.56	<0.05
Scratching_intervals	0.87	0.20
Empty_intervals	0.93	0.48

Tabella 2.3: Valori dei test di Shapiro-Wilk e i rispettivi p-value per le variabili *Licking_paw_a_intervals*, *Licking_paw_p_intervals*, *Paw_on_face_intervals*, *Up_head_intervals*, *Belly_intervals*, *Scratching_intervals*, *Empty_intervals*

2.2 Analisi bivariate

Di seguito sono analizzate le coppie di variabili al fine di osservare le relazioni e individuare quelle significative; ciò viene eseguito usufruendo di strumenti differenti in base alla natura delle variabili in questione.

È di interesse analizzare in questo studio la relazione tra il disturbo ossessivo compulsivo animale e precisi comportamenti che sono stati rilevati (frequenza e intervallo di tempo). A tale scopo, in primo luogo verranno studiate le relazioni tra la variabile che codifica la presenza o meno del disturbo (*Phatology*), che è quella di interesse primario, con tutte le altre.

Come prima coppia si considerano *Phatology* e *Type*. Nella *tabella 2.4* vengono riportate le frequenze assolute e le percentuali condizionate per colonna (valori tra parentesi). Si può affermare che tra i soggetti che hanno il disturbo ossessivo-compulsivo, il 54% sono gatti e il 46% sono cani, viceversa tra chi non è affetto da DOC sono il 46% di gatti e il 54% di cani. Il p-value del test esatto di Fischer porta a rifiutare l'ipotesi di associazione tra *Pathology* e *Type* essendo questo prossimo a uno.

Type	Pathology	
	No OCD	OCD
Cane	6 (0.46)	7 (0.54)
Gatto	7 (0.54)	6 (0.46)

Tabella 2.4: Tabella contingenza di Pathology-Type, tra parentesi le percentuali condizionate per colonna

Per le due variabili dicotomiche *Pathology* e *Gender* vengono riportate le frequenze assolute e le percentuali condizionate per colonna (valori tra parentesi) nella *tabella 2.5*. I risultati ottenuti indicano che tra i soggetti che non hanno il disturbo il 38% è di sesso femminile e il 62% è di sesso maschile; per quanto riguarda coloro che hanno questa malattia il 46% è femmina e il 54% è maschio. Il p-value del test esatto di Fischer porta a rifiutare l'ipotesi di associazione tra Pathology e Gender essendo questo prossimo a 1.

Gender	Pathology	
	No OCD	OCD
Femmina	5 (0.38)	6 (0.46)
Maschio	8 (0.62)	7 (0.54)

Tabella 2.5: Tabella di contingenza di Pathology-Gender, tra parentesi le percentuali condizionate per colonna

Per studiare coppie di variabili di natura dicotomica-quantitativa si vanno ad analizzare due gruppi suddivisi in base alle modalità della variabile dicotomica.

Viene valutata la normalità mediante il test di Shapiro-Wilk e il diagramma quantile contro quantile per entrambi i gruppi. In aggiunta vengono eseguite anche analisi grafiche tramite boxplot al fine di osservare eventuali outlier e/o asimmetrie. Infine si procede con l'applicazione di un test con lo scopo di confrontare le distribuzioni dei gruppi. Viene usato il test non parametrico di Mann-Whitney se l'ipotesi di normalità non è stata accettata, in alternativa viene

applicato il test di Welch se è accettata l'ipotesi di normalità ma non quella di omoschedasticità, oppure il test t a due campioni indipendenti se l'ipotesi di normalità e di omoschedasticità sono state accettate. Per verificare l'ipotesi di omoschedasticità si utilizza il test F.

In *figura 2.8* sono riportati i boxplot e i diagrammi quantile-quantile di *Duration* per *Pathology*. Dal boxplot si può osservare un'asimmetria verso destra per entrambi i gruppi con presenza di outlier. Il secondo gruppo, ovvero coloro che sono affetti da DOC, presenta la coda destra più lunga. I diagrammi quantile contro quantile mostrano un leggero allontanamento dalla retta di riferimento in entrambi i gruppi. Essendo entrambi i p-value < 0.05 del test di Shapiro-Wilk viene rifiutata l'ipotesi di normalità (NO DOC $SW^{oss} = 0.84$ e p-value=0.02; per DOC $SW^{oss} = 0.79$ e p-value = 0.006).

Con un p-value pari a 0.34 del test di Mann-Whitney si afferma che la distribuzione della durata in secondi del video fornito è significativamente uguale nei due gruppi.

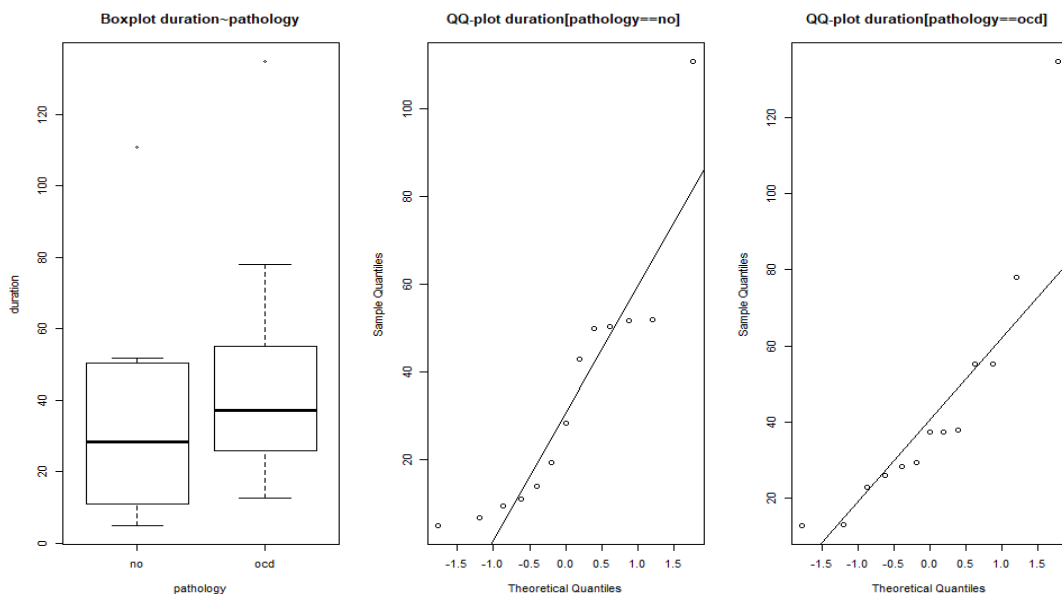


Figura 2.8: Q-Qplot e Boxplot e di Duration distintamente per DOC e NON DOC

Prendendo *Functional_n* per *Pathology* appare che in entrambi i gruppi ci sia una asimmetria: verso destra nel primo gruppo, al contrario verso sinistra nel secondo. Solo nel primo gruppo sono presenti valori anomali. Il grafico quantile-quantile di coloro che non hanno il disturbo ossessivo compulsivo mostra un allontanamento sistematico dalla retta di riferimento, in particolar modo nella coda sinistra (*figura 2.9*). Il p-value del test di Shapiro-Wilk è pari a 0.004 ($SW^{OSS}=0.78$) nel primo gruppo per il quale viene rifiutata l'ipotesi di normalità ed è uguale a 0.59 ($SW^{OSS}=0.95$) nel secondo gruppo per il quale l'ipotesi di normalità viene accettata. Con il valore osservato pari a 67 e p-value uguale a 0.38, con il test di Mann-Whitney si conclude che la distribuzione della frequenza delle azioni funzionali nei due gruppi è significativamente uguale.

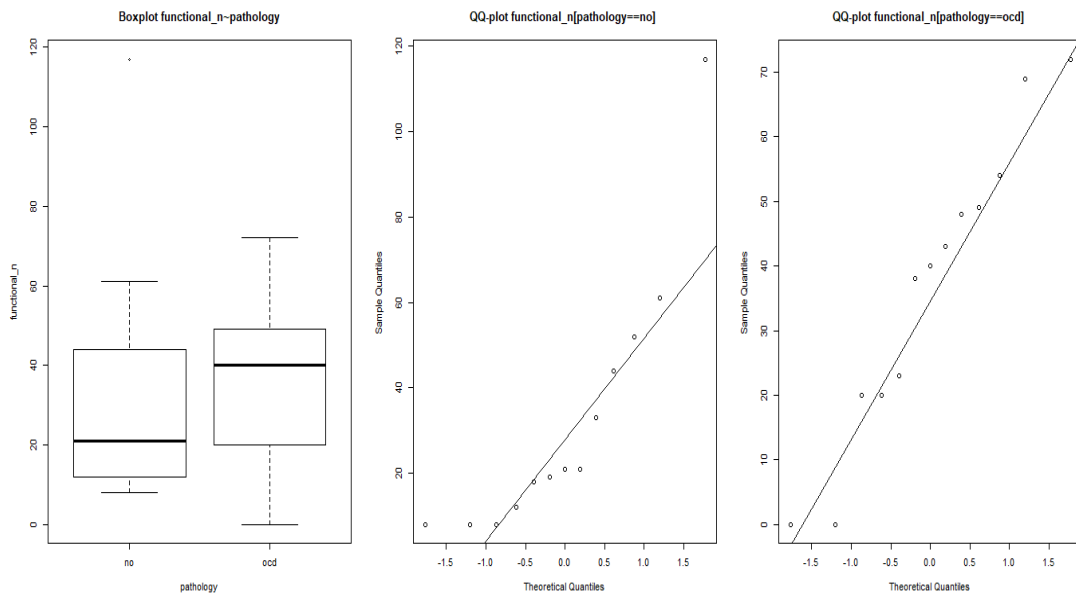


Figura 2.9: Q-Qplot e Boxplot e di *Functional_n* distintamente per DOC e NON DOC

Dalla *figura 2.10* si deduce come la frequenza delle azioni non funzionali sia inferiore negli individui che non hanno il disturbo ossessivo compulsivo (*Unfunctional_n Pathology*). Il test di Shapiro-Wilk rifiuta l'ipotesi di normalità in entrambi i gruppi risultando i p-value<0.05 (NO DOC $SW^{OSS} = 0.52$ e p-value<0.001; DOC $SW^{OSS} = 0.8$ e p-value = 0.008). I diagrammi quantile contro quantile mostrano un allontanamento sistematico dalla retta di riferimento.

Il test di Mann-Whitney non porta al rifiuto dell'ipotesi di uguaglianza delle mediane nei due gruppi ($W^{\text{OSS}} = 61.5$ e $p\text{-value} = 0.23$).

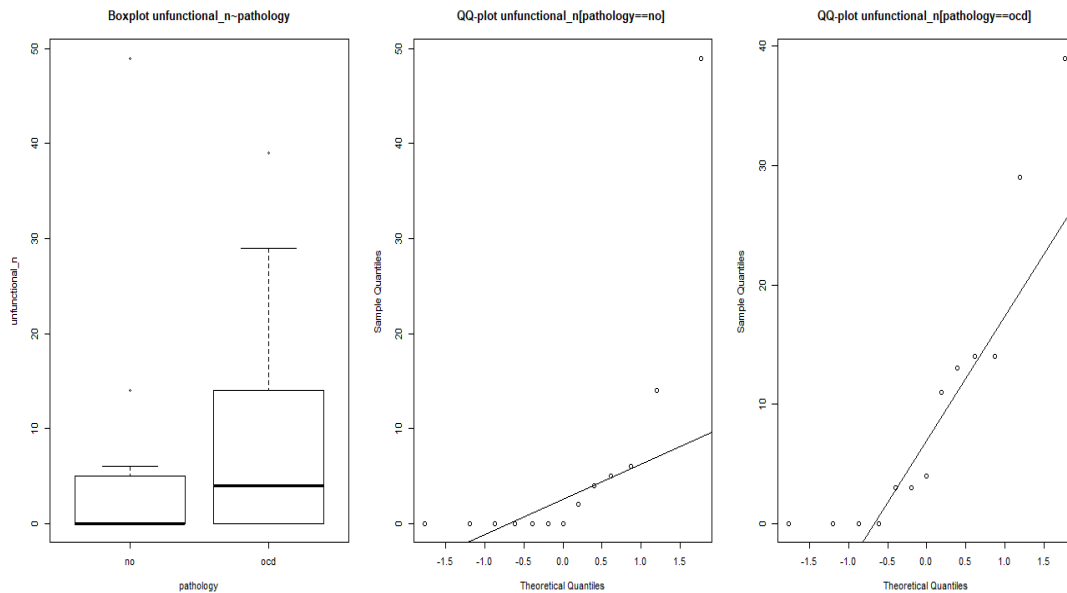


Figura 2.10: Q-Qplot e Boxplot e di *Unfunctional_n* distintamente per DOC e NON DOC

La *tabella 2.6* contiene i risultati dell'analisi riguardante la relazione tra *Phatology* e le restanti variabili quantitative che identificano la frequenza dei vari comportamenti (*Licking_paw_a_n*, *Licking_paw_p_n*, *Pow_on_face_n*, *Up_head_n*, *Belly_n*, *Scratching_n*, *Empty_n*). nella *tabella 2.7* sono riportati i p-value dei test svolti. Da questa tabella si evince che per quasi tutte le variabili la mediana risulta pari a 0 in entrambi i gruppi, ciò significa che più della metà delle unità statistiche non compiono mai l'azione d'interesse nel periodo di osservazione, indipendentemente che si tratti di soggetti con DOC oppure no. Unica eccezione è la variabile *Licking_paw_a_n* che ha mediana pari a 3.00 nel gruppo di coloro che non sono affetti dal disturbo. Per quanto riguarda il confronto delle medie nei due gruppi, la differenza più evidente è presente nella variabile *Paw_on_face_n* che ha media pari a 0.38 negli animali sani rispetto a 4.15 nei soggetti malati e nella variabile *Empty_n* che risulta essere pari a 1.23 per i soggetti senza DOC rispetto

a 5.54 per coloro con DOC. Di conseguenza si può affermare che le precedenti azioni sono mediamente più frequenti nei soggetti che hanno un disturbo ossessivo compulsivo. Questa conclusione vale anche per le altre azioni, fatta eccezione per *Licking_paw_p_n* e *Scratching_n* per le quali si osserva il viceversa. Essi infatti hanno media maggiore nel gruppo di sani rispetto al gruppo con disturbo ossessivo compulsivo. Nessun p-value del test di Shapiro-Wilk consente l'accettazione dell'ipotesi di normalità della distribuzione nei vari gruppi. Ogni test di Mann-Whitney eseguito porta all'accettazione dell'ipotesi di uguaglianza delle mediane nei due gruppi.

Variabile	Phatology	1Qu.	Mean	Median	3 Qu.	Sd
Licking_paw_a_n	NO DOC	0	9.54	3.00	14	13.07
	DOC	0	11.92	0	20	17.83
Licking_paw_p_n	NO DOC	0	11.31	0	11.31	32.06
	DOC	0	9.46	0	9.46	20.16
Paw_on_face_n	NO DOC	0	0.38	0	0	1.39
	DOC	0	4.15	0	2.00	11.03
Up_head_n	NO DOC	0	0.62	0	0	1.45
	DOC	0	0.96	0	1	1.25
Belly_n	NO DOC	0	2.77	0	7	4.30
	DOC	0	5.15	0	1	11.20
Scratching_n	NO DOC	0	4.31	0	1	10.57
	DOC	0	3.36	0	3	7.05
Empty_n	NO DOC	0	1.23	0	1	2.28
	DOC	0	5.54	0	11	7.88

Tabella 2.6: Principali statistiche di sintesi per le variabili quantitative Linking_paw_a_n, Linking_paw_p_n, Paw_on_face_n, Up_head_n, Belly_n, Scratching_n, Empty_n distintamente per DOC e NON DOC

Variabile	Phatology	p-value Shapiro- Wilk	p-value Mann- Whitney
Licking_paw_a_n	NO DOC	0.003	0.91
	DOC	0.0004	
Licking_paw_p_n	NO DOC	<0.001	0.55
	DOC	<0.001	
Paw_on_face_n	NO DOC	<0.001	0.16
	DOC	<0.001	
Up_head_n	NO DOC	<0.001	0.72
	DOC	<0.001	
Belly_n	NO DOC	0.0004	0.92
	DOC	<0.001	
Scratching_n	NO DOC	<0.001	0.81
	DOC	<0.001	
Empty_n	NO DOC	<0.001	0.21
	DOC	0.002	

Tabella 2.7: P-value relativo al test di Shapiro-Wilk e Mann-Whitney di Linking_paw_a_n, Linking_paw_p_n, Paw_on_face_n, Up_head_n, Belly_n, Scratching_n, Empty_n distintamente per DOC e NON DOC

Prendendo la variabile *Functional_intervals* per *Pathology* appare che nel primo gruppo ci sia una leggera asimmetria verso sinistra con rispettiva coda più lunga; nel secondo, invece, non si rilevano particolari tendenze. Diversamente accade per gli outlier che sono presenti in entrambi i gruppi. I due grafici quantile contro quantile mostrano un allontanamento sistematico dalla retta di riferimento, in particolar modo nella coda destra (*figura 2.11*). Il p-value del test di Shapiro-Wilk è pari a 0.0003 ($SW^{OSS}=0.67$) nel primo gruppo e uguale a 0.004 ($SW^{OSS}=0.77$) nel secondo gruppo; perciò in entrambi i casi l'ipotesi di normalità viene rifiutata con livello di significatività al 5%.

Con il valore osservato pari a 71 e p-value uguale a 1, con il test di Mann-Whitney si conclude che la distribuzione della media dell'intervallo di tempo intercorso tra azioni funzionali nei due gruppi è significativamente uguale.

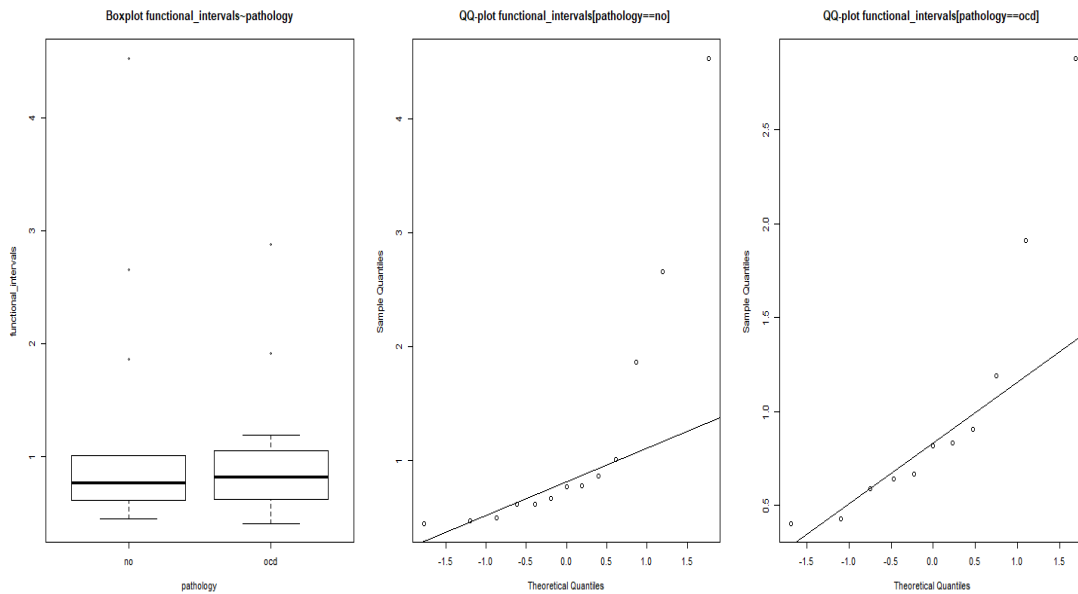


Figura 2.11: Q-Qplot e Boxplot e di *Functional_intervals* distintamente per DOC e NON DOC

Dalla figura 2.12 si deduce che la distribuzione dei valori relativi alla variabile *Unfunctional_intervals* nei due gruppi, suddivisi in base a *Pathology*, è leggermente asimmetrica verso destra nel primo gruppo e verso sinistra nel secondo gruppo. Il test di Shapiro-Wilk accetta l'ipotesi di normalità in entrambi i gruppi (NO DOC $SW^{OSS} = 0.95$ e p-value = 0.75; DOC $SW^{OSS} = 0.95$ e p-value = 0.67) e i diagrammi quantile-quantile non mostrano alcun allontanamento sistematico dalla retta di riferimento. Applicando il test F per l'uguaglianza delle varianze il p-value risulta essere pari 0.45 e di conseguenza si accetta l'ipotesi di omoschedasticità. Il test *t* di Student porta a concludere che la distribuzione dell'intervallo di tempo delle azioni non funzionali nei due gruppi è significativamente uguale ($T^{OSS} = 0.06$ e p-value = 0.96).

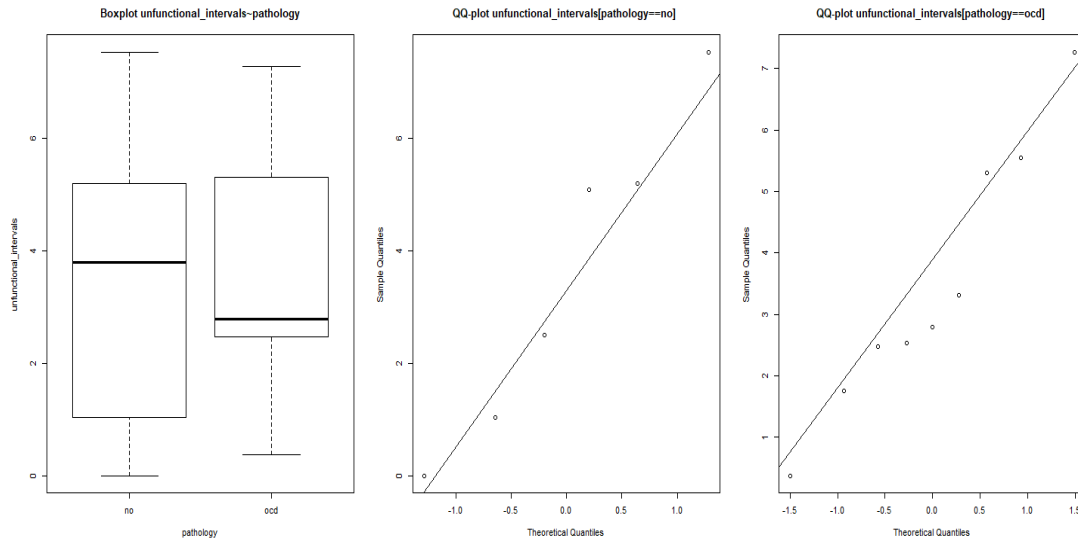


Figura 2.12: Q-Qplot e Boxplot e di *Unfunctional_intervals* distintamente per DOC e NON DOC

La *tabella 2.8* contiene i risultati dell'analisi riguardante la relazione tra *Phatology* e le restanti variabili quantitative che identificano la media dell'intervallo di tempo intercorso tra i vari atti (*Licking_paw_a_intervals* *Licking_paw_p_intervals* *Belly_intervals* *Scratching_intervals* *Empty_intervals*). Nella *tabella 2.9* sono riportati i p-value dei test svolti. Uniche due variabili escluse da queste analisi bivariate sono *Pow_on_face_intervals* e *Up_head_intervals*; ciò è causato dalla numerosità bassa presente nei gruppi. Da questa tabella si evince che valori più alti della media sono presenti nelle variabili *Empty_intervals* e *Scratching_intervals* ciò significa che queste azioni sono svolte dalle unità statistiche in modo meno ravvicinato l'una dall'altra; il contrario avviene per le variabili *Licking_paw_a_intervals*, *Licking_paw_p_intervals* e *Belly_intervals* che hanno medie più basse. Di conseguenza queste azioni vengano eseguite più volte in un arco temporale piccolo. Per quanto riguarda il confronto delle medie nei due gruppi creati, studiati rispetto alle varie variabili, la differenza più evidente è presente in *Empty_intervals* che ha media pari a 8.19 negli animali senza DOC rispetto a 3.99 per coloro con DOC. Di conseguenza, si può affermare che l'azione precedente è mediamente eseguita in modo più ravvicinato nei soggetti che hanno un disturbo ossessivo compulsivo. Questa conclusione può essere valida anche per le altre azioni, fatta eccezione per *Licking_paw_a_intervals* e *Scratching_intervals* per le quali vale il contrario. Essi

infatti hanno media minore nel gruppo di sani rispetto al gruppo con disturbo ossessivo compulsivo. Il test di Shapiro-Wilk porta all'accettazione dell'ipotesi di normalità in tutti i sottogruppi ad eccezione di quelli creati studiando le variabili Licking_pow_a_intervals e Belly_intervals. Per queste ultime il test di Mann-Whitney porta all'accettazione dell'ipotesi di uguaglianza delle mediane nei due gruppi. Applicando il test F, il p-value risulta essere pari a 0.24 per la variabile Licking_paw_p_intervals, 0.28 per la variabile Scratching_intervals e 0.49 per Empty_intervals; di conseguenza viene sempre accettata l'ipotesi di omoschedasticità. Il test *t* di Student porta a concludere che la distribuzione dell'intervallo di tempo delle varie azioni nei due gruppi è significativamente uguale (Paw_p_intervals $T^{oss} = 0.5$ e $p - value = 0.65$; Scratching_intervals $T^{oss} = -0.57$ e $p - value = 0.60$; Empty_intervals $T^{oss} = 2.5$ e $p - value = 0.09$).

Variabile	Phatology	1Qu.	Mean	Median	3 Qu.	Sd
Licking_paw_a_intervals	NO DOC	0.48	0.73	0.61	0.63	0.50
	DOC	0.70	1.23	0.83	0.9	1.26
Licking_paw_p_intervals	NO DOC	0.57	1.65	0.67	2.24	1.87
	DOC	0.41	1.06	0.5	1.4	0.96
Belly_intervals	NO DOC	0.82	1.83	1.04	2.04	1.82
	DOC	0.66	0.67	0.67	0.68	0.02
Scratching_intervals	NO DOC	1.71	1.98	2.39	2.45	0.81
	DOC	1.80	2.61	2.39	3.20	2.03
Empty_intervals	NO DOC	6.77	8.19	7.53	9.29	2.58
	DOC	2.55	3.99	3.83	3.99	1.89

Tabella 2.8: Principali statistiche di sintesi per le variabili quantitative Linking_paw_a_intervals, Linking_paw_p_intervals, Belly_n, Scratching_intervals, Empty_intervals distintamente per DOC e NON DOC

Variabile	Phatology	p-value Shapiro- Wilk	p-value Mann- Whitney	p-value del test F	p-value del test <i>t</i>
Licking_paw_a_intervals	NO DOC	0.0003	0.27	/	/
	DOC	0.02			
Licking_paw_p_intervals	NO DOC	0.1	/	0.24	0.65
	DOC	0.07			
Belly_intervals	NO DOC	0.03	0.06	/	/
	DOC	0.72			
Scratching_intervals	NO DOC	0.13	/	0.28	0.60
	DOC	0.70			
Empty_intervals	NO DOC	0.58	/	0.49	0.09
	DOC	0.23			

Tabella 2.9: P-value relativo al test di Shapiro-Wilk e Mann-Whitney di Linking_paw_a_intervals, Linking_paw_p_intervals, Belly_n, Scratching_intervals, Empty_intervals distintamente per DOC e NON DOC

Successivamente verranno eseguite ulteriori analisi bivariate prendendo tutte le restanti combinazioni di coppie di variabili. Essendo una numerosità elevata di confronti verranno riportati soltanto i risultati significativi.

Analizzando tutte le combinazioni di coppie di variabili quantitative viene studiata la presenza o meno di correlazione. A tal fine, ci si serve del coefficiente di correlazione di Spearman (ρ). Viene anche eseguito il test che studia l'ipotesi che la correlazione sia significativamente uguale a 0.

Nella *tabella 2.8* sono riportate le coppie di variabili per le quali le misure di ρ risultano essere significative al 5%. Dalla tabella si può notare che ogni valore di ρ ottenuto ha segno positivo, di conseguenza tra le variabili prese in considerazione si ha sempre una relazione direttamente proporzionale.

Coppie di variabili	Rho	P-value test di Spearman
Empty_intervals Unfunctional_intervals	0.73	0.031
Up_head_n Empty_n	0.59	0.002
Paw_on_face_n Empty_n	0.58	0.002
Paw_on_face_n Functional_intervals	0.48	0.02
Licking_paw_p_n Scratching_n	0.43	0.03
Licking_paw_p_n Empty_n	0.43	0.03
Licking_paw_a_n Empty_n	0.41	0.04
Unfunctional_n Licking_paw_p_n	0.57	0.003
Unfunctional_n Up_head_n	0.62	<0.05
Unfunctional_n Scratching_n	0.61	<0.05
Unfunctional_n Empty_n	0.73	<0.05
Functional_n Unfunctional_n	0.53	0.005
Functional_n Licking_paw_p_n	0.44	0.02
Functional_n Scratching_n	0.53	0.005
Functional_n Empty_n	0.43	0.03
Duration Functional_n	0.40	0.04
Duration Unifonctinal_n	0.64	<0.05
Duration Up_head_n	0.44	0.02
Duration Empty_n	0.53	0.005
Duration Functional_intervals	0.49	0.02

Tabella 2.8: coefficiente di correlazione di Spearman (ρ) e rispettivo p-value del test relativo alle coppie di variabili per la quale le misure di ρ risultano essere significative

Il prossimo capitolo verte sull'implementazione e lo studio dei modelli di regressione specifici dataset di piccole dimensioni. L'obiettivo è di stimare le relazioni tra le covariate e la presenza o meno del disturbo ossessivo compulsivo

CAPITOLO 3

MODELLI DI REGRESSIONE LOGISTICA

3.1 La regressione logistica

I modelli lineari generalizzati sono una ampia classe di modelli che hanno come caratteristica principale quella di avere la variabile risposta Y non gaussiana. A seconda della distribuzione della risposta sono possibili diversi livelli di specificazione, i più utilizzati nell'ambito inferenziale sono i modelli di regressione logistica (o modelli logit), per variabili risposta dicotomiche (Salvan *et al.*, 2020).

I modelli di regressione logistica sono quindi impiegati quando la variabile risposta è dicotomica. Tali modelli vengono utilizzati al fine di esprimere una relazione tra la media di Y , che corrisponde alla probabilità di successo π , e le p variabili esplicative $(x_1, x_2, \dots, x_p)$.

Questi modelli hanno specifiche caratteristiche:

- La variabile risposta Y ha distribuzione di Bernoulli, può dunque assumere valori $\{0,1\}$.
- Si assume $Y = (Y_1, Y_2, \dots, Y_n)$, con Y_i variabili casuali indipendenti.
- Essendo $Y_i \sim Be(\pi_i)$, vale

$$P(Y_i = y_i) = \begin{cases} \pi_i & y_i=1 \\ 1 - \pi_i & y_i=0 \end{cases}$$

- Il modello descrive la dipendenza di π (probabilità di successo) dai valori che le variabili esplicative $x_1, \dots, x_p$ assumono.
- La funzione di legame del modello prevede

$$g(\pi_i) = \beta_1 x_{i1} + \dots + \beta_p x_{ip}.$$

Più specificatamente per il modello in questione si usa la funzione di legame logit:

$$g(\pi_i) = \text{logit}(\pi_i) = \log(\pi_i / (1 - \pi_i)),$$

la cui inversa corrisponde a

$$\pi_i = \Pr(Y_i = 1) = (e^{\beta^T x_i}) / (1 + e^{\beta^T x_i}),$$

dove $\beta = (\beta_1, \dots, \beta_p) \in R^p$ sono parametri ignoti.

I modelli di regressione logistica sono maggiormente utilizzati perché permettono una facile interpretazione del predittore lineare $\eta_i = g(\pi_i) = \log(\pi_i / (1 - \pi_i)) = \beta^T x_i$. Ciò infatti è inteso come logaritmo della quota $(\pi_i / (1 - \pi_i))$ che è il rapporto tra probabilità di successo e di insuccesso. Di conseguenza, un generico coefficiente β_r esprime l'effetto che ha un incremento unitario di x_{ir} sul logaritmo della quota (Salvan *et al.*, 2020).

3.2 La correzione di Firth

I modelli di regressione logistica possono imbattersi in un fenomeno denominato “problema di separazione” che è caratterizzato dalla monotonia della funzione di verosimiglianza (Heinze e Schemper, 2002).

Questo fenomeno si presenta maggiormente in campioni di piccole dimensioni e può causare problemi di stima dei coefficienti di regressione.

Esistono due tipologie di separazione:

- quella completa, che avviene quando la combinazione dei predittori genera una previsione perfetta;
- quella quasi-completa, che avviene quando quasi tutte le osservazioni vengono classificate correttamente.

Per risolvere i problemi creati dal fenomeno di separazione si può condurre una regressione logistica con la correzione di Firth (1993). Firth scoprì che, tramite una apposita modifica della funzione punteggio, il termine che causa la distorsione dello stimatore di massima verosimiglianza può essere rimosso.

Il pacchetto di R “*logistf*” fornisce uno strumento adatto a facilitare l’applicazione della correzione di Firth nella stima dei modelli di regressione logistica. Per eseguire questa implementazione viene utilizzata la funzione *logistf(formula, data,)* passando come *data* il dataset nella quale sono presenti i dati e come *formula* una descrizione del modello da adattare (Heinze, 2022).

3.3 Modello logistico con correzione di Firth

Attraverso il modello di regressione logistica si vuole studiare l’effetto che le variabili esplicative hanno su quella dipendente. Per quanto riguarda lo studio presente nell’elaborato, la variabile risposta è la variabile dicotomica "Pathology" che registra la presenza o meno del disturbo ossessivo compulsivo nell’animale. In particolare, vale 0 se il l’unità statistica non ha il DOC (sano); vale 1 se il soggetto presenta la patologia (malato).

Nell’implementazione del modello logistico viene eseguito un approccio backward, considerando un livello di significatività del 10%, per il quale viene data maggiore attenzione alle variabili relative alla frequenza dei comportamenti osservati. Nell’applicare al dataset il modello logistico sono stati ottenuti, come stime dei coefficienti e dei relativi standard error, valori particolarmente elevati; perciò si è deciso di stimare il modello logistico tramite la correzione di Firth.

Durante lo svolgimento della selezione all’indietro viene inizialmente considerato il modello con tutte le variabili esplicative, eccetto le variabili che misurano la media dell’intervallo di tempo intercorso tra gli atti. Successivamente sono state eliminate, di volta in volta, le variabili che risultano avere un minor valore della statistica test del rapporto di verosimiglianza penalizzata (corretta secondo il metodo di Firth).

Si è giunti a un modello finale che comprende le sole variabili esplicative *Unfunctional_n* (frequenza degli atti non funzionali) e *Scratching_n* (frequenza di volte in cui l’animale ha graffiato).

	Coef	Std.Error	Chisq	p-value
Intercept	-0.33679	0.48633	0.5102	0.4750
Unfunctional_n	0.12593	0.08991	3.12321	0.07719
Scratching_n	-0.16525	0.12403	0.09895	0.09895

Tabella 3.1: Tabella dei coefficienti del primo modello

Il modello finale stimato risulta essere il seguente:

$$\log (\hat{\pi}_i / (1-\hat{\pi}_i)) = - 0.33679 + 0.12593 \ x_{i1} - 0.16525 \ x_{i2} ,$$

con $x_{i1} = \text{Unfunctional_n}$ e $x_{i2} = \text{Scratching_n}$, $i=1, \dots, 26$.

Osservando i coefficienti stimati si può dedurre che la probabilità che un animale abbia il disturbo ossessivo compulsivo aumenti all'aumentare della frequenza delle azioni non funzionali. Questo risultato concorda con ciò che si conosce del DOC, ossia che un soggetto affetto da questo disturbo, al fine di alleviare l'ansia creatasi dalla ossessione, è portato a compiere una compulsione ovvero un'azione non funzionale che plachi il pensiero intrusivo dell'ossessione, pertanto i comportamenti sono ripetuti in modo frequente.

È anche vero che dal modello è emerso che la probabilità che un animale abbia il disturbo ossessivo compulsivo diminuisca all'aumentare della frequenza con la quale il soggetto tende a graffiare. Si può dunque affermare che più l'animale compie azioni non funzionali più aumenta la probabilità che esso abbia il disturbo.

Mediante la costruzione della matrice di corretta classificazione (*Tabella 3.2*) si nota che il modello ha sensibilità pari a $5/(5+8)=0.385$, specificità pari a $12/13=0.923$ e predice correttamente il 65.4% delle osservazioni.

	FALSE	TRUE
NO DOC	12	1
DOC	8	5

Tabella 3.2: Matrice di confusione per modello stimato

Come misura della bontà di adattamento del modello e della sua capacità di classificazione viene riportata di seguito la curva ROC, ovvero un grafico che mette in relazione la sensibilità e la specificità, al variare del valore della soglia considerata. Viene, inoltre, calcolata l'area sotto la curva per la quale il valore unitario rappresenta la massima capacità previsiva (assenza di errori di classificazione).

La *figura 3.1* riporta la curva ROC relativa al modello stimato con l'approssimazione di Firth; viene calcolata l'area sotto la curva ed essa risulta essere pari a 0.624. Di conseguenza la capacità predittiva del modello risulta essere bassa.

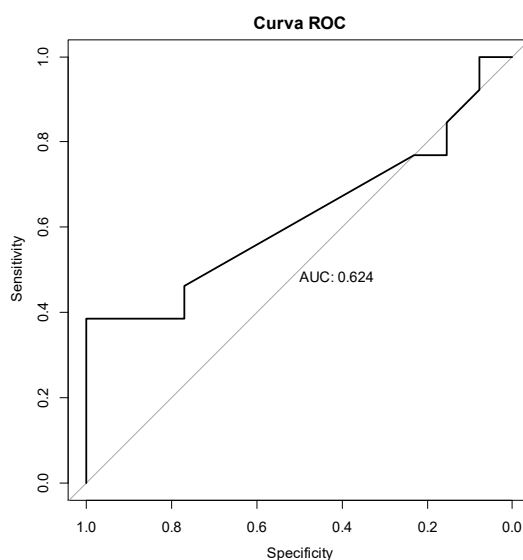


Figura 3.1: Curva ROC per l'accuratezza del modello

3.4 Metodi asintotici di ordine superiore

La maggior parte delle applicazioni d'inferenza di verosimiglianza si fondano sui risultati asintotici del primo ordine. Infatti, la funzione di verosimiglianza rappresenta la base di molti metodi statistici utilizzati per la stima, la verifica d'ipotesi e il calcolo degli intervalli di confidenza.

Se la dimensione del campione è ridotta queste approssimazioni, tuttavia, possono essere inaccurate. Perciò sono state apportate delle modifiche ad approcci consolidati, che portano a inferenze più accurate e a fornire informazioni relative a quando fare affidamento sui metodi del primo ordine. Questi sviluppi sono comunemente denominati come metodi asintotici di ordine superiore (Brazzale *et al.*, 2007).

Per l'inferenza su un parametro scalare ψ ci si serve di varie statistiche classiche di verosimiglianza, quali:

- Statistica alla Wald profilo
- Radice con segno profilo $r_p(\psi)$

Queste statistiche hanno distribuzione nulla normale standard al primo ordine. Ma poiché per i campioni di piccole dimensioni le approssimazioni al primo ordine possono dare risultati imprecisi, sono state sviluppate delle statistiche di ordine elevato.

Un importante statistica è la radice con segno profilo modificata, che è asintoticamente normale al terzo ordine, e che ha la seguente espressione

$$r_p^*(\psi) = r_p(\psi) + 1/r_p(\psi) * \log\{q(\psi)/r_p(\psi)\},$$

dove $q(\psi)$ è una opportuna funzione (Brazzale *et al.*, 2007).

Per applicare l'inferenza di ordine superiore viene utilizzato il pacchetto “*hoa*” di R che implementa tre modelli: regressione logistica, modelli lineari non normali e regressione non lineare con varianza non omogenea. Le funzioni corrispondenti da utilizzare per ogni modello sono rispettivamente: *cond*, *marg* e *nlreg*.

3.5 Modello logistico di ordine superiore

Dopo aver analizzato il modello di regressione logistica, si vuole valutare se lo stesso modello può essere validato in egual maniera al terzo ordine. Per fare ciò si fa uso del pacchetto “*hoa*” di R, più precisamente della funzione *cond(object,offset...)* alla quale viene passato il modello precedentemente stimato (*object*) e la covariata, che è presente all’interno del medesimo modello, il cui coefficiente rappresenta il parametro di interesse (*offset*). La funzione permette di calcolare gli intervalli di confidenza di primo ordine, ovvero della statistica alla Wald e della radice con segno profilo, e gli intervalli di confidenza di ordine superiore, cioè la radice con segno profilo modificata.

Questa procedura è stata eseguita per ogni covariata risultata significativa durante la regressione logistica al primo ordine, quindi sono stati confrontati gli intervalli calcolati dalla funzione al fine di verificare se il modello potesse essere validato al terzo ordine oppure se fosse necessario ridurlo ulteriormente.

Al termine delle varie verifiche sono stati ottenuti i medesimi risultati raggiunti con lo svolgimento delle analisi al primo ordine, pertanto si è potuto appurare che il modello logistico, ottenuto in precedenza, è validato anche ad ordini superiori.

CAPITOLO 4

CONCLUSIONI

In questo elaborato sono state condotte delle analisi al fine di verificare se alcuni comportamenti attuati dagli animali domestici (cani e gatti), possono essere riconducibili ad un disturbo ossessivo compulsivo (DOC). Per analizzare ciò è stato eseguito uno studio pilota su un campione di 26 unità statistiche. Dei soggetti in questione sono stati osservati i comportamenti, rilevando le frequenze dei gesti e l'intervallo di tempo trascorso tra l'azione e la ripetizione della stessa. Ogni unità è stata catalogata in un gruppo specifico (casi o controlli) in base alla diagnosi di presenza o assenza del disturbo; sono stati posti a confronto il gruppo dei casi con quello dei controlli al fine di rispondere a due quesiti:

- si osservano frequenze maggiori di attività funzionali nel gruppo dei casi rispetto a quello dei controlli?
- i comportamenti non funzionali sono attuati con più frequenza e in modo più ravvicinato nel gruppo dei casi rispetto a quello dei controlli?

In primo luogo sono state svolte analisi esplorative univariate e bivariate (Capitolo 2). Nelle analisi univariate sono state osservate, per ogni variabile, le statistiche descrittive, con le quali si è notato che: il campione è suddiviso equamente tra soggetti con DOC e senza DOC; il campione è omogeneo rispetto alla tipologia di animale; sono presenti più soggetti di sesso maschile rispetto a quelli di sesso femminile. Per quanto riguarda le variabili di tipo quantitativo sono state svolte, in aggiunta, analisi grafiche ed è stata valutata l'ipotesi di normalità mediante il test di Shapiro-Wilk; solo per le variabili *Unfunctional_intervals*, *Paw_on_face_intervals*, *Up_head_intervals*, *Scratching_intervals* e *Empty_intervals* è stata accettata l'ipotesi.

Durante lo studio si è osservata la stretta relazione tra le variabili che descrivono la frequenza con la quale avviene una specifica azione e le variabili che descrivono la media dell'intervallo di tempo intercorso tra specifici comportamenti. Queste ultime

infatti non sono sempre rilevate, ma solamente nel caso in cui un soggetto attua un'azione con frequenza maggiore di 1.

Nelle analisi bivariate è stata data maggiore importanza all'osservazione delle eventuali associazioni tra la presenza o meno del disturbo ossessivo compulsivo e tutte le altre variabili. Si osserva che il DOC è presente maggiormente nei cani rispetto ai gatti e nei maschi rispetto alle femmine. Nessuna relazione tra la patologia e tutte le altre variabili risulta essere significativa. Successivamente sono state eseguite ulteriori analisi bivariate prendendo in considerazione tutte le restanti combinazioni di coppie di variabili e riportando i soli risultati significativi in *tabella 2.8*.

Nel capitolo 3 sono stati studiati e implementati dei modelli di regressione logistica, attraverso i quali si è voluto analizzare l'effetto che le variabili esplicative hanno sulla presenza o meno del disturbo ossessivo compulsivo. In primo luogo è stato utilizzato un approccio backward al fine di generare un modello al primo ordine. Il modello in questione è di tipo logistico, essendo la variabile risposta "Phatology" dicotomica. Durante la procedura di selezione del modello si ottenevano, come stime dei coefficienti e dei relativi standard error, valori particolarmente elevati, perciò si è deciso di stimare il modello logistico tramite la correzione di Firth.

Al termine dei vari test è stato realizzato un modello stimato che comprendeva *Unfunctional_n* e *Scratching_n* come variabili concomitanti. Dal modello stimato risulta che maggiori sono le azioni non funzionali eseguite dall'animale maggiore è la probabilità che esso presenti il disturbo ossessivo compulsivo; vale invece il contrario per quanto riguarda l'azione "graffiare".

Considerando le varie analisi del dataset fornito, è possibile affermare che la modalità con cui l'animale esegue atti funzionali non varia in base alla presenza o meno del disturbo. Al contrario la frequenza degli atti non funzionali sembrerebbe essere rilevante, infatti mediante l'implementazione del modello si è notato che maggiori sono le azioni non funzionali eseguite dall'unità statistica, maggiore è la probabilità che essa presenti il disturbo ossessivo compulsivo. Per quanto

concerne i restanti comportamenti non sembrerebbe esserci alcuna relazione con il DOC.

Durante lo studio sono state riscontrate alcune problematiche causate dalla limitata quantità di dati forniti, pertanto sarebbe auspicabile un nuovo studio campionando una quantità più elevata di cani e gatti al fine di rendere maggiormente rappresentativo il campione. Con una nuova analisi si potrebbe valutare se le conclusioni fornite da questo studio siano confermate oppure se, campionando più soggetti, risultano esserci ulteriori relazioni tra la patologia ed altri tipi di comportamenti, ottenendo quindi risultati differenti.

BIBLIOGRAFIA

Tonna M., Ottoni R., Pellegrini C, Mora L, Gambolo L, Di Donna A, Parmigiani S, and Marchesi C (2022), *The motor profile of obsessive-compulsive rituals: psychopathological and evolutionary implications*, Cambridge University press.

Bruno Bara ed Bollati Boringhieri, (2005), *Il Disturbo Ossessivo Compulsivo*
Francesco Mancini Una versione successiva è stata pubblicata sul Nuovo
Manuale di Psicoterapia Cognitiva, Torino.

Gabriele Melli, (2018), *Vincere le ossessioni capire e affrontare il disturbo ossessivo-compulsivo*, Edizioni Centro Studi Erickson S.p.A.

Alessandra R. Brazzale (2005), *hoa: An R Package for Higher Order Likelihood Inference*, May, pp. 20–27

Alessandra R. Brazzale (2019) *Package ‘hoa’, Higher Order Likelihood Inference*. The function tem is based on work by Anthony Davison.

Alessandra Salvan, Nicola Sartori e Luigi Pace (2020), *Modelli lineari generalizzati*. Springer Milan, Milano.

Heinze, G. e Schemper, M. (2002). *A solution to the problem of separation in logistic regression*. Statistics in Medicine.

Georg Heinze, Meinhard Ploner, Daniela Dunkler, Harry Southworth, Lena Jiricka, (2022). *Package ‘logistf’*.

A.R. Brazzale, A. C. Davison, N. Red, (2007), *Applied Asymptotics: Case Studies in Small-Sample Statistics*,

RINGRAZIAMENTI

Al termine di questo mio primo traguardo universitario, vorrei dedicare alcune righe a tutte quelle persone che in questi tre anni mi hanno supportata e sopportata. Grazie per ciò che è stato fatto con me e per me. Il grazie più grande va alla mia famiglia: mamma, babbo e sorellona. Senza la loro presenza non sarei arrivata fino a qui. Grazie di esserci in ogni momento, nel bene e nel male. Voglio ringraziare tutti i miei amici, senza alcuna distinzione, perché a volte, anche inconsciamente, hanno reso più leggere le mie giornate con una risata e una chiacchiera. Grazie anche per tutti i momenti nei quali siete stati pazienti ed avete ascoltato con attenzione i miei lunghi monologhi. Vorrei fare un ringraziamento speciale ai miei nonni perché mi hanno cresciuta e coccolata, mia nonna dice sempre “studia tu perché io non ho potuto”, ed è quello che sto facendo sperando di rendere tutti orgogliosi.