

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA TRIENNALE IN
STATISTICA PER LE TECNOLOGIE E LE SCIENZE



RELAZIONE FINALE

**Dall'AUC al VUS: validazione di test diagnostici
per la Leishmaniosi canina**

RELATRICE: PROF.SSA LAURA VENTURA
DIPARTIMENTO DI SCIENZE STATISTICHE

LAUREANDA: GRETA BASSO
MATRICOLA N. 2036016

ANNO ACCADEMICO 2023/2024

Indice

1	Introduzione	1
1.1	Leishmaniosi canina	1
1.2	Obiettivi e dati	4
2	Analisi esplorative	7
2.1	Analisi univariate	7
2.1.1	Variabili qualitative	7
2.1.2	Variabili quantitative	9
2.2	Analisi bivariate	12
2.2.1	Variabili qualitative rispetto a qualitative	12
2.2.2	Variabili quantitative rispetto a qualitative	13
2.2.3	Variabili quantitative rispetto a quantitative	21
2.3	Considerazioni finali	23
3	Analisi delle curve ROC	25
3.1	Classificazione tra sani e malati	25
3.1.1	Test diagnostici analizzati	26
3.1.2	Considerazioni finali	30
3.2	Classificazione per gravità della malattia	31
3.2.1	Test diagnostici analizzati	34
3.2.2	Considerazioni finali	43
4	Regressione logistica	45
4.1	Stima del modello	46
4.2	Metodi alternativi per la selezione delle variabili	49
4.2.1	Selezione stepwise basata sull'AIC	49
4.2.2	Metodo Lasso	50
4.3	Considerazioni finali	53
5	Risultati e conclusioni	55
	Appendice	59
	Bibliografia	61

Capitolo 1

Introduzione

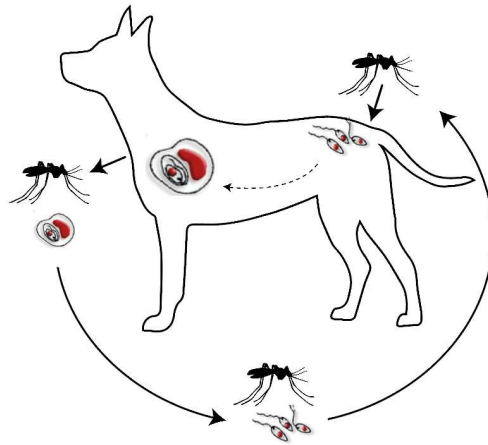
1.1 Leishmaniosi canina

La leishmaniosi è una malattia provocata da un microscopico parassita, un protozoo dal nome *Leishmania infantum*, che colpisce principalmente il cane e raramente anche l'uomo. È molto diffusa in tutto il Centro e il Sud Italia, ma dagli anni '90 ha cominciato a espandersi anche al Nord.

La malattia ha andamento cronico e, se non curata, porta spesso alla morte del cane. Si trasmette attraverso la puntura dei flebotomi, insetti color sabbia conosciuti anche con il nome di “pappataci” (Figura 1.1). Esistono circa 600 specie di flebotomi, ma solo una sessantina di esse è in grado di trasmettere leishmania.

In Italia i vettori principali sono il *Phlebotomus perniciosus* e il *P. perfiliewi*, entrambi ad attività crepuscolare e notturna, che si estende dalla tarda primavera al tardo autunno. Solo la femmina si nutre di sangue quando deve portare a maturazione le uova.

Sono state segnalate altre rare vie di trasmissione: attraverso donazioni di sangue e organi, per via verticale (dalla madre al feto), attraverso l'accoppiamento. La malattia, tuttavia, si mantiene nell'ambiente solo se sono presenti i flebotomi vettori.

FIGURA 1.1: Ciclo della *Leishmania*.

Sintomi e cure

La malattia può rimanere inapparente per molto tempo (da 3 mesi a 7 anni). I sintomi possono comparire in tempi diversi e associarsi nella più varia combinazione; l'andamento è prevalentemente di tipo sub-acute o cronico e solo raramente vengono segnalate forme acute con insorgenza di febbre remittente o intermittente.

I primi sintomi possono essere l'appetito capriccioso, il dimagrimento e la stanchezza; poi possono insorgere alterazioni della pelle non pruriginose e spesso simmetriche e alterazioni del mantello (dermatite con forfora, perdita di pelo, ulcere, pustole e ipercheratosi nasale e/o digitale). Esse sono tra i segni più importanti della malattia.

Seguono ingrossamento dei linfonodi, problemi agli occhi (congiuntivite, uveite, etc.), problemi renali, alle articolazioni e ad altri organi interni, questi ultimi in genere rilevabili dal veterinario tramite approfondimenti diagnostici.

Per la Leishmaniosi canina esiste una cura, anche se lunga e impegnativa; le probabilità di successo sono tanto maggiori quanto più il cane viene curato precocemente. Inoltre, anche se i sintomi scompaiono, spesso ciò non è associato all'eliminazione del parassita, che continua ad albergare nei macrofagi del cane infettando i vettori ed esponendo l'animale a eventuali recidive.

La terapia per la leishmaniosi si basa sull'utilizzo di terapia antimoniale prolungata con cicli anche di 6-8 settimane senza interruzione. Uno dei principi attivi più utilizzati è l'antimoniato di *N-metilglucamina*, somministrato per via sottocutanea, due volte al giorno allo scopo di mantenere costante la concentrazione del farmaco in circolo, associato ad allopurinolo per via orale, la cui somministrazione può protrarsi per 4-6 mesi.

Test diagnostici utilizzati

Il paziente con segni clinici e/o alterazioni di laboratorio compatibili con leishmaniosi deve essere considerato malato se il parassita è evidenziabile citologicamente nelle lesioni e/o il titolo anticorpale è 4 volte superiore al limite di positività del laboratorio di riferimento.

Se il titolo anticorpale è medio-basso e l'esame citologico negativo, vanno eseguiti esami istologici/immunoistochimici o PCR su midollo e/o linfonodo. Se solamente quest'ultimo esame risulta positivo il soggetto deve essere considerato infetto e, anche malato, solo nel caso emerga una chiara correlazione tra la positività e la lesione. Se anche la PCR risulta negativa il soggetto va considerato esposto e monitorato sierologicamente.

La proteinuria deve essere quantificata mediante il rapporto proteine/creatinina urinario. Dato che il grado di proteinuria si associa alla possibilità di sviluppare insufficienza renale, è importante non solo instaurare una terapia anti-*Leishmania* ma anche contrastare la proteinuria. A tale scopo possono essere utilizzati diversi farmaci, tra i quali gli ACE-inibitori. Infine, è necessario modificare la dieta, soprattutto nei soggetti con contemporanea insufficienza renale.

I metodi diagnostici utilizzati in maggior misura sono descritti di seguito.

- **Esame citologico:** viene effettuato un campionamento a livello di organi lesionati e si osserva se sono presenti alterazioni citologiche compatibili con la leishmaniosi.
- **Esame istologico:** il parassita può essere evidenziato in sezioni allestite da lesioni colorate routinariamente con *ematossilina-eosina*.
- ***Polymerase chain reaction (PCR)*:** questa tecnica permette di amplificare sequenze specifiche del genoma di *Leishmania*. È in grado di identificare piccolissime quantità di DNA dei protozoi presenti nel materiale biologico esaminato. Le tre tecniche più utilizzate sono la PCR *convenzionale*, la PCR *nested* e la PCR *real-time*.
- **Test ELISA (*Enzyme-Linked ImmunoSorbent Assay*):** è una tecnica immunoenzimatica che permette di rilevare e quantificare gli anticorpi specifici (prevalentemente di tipo G) nei confronti di *Leishmania infantum*. È un test specifico e ha sensibilità medio-alta (70-100%); inoltre, permette di quantificare gli anticorpi specifici.

- **Test IFAT (*Indirect immunofluorescent antibody test*):** viene eseguito ponendo il siero in esame su vetrini su cui sono presenti promastigoti di *Leishmania*. Gli anticorpi eventualmente presenti si legano ai promastigoti e la positività viene evidenziata utilizzando anti-anticorpi fluorescenti. La sensibilità e specificità del test sono prossime al 100%.

1.2 Obiettivi e dati

Le seguenti analisi fanno riferimento a uno studio svolto presso la clinica veterinaria San Marco, Veggiano (PD).

Il dataset contiene informazioni su un campione di 72 cani che sono stati sottoposti a un test sierologico con lo scopo di stabilire se siano affetti o meno da Leishmaniosi canina.

Obiettivo dello studio è valutare la performance e l'accuratezza di una serie di test diagnostici, al fine di individuare quelli che discriminano meglio tra sani e malati, e di modellare lo status del paziente in funzione delle altre variabili.

Le variabili prese in esame sono:

- **GROUP:** variabile qualitativa dicotomica che assume valore 0 per gli individui sani e valore 1 per gli individui malati;
- **SEROLOGY:** variabile quantitativa continua che rappresenta l'esito del test sierologico di riferimento;
- **Leishvet:** variabile qualitativa a cinque livelli che rappresenta una classificazione degli stadi corrispondenti alla gravità di sviluppo della malattia (H, IIa, IIb, III, IV);
- **Leishvet_2:** variabile qualitativa a tre livelli che rappresenta una suddivisione meno dettagliata di *Leishvet*, raggruppando i malati in due gruppi (H, IIa, IIb+III+IV);
- **IRIS:** variabile qualitativa a cinque livelli che rappresenta un'altra classificazione degli stadi di gravità di sviluppo della malattia (H, I, II, III, IV);
- **IRIS_2:** variabile qualitativa a tre livelli che rappresenta una suddivisione meno dettagliata di *IRIS*, raggruppando i malati in due gruppi (H, I, II+III+IV);

- **MIX.BREED:** variabile qualitativa dicotomica che assume valore *YES* se il cane è di razza mista e valore *NO* se il cane è di razza;
- **SEX:** variabile qualitativa dicotomica che assume valore *F* per le femmine e *M* per i maschi;
- **SEXUAL.STATUS:** variabile qualitativa dicotomica che assume valore *C* se il cane è castrato e valore *I* se è intero;
- **AGE:** variabile quantitativa continua che indica l'età in mesi;
- **BW:** variabile quantitativa continua che indica il peso in kg;
- **HR:** variabile quantitativa continua che indica la frequenza cardiaca in bpm;
- **RR:** variabile quantitativa continua che indica la frequenza respiratoria, ovvero il numero di respiri al minuto;
- **SBP:** variabile quantitativa continua che indica la pressione sistolica in mmHg;
- **MED_U_POD:** variabile quantitativa continua che indica la quantità di podocina espressa in mg/dl;
- **CREA_U:** variabile quantitativa continua che indica la quantità di creatinina presente nelle urine, espressa in mg/dl;
- **MED_U_NEF:** variabile quantitativa continua che indica la quantità di nefrina espressa in mg/dl;
- **UPC:** variabile quantitativa continua che rappresenta il rapporto tra proteine urinarie e creatinina urinaria;
- **USG:** variabile quantitativa continua che rappresenta il peso specifico, espresso in g/ml di urina;
- **BUN:** variabile quantitativa continua che indica la quantità di urea nel sangue espressa in mg/dl;
- **CREA_S:** variabile quantitativa continua che indica la quantità di creatinina presente nel sangue, espressa in mg/dl;
- **SDMA:** variabile quantitativa continua che indica la quantità di dimetilarginina simmetrica, un marker della funzionalità renale, espressa in g/dl;

- **uNuC**: variabile quantitativa continua che rappresenta il rapporto tra nefrina e creatinina urinaria, moltiplicato per 100.000;
- **uPuC**: variabile quantitativa continua che rappresenta il rapporto tra podocina e creatinina urinaria, moltiplicato per 100.000;

Le analisi presentate in questo studio sono state eseguite con il software R nella versione 4.1.3. Il livello di significatività è fissato al 5%.

Capitolo 2

Analisi esplorative

In questo capitolo si effettua una prima analisi esplorativa di tutte le variabili osservate per le 72 unità statistiche. Ai metodi grafici, quali boxplot, si affiancano opportuni test. Per approfondimenti si rimanda a Ventura e Racugno (2017).

2.1 Analisi univariate

2.1.1 Variabili qualitative

Nella Tabella 2.1 sono riportate le frequenze assolute e percentuali delle variabili qualitative dicotomiche.

TABELLA 2.1: Frequenze assolute e percentuali delle variabili qualitative dicotomiche.

Variabile	Categorie	N	%
GROUP	Sani	35	48.6
	Malati	37	51.4
MIX.BREED	YES	22	30.6
	NO	50	69.4
SEX	Femmine	32	44.4
	Maschi	40	55.6
SEXUAL.STATUS	C	36	50.0
	I	36	50.0

Il campione è composto da 35 soggetti sani (48.6%) e da 37 soggetti malati (51.4%). Quasi il 70% dei cani è di razza e il restante 30% è di razza mista. Per quanto riguarda

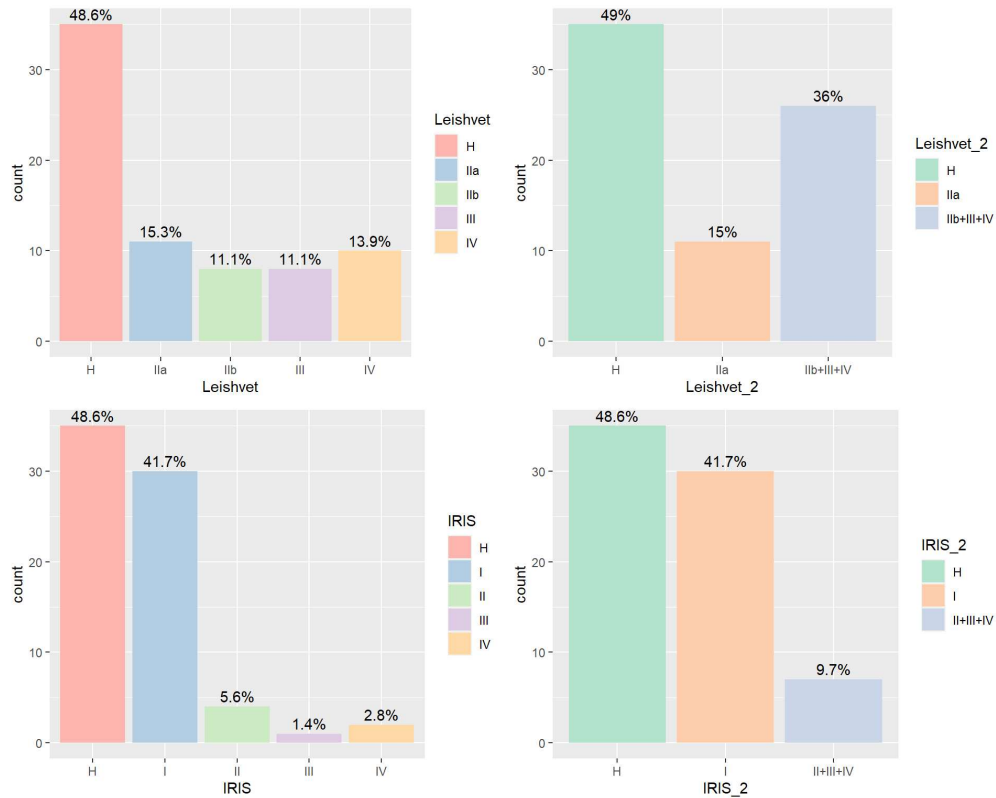


FIGURA 2.1: Diagrammi a barre delle variabili qualitative relative alle classificazioni della gravità della malattia, con relative frequenze percentuali.

il sesso, le femmine sono il 44.4% e i maschi il 55.6%; inoltre, esattamente la metà del campione è castrata.

Nella Figura 2.1 sono rappresentati i diagrammi a barre delle variabili qualitative che rappresentano le diverse classificazioni degli stadi di gravità della malattia. Per tutte e quattro le variabili il gruppo *H* contiene il 48.6% dei soggetti; esso corrisponde, infatti, al gruppo dei sani individuato dalla variabile *GROUP*.

Per la variabile *Leishvet* i malati sono suddivisi tra il gruppo *Ila* (15.3%), il gruppo *Ilb* (11.1%), il gruppo *III* (11.1%) e il gruppo *IV* (13.9%); questi ultimi tre sono poi raggruppati insieme per la variabile *Leishvet_2* (36%).

Per la variabile *IRIS* i malati sono suddivisi tra il gruppo *I* (41.7%), il gruppo *II* (5.8%), il gruppo *III* (1.4%) e il gruppo *IV* (2.8%); questi ultimi tre sono poi raggruppati insieme per la variabile *IRIS_2* (9.7%). Poiché per la variabile *IRIS* si ottengono alcuni gruppi con numerosità campionaria pari a uno o due, per le successive analisi si considererà solo il suo raggruppamento *IRIS_2*.

2.1.2 Variabili quantitative

Nella Tabella 2.2 sono riportate le statistiche descrittive di sintesi per le variabili continue e in Figura 2.2 si osservano i relativi boxplot.

TABELLA 2.2: Statistiche descrittive per le variabili quantitative.

Variabile	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.Dev.	IQR
AGE	10.0	46.8	67.8	71.3	97.5	153.6	33.091	50.7
BW	4.8	14.9	21.4	23.4	30.7	60.0	12.174	15.8
HR	60.0	92.0	100.0	110.0	120.0	190.0	26.477	20.0
RR	16.0	24.0	30.0	36.1	40.0	120.0	20.718	16.0
SBP	120.0	134.2	145.0	150.1	159.2	230.0	23.368	25.0
SEROLOGY	1.3	2.3	13.1	20.4	39.1	56.3	18.970	36.8
MED_U_POD	3.8	8.2	11.9	13.8	16.8	50.0	8.184	8.6
CREA_U	28.0	130.5	179.5	199.1	256.0	545.0	102.598	125.5
MED_U_NEF	1.0	2.2	2.9	3.4	3.9	9.0	1.630	1.7
UPC	0.1	0.2	0.3	2.0	0.8	27.7	4.618	0.6
USG	1011	1028	1038	1038	1046	1070	13.839	18.0
BUN	11.0	24.0	33.0	42.8	39.0	331.0	46.580	15.0
CREA_S	0.3	0.8	1.0	1.2	1.2	5.8	0.842	0.4
SDMA	2.4	8.3	12.0	13.0	17.0	31.0	5.692	8.7
uNuC	0.0	0.1	0.2	0.2	0.3	0.7	0.134	0.2
uPuC	0.2	0.5	0.6	0.8	0.9	2.9	0.432	0.4

Al fine di valutare l'ipotesi di normalità per tali variabili si è svolto il test di Shapiro-Wilk, i cui risultati sono riportati in Tabella 2.3, e si sono osservati i diagrammi quantile-quantile. Risulta che la normalità è verificata per le sole variabili relative all'età e al peso specifico.

L'età media dei soggetti è pari a 71.3 mesi (deviazione standard 33.091 mesi), che corrisponde a circa 6 anni. Dal boxplot non si osservano valori anomali.

Il peso mediano è di 21.4 kg (IQR 15.8 kg). Nel boxplot è presente un outlier e si osserva una lieve forma di asimmetria a destra.

La frequenza cardiaca mediana è di 100 bpm (IQR 20.0 bpm). Nel boxplot sono presenti due valori anomali e vi è una lieve forma di asimmetria a destra.

Per la variabile *RR* vi è una mediana di 30.0 respiri al minuto (IQR 16.0) e si osserva una forma di asimmetria a destra con numerosi valori anomali.

La pressione sistolica mediana (*SBP*) è di 145.0 mmHg (IQR 25.0 mmHg) e sono presenti tre valori anomali e una forma di asimmetria a destra.

La concentrazione mediana di anticorpi rilevata da *SEROLOGY* è 13.1 (IQR 36.8).

TABELLA 2.3: Statistica osservata e p-value del test di Shapiro-Wilk per le variabili quantitative.

Variabile	W^{oss}	p-value
AGE	0.98	0.478
BW	0.96	0.025
HR	0.93	0.001
RR	0.60	< 0.001
SBP	0.88	< 0.001
SEROLOGY	0.83	< 0.001
MED_U_POD	0.81	< 0.001
CREA_U	0.96	0.019
MED_U_NEF	0.88	< 0.001
UPC	0.46	< 0.001
USG	0.98	0.553
BUN	0.48	< 0.001
CREA_S	0.55	< 0.001
SDMA	0.95	0.004
uNuC	0.87	< 0.001
uPuC	0.80	< 0.001

La quantità mediana di podocina (*MED_U_POD*) è di 11.9 mg/dl (IQR 8.6 mg/dl) e si osserva una forma di asimmetria a destra con tre valori anomali.

La quantità mediana di creatinina nelle urine (*CREA_U*) è di 179.5 mg/dl (IQR 125.5 mg/dl) e si osserva una forma di asimmetria a destra con due valori anomali.

La quantità mediana di nefrina (*MED_U_NEF*) è di 2.9 mg/dl (IQR 1.7 mg/dl) e si osserva una forma di asimmetria a destra con tre valori anomali.

La mediana del rapporto tra proteine urinarie e creatinina urinaria (*UPC*) è pari a 0.3 (IQR 0.6). La distribuzione è fortemente asimmetrica a destra con un elevato numero di valori anomali.

Il peso specifico medio (*USG*) è di 1038 g/ml (deviazione standard 13.839 g/ml) e non sono presenti valori anomali.

La quantità mediana di urea (*BUN*) è 33.0 mg/dl (IQR 15.0 mg/dl). La distribuzione è fortemente asimmetrica a destra con un elevato numero di valori anomali.

La quantità mediana di creatinina nel sangue (*CREA_S*) è di 1.0 mg/dl (IQR 0.4 mg/dl). La distribuzione è fortemente asimmetrica a destra con un elevato numero di valori anomali.

La quantità mediana di dimetilargina simmetrica (*SDMA*) è 12.0 g/dl (IQR 8.7 g/dl) e si osserva un valore anomalo.

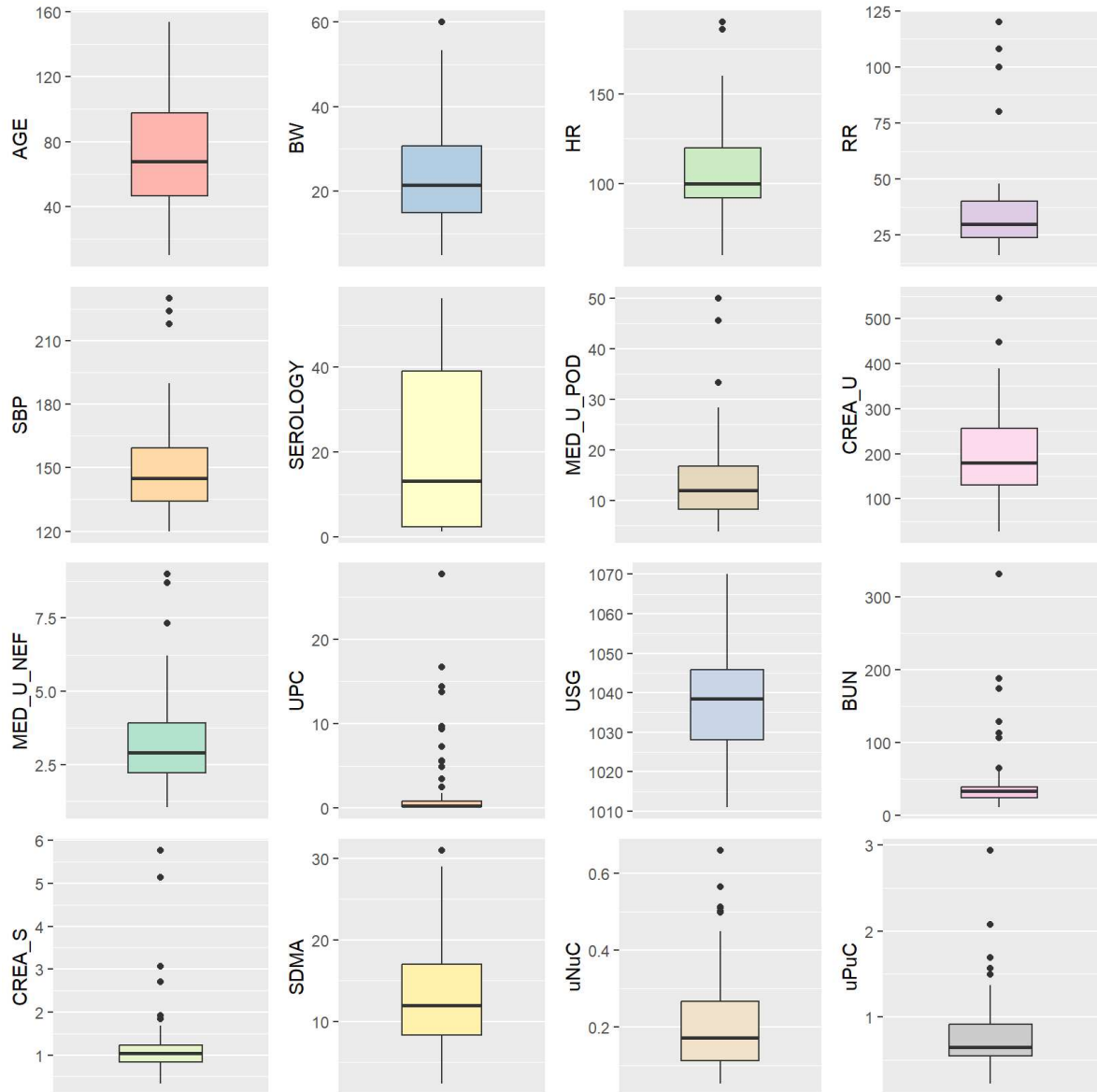


FIGURA 2.2: Boxplot delle variabili quantitative.

La mediana del rapporto tra nefrina e creatinina urinaria ($uNuC$) è 0.2 (IQR 0.2) e si osservano numerosi valori anomali.

La mediana del rapporto tra podocina e creatinina urinaria ($uPuC$) è 0.6 (IQR 0.4) e si osservano numerosi valori anomali.

2.2 Analisi bivariate

Di seguito si riportano le analisi esplorative bivariate. In particolare vengono inserite solamente le relazioni tra variabili che sono risultate significative dai test di verifica di ipotesi, quali il χ^2 di Pearson, il test t a due campioni e i test non parametrici di Mann-Whitney e Kruskal-Wallis basati sui ranghi. Si esegue, inoltre, un'analisi delle correlazioni tra le variabili quantitative in modo da determinare se esiste un legame tra tali variabili e quantificarne la relazione.

2.2.1 Variabili qualitative rispetto a qualitative

Per verificare se vi è associazione tra le coppie di variabili qualitative si è applicato il test χ^2 di Pearson, con eventuale correzione di Yates per le tabelle 2×2 e con eventuale p-value simulato in presenza di frequenze attese inferiori a 5.

Gli unici confronti che non sono stati presi in considerazione sono quelli tra la variabile *GROUP* e le variabili *Leishvet*, *Leishvet_2*, *IRIS* e *IRIS_2* perché sono tutte diverse classificazioni della malattia.

Dall'applicazione del test risulta che la variabile di interesse *GROUP* non è associata ad alcuna delle variabili qualitative che descrivono le caratteristiche individuali, poiché i relativi p-value sono tutti maggiori di 0.05.

Per quanto riguarda gli altri confronti, l'unico che risulta significativo è quello tra il sesso e lo stato sessuale del soggetto. Si è ottenuto un valore osservato della statistica test pari a $\chi^2 = 9.51$, con p-value = 0.002. Si conclude, quindi, che vi è associazione tra queste due variabili. Dalla Tabella 2.4 si può notare che le femmine sono per oltre il 70% castrate, mentre per i maschi questa percentuale si riduce a circa il 30%.

TABELLA 2.4: Tabella di contingenza della variabile sesso rispetto allo stato sessuale.

	C	I	Totale
F	23	9	32
M	13	27	40
Totale	36	36	72

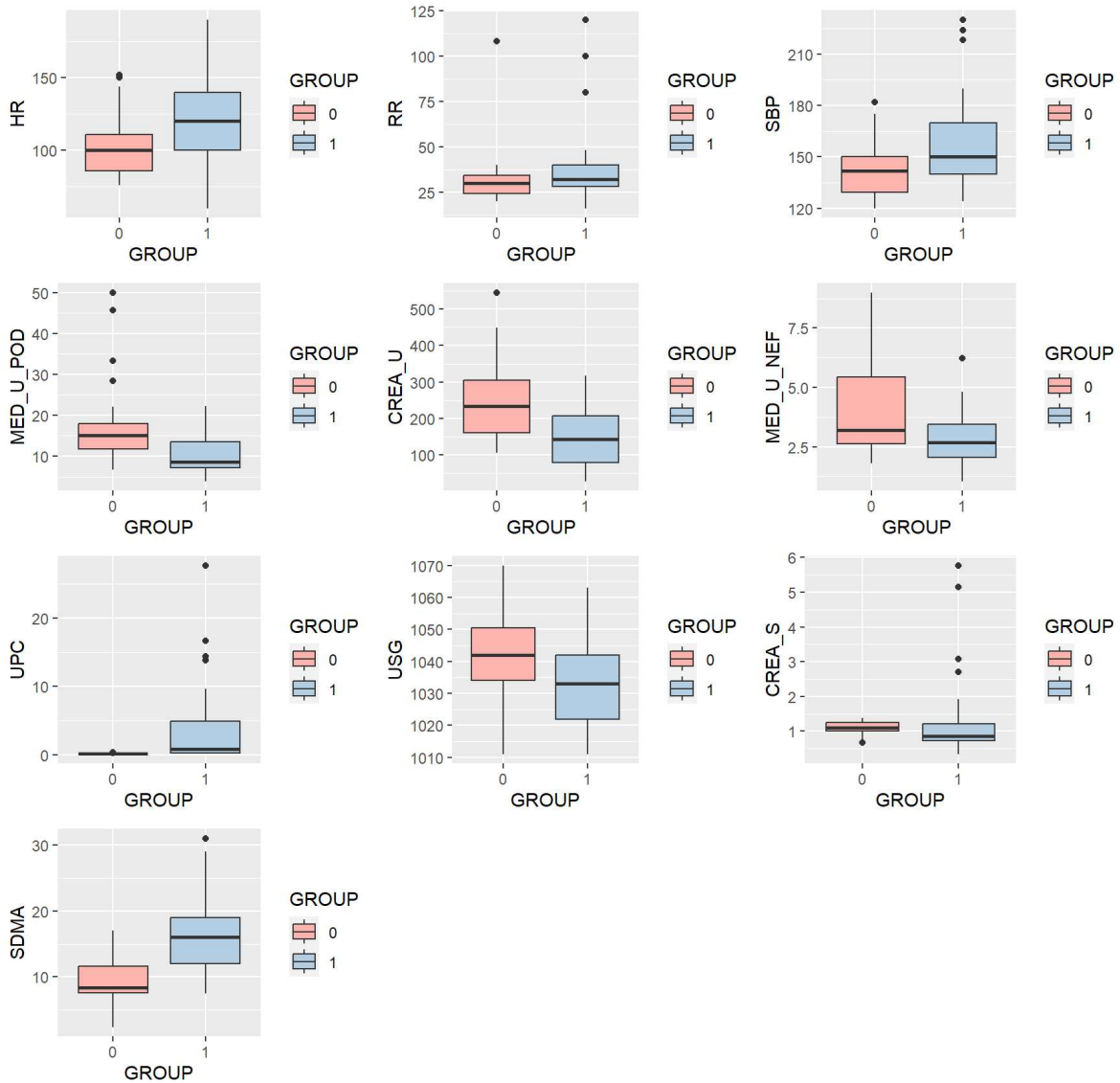


FIGURA 2.3: Boxplot delle variabili quantitative rispetto a *GROUP*.

2.2.2 Variabili quantitative rispetto a qualitative

Per verificare se vi sono differenze significative tra i gruppi identificati dalle modalità delle variabili qualitative si utilizzano i test a due campioni per i confronti tra due gruppi e l'analisi della varianza per quelli tra più di due gruppi. Si riportano soltanto i confronti che sono risultati significativi.

Confronti per la variabile *GROUP*

Si esaminano in primo luogo i confronti tra il gruppo dei sani e il gruppo dei malati. Nella Figura 2.3 sono riportati i boxplot delle variabili quantitative differenziati per i due gruppi.

Si valuta la normalità delle variabili quantitative all'interno del gruppo dei sani e di quello dei malati attraverso il test di Shapiro-Wilk, i cui risultati sono riportati in Tabella 2.5.

TABELLA 2.5: Statistica osservata e p-value del test di Shapiro-Wilk per ciascun valore di *GROUP*.

Variabile	GROUP = 0		GROUP = 1	
	W^{oss}	p-value	W^{oss}	p-value
HR	0.89	0.002	0.96	0.283
RR	0.55	< 0.001	0.65	< 0.001
SBP	0.95	0.106	0.88	0.001
MED_U_POD	0.74	< 0.001	0.89	0.002
CREA_U	0.94	0.042	0.96	0.222
MED_U_NEF	0.88	0.001	0.95	0.127
UPC	0.80	< 0.001	0.64	< 0.001
USG	0.99	0.995	0.97	0.447
CREA_S	0.97	0.409	0.61	< 0.001
SDMA	0.92	0.016	0.96	0.156

L'ipotesi di normalità si rifiuta per almeno un gruppo per ogni variabile, ad esclusione di *USG*; di conseguenza si è utilizzato il test t a due campioni per quest'ultima e il test non parametrico di Mann-Whitney per tutte le restanti variabili.

Per poter applicare il test t parametrico per la variabile *USG* è necessario prima verificare l'ipotesi di omoschedasticità nei due gruppi; il test F fornisce un valore osservato della statistica test pari a $F^{oss} = 0.97$, con relativo p-value = 0.931, che porta ad accettare l'omoschedasticità. Effettuando il test t a due campioni si ottiene il valore $t^{oss} = 2.33$ con p-value = 0.023; vi è, quindi, una differenza significativa tra le medie di *USG* nei due gruppi. In particolare, il peso specifico tra i sani è in media maggiore di quello tra i malati (Tabella 2.6).

TABELLA 2.6: Media e deviazione standard di *USG* per sani e malati.

Variabile	GROUP = 0		GROUP = 1	
	Mean	Std. Dev.	Mean	Std. Dev.
USG	1041	13.322	1034	13.527

Si è applicato poi il test di Mann-Whitney per le restanti variabili, i cui risultati sono riportati in Tabella 2.7, insieme alle statistiche di sintesi.

TABELLA 2.7: Mediana e scarto interquartilico al variare di *GROUP* e risultati del test di Mann-Whitney.

Variabile	GROUP = 0		GROUP = 1		t^{MW}	p-value
	Median	IQR	Median	IQR		
HR	100	25	120	40	402.5	0.006
RR	30	10	32	12	460.0	0.033
SBP	142	20.5	150	30	434.0	0.016
MED_U_POD	15.06	6.13	8.63	6.47	984.5	< 0.001
CREA_U	233	143.5	144	128	964.0	< 0.001
MED_U_NEF	3.2	2.81	2.7	1.37	889.5	0.007
UPC	0.2	0.05	0.8	4.6	102.0	< 0.001
CREA_S	1.11	0.23	0.86	0.5	882.5	0.008
SDMA	8.4	4.1	16.0	7.0	198.0	< 0.001

Si può quindi affermare che vi è differenza significativa tra le mediane di sani e malati per le variabili riportate in tabella. In particolare, la mediana è minore nel gruppo dei sani rispetto a quello dei malati per le variabili riferite alla frequenza cardiaca (*HR*), alla frequenza respiratoria (*RR*), alla pressione sistolica (*SBP*), al rapporto tra proteine urinarie e creatinina urinaria (*UPC*) e alla dimetilarginina simmetrica (*SDMA*).

Al contrario, la mediana è maggiore nel gruppo dei sani rispetto a quello dei malati per le variabili riferite alla podocina (*MED_U_POD*), alla creatinina nelle urine (*CREA_U*), alla nefrina (*MED_U_NEF*) e alla creatinina nel sangue (*CREA_S*).

Confronti per le altre variabili dicotomiche

Dopo aver svolto i confronti per la variabile *GROUP*, si è ripetuta l'analisi per le altre tre variabili dicotomiche (razza, sesso e stato sessuale).

Per *MIX.BREED* si riportano i boxplot dei due confronti risultati significativi in Figura 2.4.

Per la variabile *BW* l'ipotesi di normalità, valutata tramite il test di Shapiro-Wilk, è verificata sia per il gruppo dei cani di razza ($W^{oss} = 0.98$, p-value = 0.369), che per quello dei cani di razza mista ($W^{oss} = 0.92$, p-value = 0.072). Inoltre, si accetta l'ipotesi di omoschedasticità nei due gruppi, valutata tramite il test F ($F^{oss} = 3.51$, p-value = 0.226).

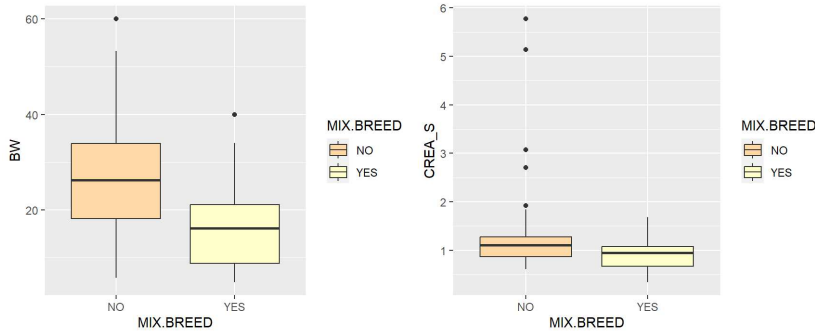


FIGURA 2.4: Boxplot delle variabili quantitative rispetto a *MIX.BREED*.

Si applica, quindi, il test t a due campioni che porta al rifiuto dell'ipotesi di uguaglianza delle medie ($t^{oss} = 0.92$, $p\text{-value} < 0.001$). In particolare, il peso dei cani di razza è in media maggiore di quello dei cani di razza mista (Tabella 2.8).

TABELLA 2.8: Media e deviazione standard di *BW* per cani di razza e di razza mista.

	MIX.BREED = NO		MIX.BREED = YES	
Variabile	Mean	Std. Dev.	Mean	Std. Dev.
BW	26.51	12.019	16.35	9.432

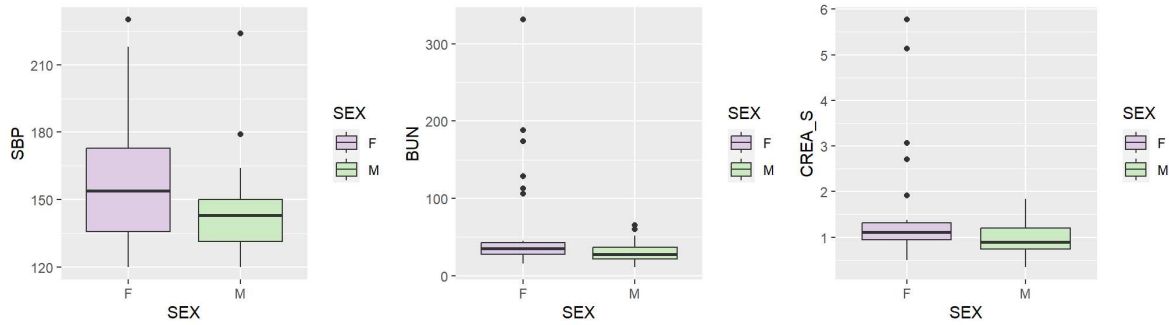
Per la variabile *CREA_S* l'ipotesi di normalità, valutata tramite il test di Shapiro-Wilk, è rifiutata per il gruppo dei cani di razza ($W^{oss} = 0.54$, $p\text{-value} < 0.001$) ed è accettata per quello dei cani di razza mista ($W^{oss} = 0.98$, $p\text{-value} = 0.905$).

Si applica, quindi, il test di Mann-Whitney che porta al rifiuto dell'ipotesi di uguaglianza delle mediane ($W^{oss} = 722.5$, $p\text{-value} = 0.035$). In particolare, la quantità di creatinina nel sangue nei cani di razza è in mediana maggiore di quella nei cani di razza mista (Tabella 2.9).

TABELLA 2.9: Mediana e scarto interquartile di *CREA_S* per cani di razza e di razza mista.

	MIX.BREED = NO		MIX.BREED = YES	
Variabile	Median	IQR	Median	IQR
CREA_S	1.10	0.402	0.95	0.412

Per *SEX* si riportano i boxplot dei tre confronti risultati significativi in Figura 2.5.

FIGURA 2.5: Boxplot delle variabili quantitative rispetto a *SEX*.

L'ipotesi di normalità, valutata tramite il test di Shapiro-Wilk, è rifiutata per almeno un gruppo per tutte e tre le variabili; di conseguenza si è applicato il test non parametrico di Mann-Whitney, il quale porta al rifiuto dell'ipotesi di uguaglianza delle mediane nei tre casi (Tabella 2.10). In particolare, la mediana è maggiore nel gruppo delle femmine rispetto a quello dei maschi per le variabili riferite alla pressione sistolica (*SBP*), alla quantità di urea (*BUN*) e alla quantità di creatinina nel sangue (*CREA_S*).

TABELLA 2.10: Mediana e scarto interquartile al variare di *SEX* e risultati del test di Mann-Whitney.

Variabile	SEX = F		SEX = M		t^{MW}	p-value
	Median	IQR	Median	IQR		
SBP	154	37	143	18.5	831.5	0.030
BUN	35.0	15.0	27.5	15.2	851.5	0.017
CREA_S	1.11	0.37	0.89	0.46	862.5	0.012

Nessun confronto è risultato significativo per la variabile riferita allo stato sessuale.

Confronti per le variabili con più di due modalità

Poiché le variabili *Leishvet*, *Leishvet_2*, *IRIS* e *IRIS_2* sono classifiche ordinali della gravità della malattia, nel seguito dell'analisi si utilizzano i test non parametrici basati sui ranghi. In particolare, viene svolta l'ANOVA non parametrica a una via applicando il test di Kruskal-Wallis.

Come spiegato nella Sezione 2.1.1, non vengono svolti i test per la variabile *IRIS*, perché presenta gruppi con numerosità campionaria pari a uno o due. Nel seguito si riportano soltanto i confronti che sono risultati significativi.

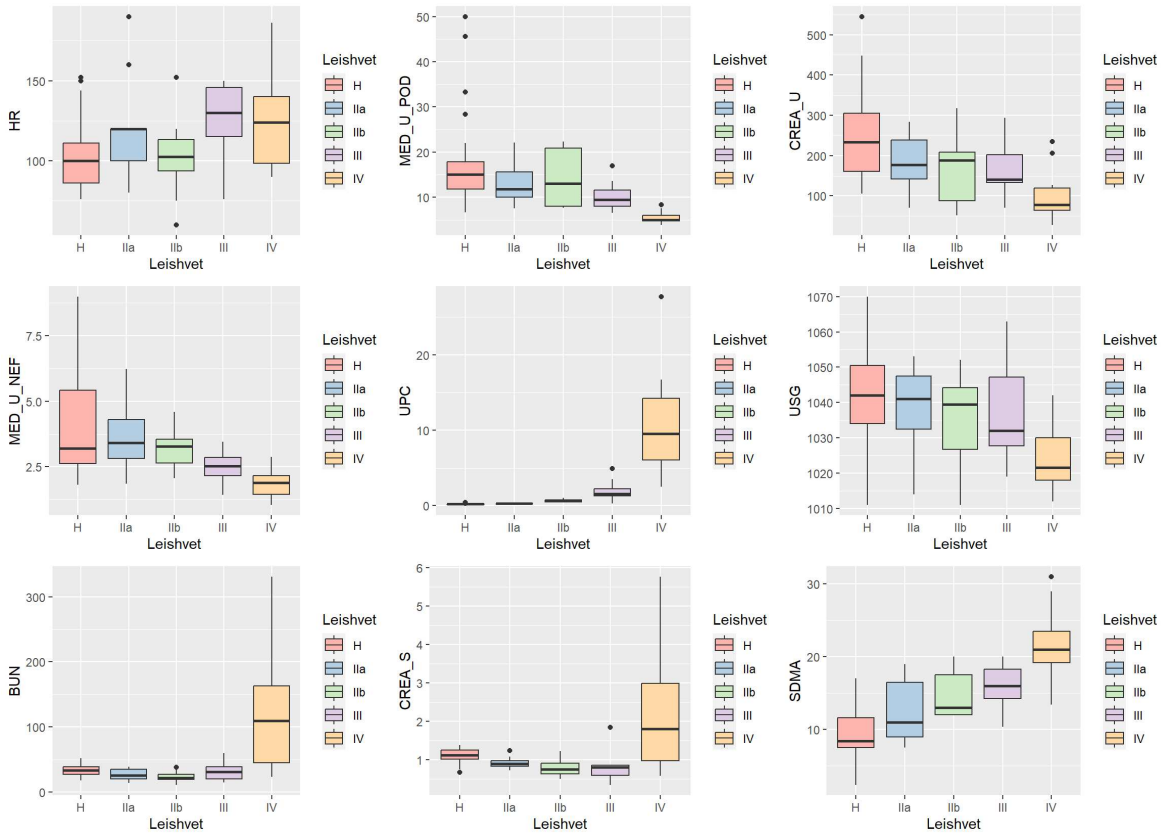


FIGURA 2.6: Boxplot delle variabili quantitative rispetto a *Leishvet*.

In Figura 2.6 sono rappresentati i boxplot delle variabili quantitative rispetto a *Leishvet* e in Figura 2.7 quelli rispetto a *Leishvet_2*.

Nella Tabella 2.11 si osservano i p-value del test di Kruskal-Wallis per la variabile *Leishvet* e per il suo raggruppamento *Leishvet_2*.

TABELLA 2.11: Statistiche osservate e p-value del test di Kruskal-Wallis per le variabili quantitative differenziate per *Leishvet* e *Leishvet_2*.

Variabile	Leishvet		Leishvet_2	
	KW_{oss}	p-value	KW_{oss}	p-value
HR	10.22	0.037	7.72	0.021
MED_U_POD	29.47	< 0.001	18.20	< 0.001
CREA_U	18.42	0.001	14.91	< 0.001
MED_U_NEF	22.25	< 0.001	14.90	< 0.001
UPC	55.64	< 0.001	52.82	< 0.001
USG	10.89	0.028	6.58	0.037
BUN	20.01	< 0.001	3.08	0.215
CREA_S	21.94	< 0.001	7.42	0.025
SDMA	35.05	< 0.001	31.75	< 0.001

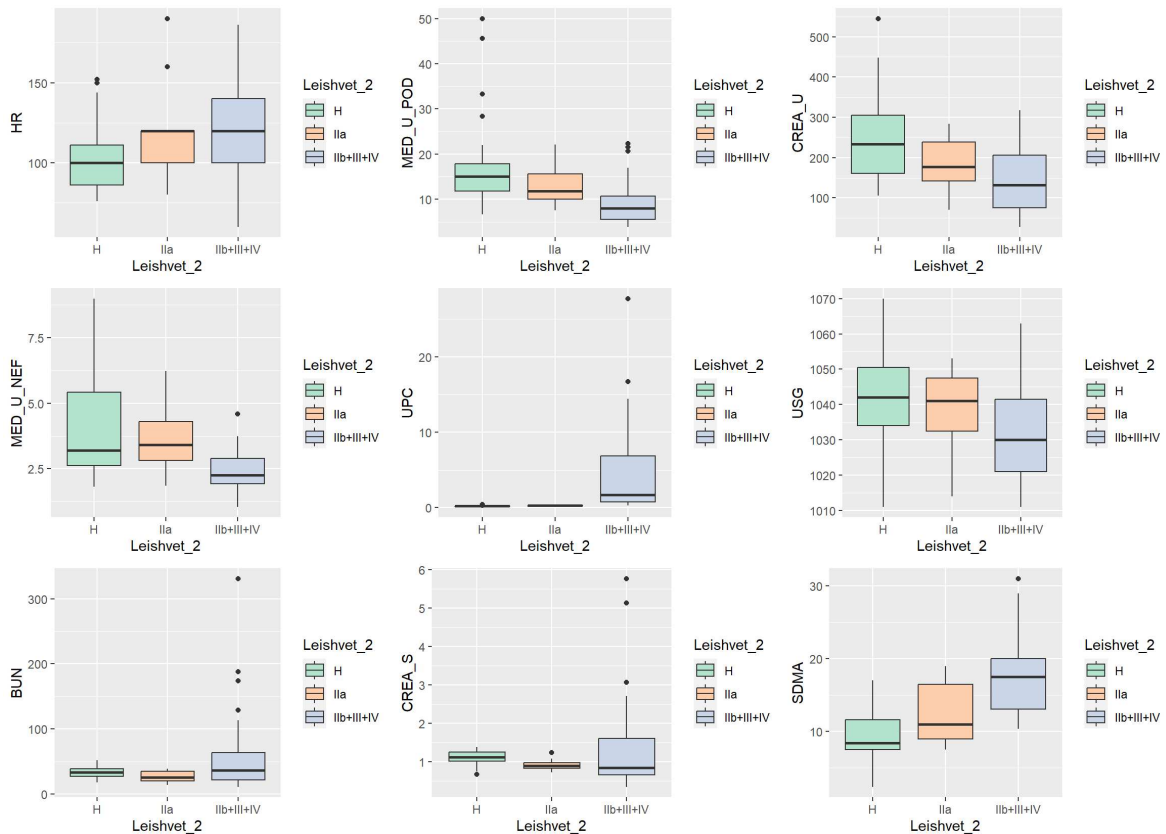


FIGURA 2.7: Boxplot delle variabili quantitative rispetto a *Leishvet_2*.

Data la significatività del test per queste variabili, si procede con un'analisi post-hoc, con correzione di Holm, per stabilire quali confronti hanno portato al rifiuto dell'ipotesi di uguaglianza tra mediane.

Per quanto riguarda la variabile *HR*, nonostante l'ANOVA sia risultata significativa per la variabile *Leishvet*, l'analisi post-hoc non evidenzia alcun confronto significativo tra i vari gruppi. Questo può essere dovuto al ridotto numero di unità statistiche nei gruppi oppure all'elevato numero di confronti. Per *Leishvet_2*, invece, è risultato significativo il confronto tra H e I Ib+III+IV.

Per la variabile *MED_U_POD*, dall'analisi post-hoc risultano significativi i confronti tra il livello IV e tutti gli altri livelli e tra H e III per *Leishvet*; i confronti tra I Ib+III+IV e gli altri livelli per *Leishvet_2*.

Per la variabile *CREA_U*, risulta significativo il confronto tra H e IV per *Leishvet*; il confronto tra H e I Ib+III+IV per *Leishvet_2*.

Per la variabile *MED_U_NEF*, risulta significativo il confronto tra IV e H, tra IV e Ila, tra IV e I Ib per *Leishvet*; il confronto tra H e I Ib+III+IV e tra Ila e I Ib+III+IV per *Leishvet_2*.

Per la variabile *UPC* tutti i confronti post-hoc sono significativi per entrambe le variabili.

Per la variabile *USG*, risulta significativo il confronto tra H e IV per *Leishvet*; il confronto tra H e IIb+III+IV per *Leishvet_2*.

Per *BUN* risultano significativi i confronti tra H e IV, tra IIa e IV e tra IIb e IV per *Leishvet*.

Per *CREA_S*, risultano significativi i confronti tra H e IIa, tra H e IIb e tra H e III per *Leishvet*; il confronto tra H e IIa per *Leishvet_2*.

Infine, per *SDMA* risultano significativi i confronti tra H e IIb, tra H e III, tra H e IV, tra IIa e IV e tra IIb e IV per *Leishvet*; il confronto tra IIb+III+IV e gli altri due livelli per *Leishvet_2*.

In Figura 2.8 sono rappresentati i boxplot delle variabili quantitative rispetto a *IRIS_2* e nella Tabella 2.12 si osservano i p-value del test di Kruskal-Wallis per i medesimi confronti.

TABELLA 2.12: Statistiche osservate e p-value del test di Kruskal-Wallis per le variabili quantitative differenziate per *IRIS_2*.

Variabile	IRIS_2	
	KW_{oss}	p-value
HR	7.96	0.019
SBP	9.37	0.009
MED_U_POD	18.42	< 0.001
CREA_U	12.76	0.002
MED_U_NEF	17.96	0.001
UPC	42.73	< 0.001
BUN	2018.47	< 0.001
CREA_S	36.77	< 0.001
SDMA	31.98	< 0.001

Si procede ora con l'analisi post-hoc con correzione di Holm. Per la variabile *HR* risulta significativo il confronto tra H e I; per *SBP* il confronto tra H e II+III+IV.

Per la variabile *MED_U_POD* sono significativi tutti i confronti; per *CREA_U* il confronto tra H e I; per *MED_U_NEF* il confronto tra H e II+III+IV e tra I e II+III+IV; per *UPC* tutti i confronti; per *BUN* i confronti tra H e II+III+IV e tra I e II+III+IV; infine, per *CREA_S* e per *SDMA* tutti i confronti.

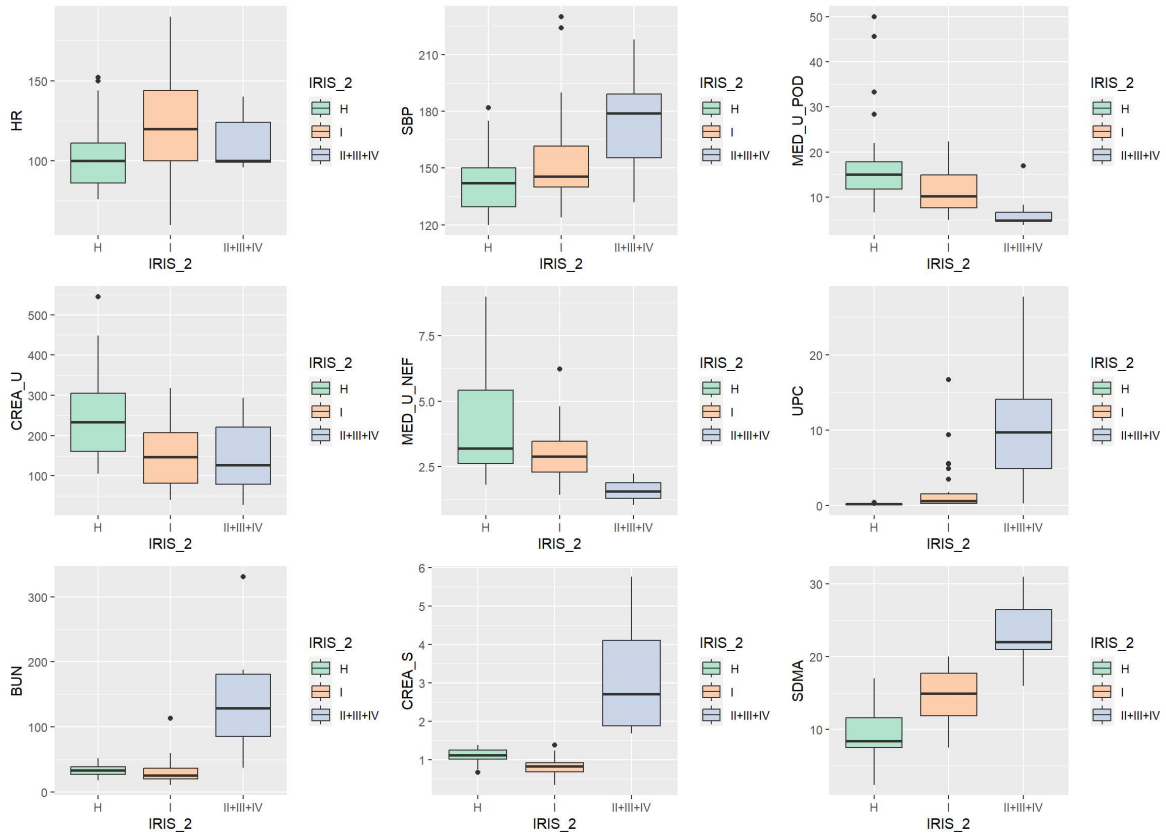


FIGURA 2.8: Boxplot delle variabili quantitative rispetto a *IRIS_2*.

2.2.3 Variabili quantitative rispetto a quantitative

Si studiano ora le correlazioni tra le variabili quantitative. Poiché l'assunzione di normalità è verificata soltanto per le variabili *AGE* e *USG*, si utilizza il coefficiente di correlazione di Spearman, basato sui ranghi, con relativo test per valutare la sua nullità.

In Figura 2.9 si osservano le correlazioni di Spearman per tutte le coppie di variabili quantitative, mentre in Figura 2.10 soltanto quelle che sono risultate significative al livello 5%.

Le correlazioni positive più alte si hanno nelle coppie *CREA_U* - *USG* ($\rho = 0.79$), *MED_U_POD* - *USG* ($\rho = 0.72$), *MED_U_POD* - *CREA_U* ($\rho = 0.69$) e *uNuC* - *uPuC* ($\rho = 0.68$).

Le correlazioni negative più alte si hanno nelle coppie *CREA_U* - *uNuC* ($\rho = -0.7$), *CREA_U* - *UPC* ($\rho = -0.58$) e *MED_U_POD* - *SDMA* ($\rho = -0.54$).

Per queste coppie si riportano in Figura 2.11 i diagrammi di dispersione con le relative rette di regressione, stimate mediante il metodo della regressione locale.

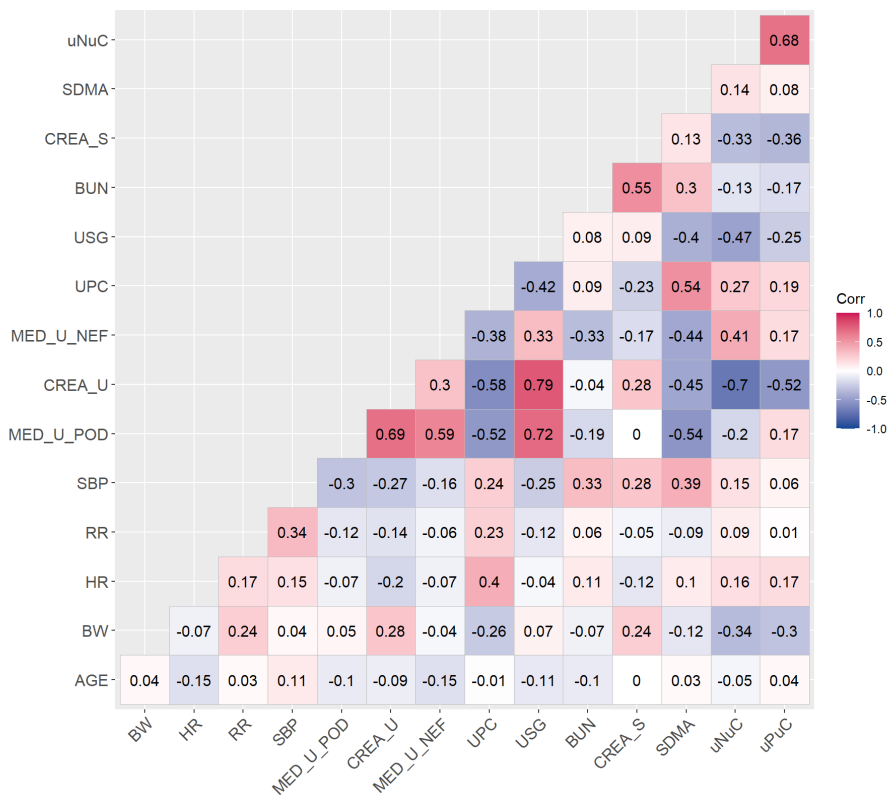


FIGURA 2.9: Correlazioni di Spearman calcolate per le variabili quantitative.

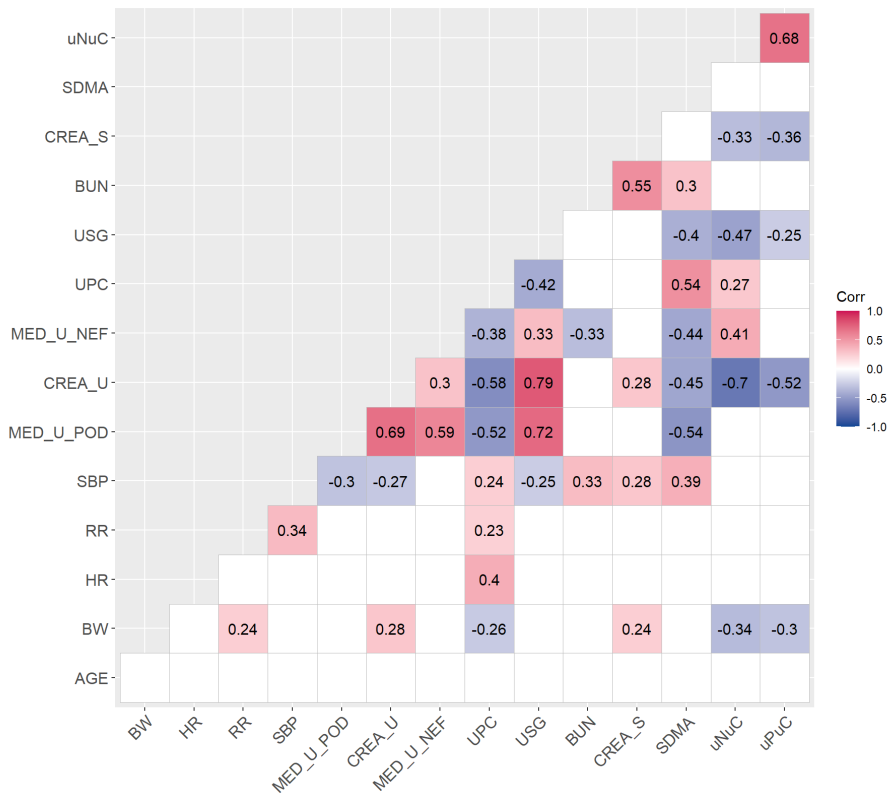


FIGURA 2.10: Correlazioni di Spearman significative al livello 5%.

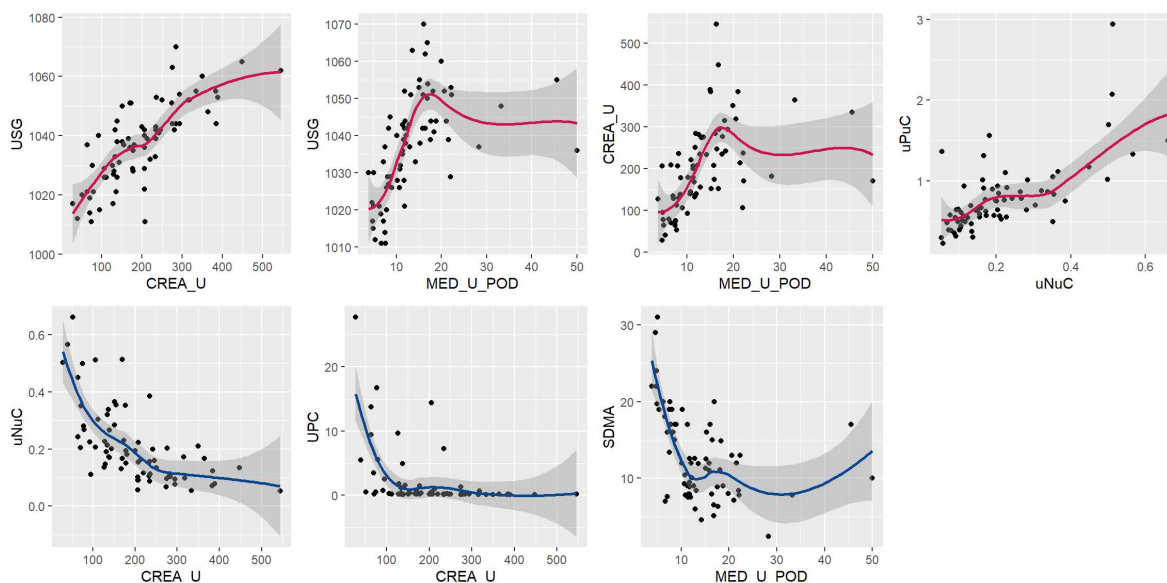


FIGURA 2.11: Diagrammi di dispersione per le variabili maggiormente correlate, con relative rette di regressione stimate.

2.3 Considerazioni finali

In conclusione, le analisi esplorative hanno portato alla rilevazione di una sola associazione significativa tra le variabili qualitative, ovvero quella tra il sesso e lo stato sessuale dei soggetti.

Per quanto riguarda i confronti tra sani e malati, si è osservato che la mediana è minore nel gruppo dei sani rispetto a quello dei malati per le variabili riferite alla frequenza cardiaca, alla frequenza respiratoria, alla pressione sistolica, al rapporto tra proteine urinarie e creatinina urinaria e alla dimetilarginina simmetrica.

Per queste variabili si nota, infatti, un andamento crescente all'aumentare della gravità della malattia, codificata dalle variabili *Leishvet_2* e *IRIS_2*.

Al contrario, la mediana è maggiore nel gruppo dei sani rispetto a quello dei malati per le variabili riferite alla podocina, alla creatinina nelle urine, alla nefrina e alla creatinina nel sangue, e la media è maggiore nei sani per la variabile riferita al peso specifico.

Per queste variabili si nota, infatti, un andamento decrescente all'aumentare della gravità della malattia, codificata dalle variabili *Leishvet_2* e *IRIS_2*, ad esclusione della creatinina nel sangue che decresce nel gruppo I ma è notevolmente più alta nel gruppo II+III+IV.

Si è riscontrato, inoltre, che il peso e la quantità di creatinina nel sangue dei cani di razza sono, rispettivamente, in media e in mediana maggiori di quelli dei cani di razza mista e che la mediana è maggiore nel gruppo delle femmine rispetto a quello dei maschi

per le variabili riferite alla pressione sistolica, alla quantità di urea e alla quantità di creatinina nel sangue.

Infine, si è osservato che le correlazioni positive più alte si hanno tra la creatinina urinaria e il peso specifico USG, tra la podocina e il peso specifico USG, tra la podocina e la creatinina urinaria e tra il rapporto nefrina/creatinina urinaria e il rapporto podocina/creatinina urinaria. Le correlazioni negative più alte si hanno tra la creatinina urinaria e il rapporto nefrina/creatinina urinaria, tra la creatinina urinaria e il rapporto proteine urinarie/creatinina urinaria e tra la podocina e la dimetilarginina simmetrica.

Capitolo 3

Analisi delle curve ROC

La performance di un test diagnostico può essere valutata mediante un'analisi statistica basata sulla curva ROC (Ventura and Racugno, 2017). In questo capitolo verrà valutata l'accuratezza dei test diagnostici considerati nel discriminare, in primo luogo, tra pazienti sani e pazienti malati e, successivamente, tra tre livelli di gravità della malattia.

3.1 Classificazione tra sani e malati

Si definiscono, in via preliminare, alcuni indici della qualità della classificazione di un test diagnostico su due gruppi. In tabella 3.1 è riportata la matrice di confusione con la classificazione dei soggetti.

TABELLA 3.1: Matrice di confusione per la classificazione dei soggetti.

	Malati	Sani	Totale
Positivi	TP (Veri Positivi)	FP (Falsi Positivi)	TP + FP
Negativi	FN (Falsi negativi)	TN (Veri Negativi)	FN + TN
Totale	TP + FN	FP + TN	

- L'**accuratezza** è la proporzione di veri positivi e veri negativi sul totale.
- La **sensibilità** esprime la proporzione di veri positivi sul totale dei pazienti realmente malati.
- La **specificità** esprime la proporzione di veri negativi sul totale dei pazienti realmente sani.

I valori di specificità e sensibilità dipendono da una soglia k di discriminazione fissata (*cut-off*). Poiché il fine di quest'analisi è l'individuazione del maggior numero di malati, nella scelta dei valori di *cut-off* verranno preferiti quelli con sensibilità più elevata.

La curva ROC consiste nella rappresentazione grafica dei punti di coordinate

$$(1 - \text{specificità}, \text{sensibilità}),$$

calcolati al variare della soglia k .

L'area sottesa alla curva ROC, indicata con AUC (*Area Under the ROC Curve*), è il metodo più comunemente utilizzato per sintetizzare l'accuratezza complessiva di un test diagnostico.

Sia X la variabile che rappresenta la misura nel gruppo dei pazienti sani e Y quella nel gruppo dei pazienti malati, l'AUC è data da

$$AUC = P(X < Y).$$

Sotto ipotesi non parametriche, l'AUC è strettamente legato alla statistica test di Mann-Whitney tramite la relazione

$$AUC = 1 - \frac{T_{MW}}{n_x n_y},$$

dove n_x e n_y rappresentano le numerosità dei due campioni casuali semplici tratti da X e Y .

3.1.1 Test diagnostici analizzati

Si è svolta l'analisi delle curve ROC solo per le variabili quantitative per le quali i confronti tra sani e malati sono risultati significativi, ovvero *UPC*, *SDMA*, *MED_U_POD*, *CREA_U*, *HR*, *MED_U_NEF*, *CREA_S*, *SBP*, *USG* e *RR*.

In primo luogo, per ogni variabile considerata si sono rappresentate le curve di densità stimate per il gruppo dei sani e per il gruppo dei malati (Figura 3.1) e le corrispondenti curve ROC (Figure 3.2 e 3.3). In entrambi i grafici le variabili sono state ordinate per valori di AUC decrescenti.

Si è poi calcolato per ciascuna di esse il valore di AUC, con il relativo intervallo di confidenza di livello 95%, e si è svolto il test con ipotesi nulla

$$H_0 : AUC = 0.5.$$

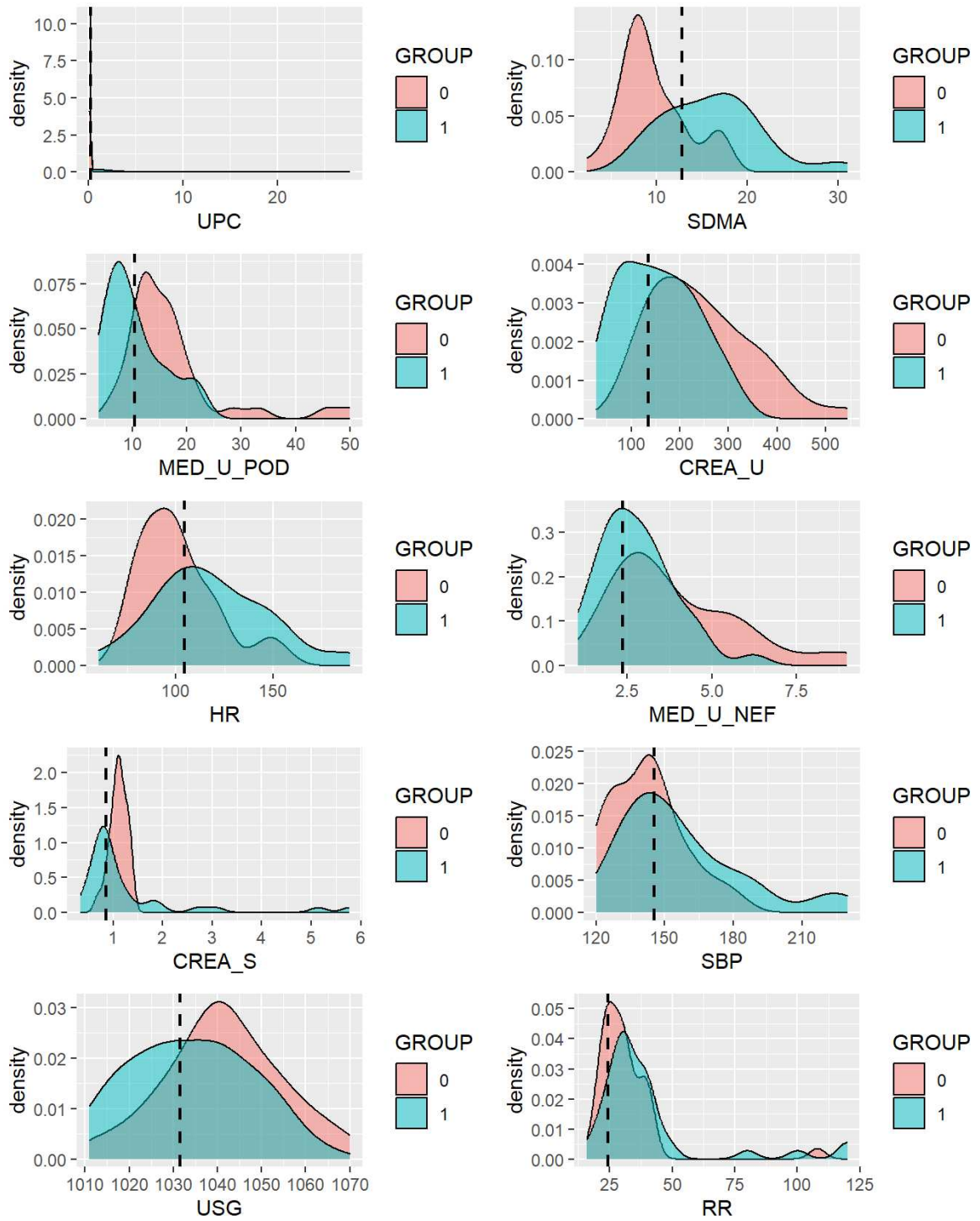


FIGURA 3.1: Densità stimate per le variabili quantitative con relativa soglia di *cut-off*.

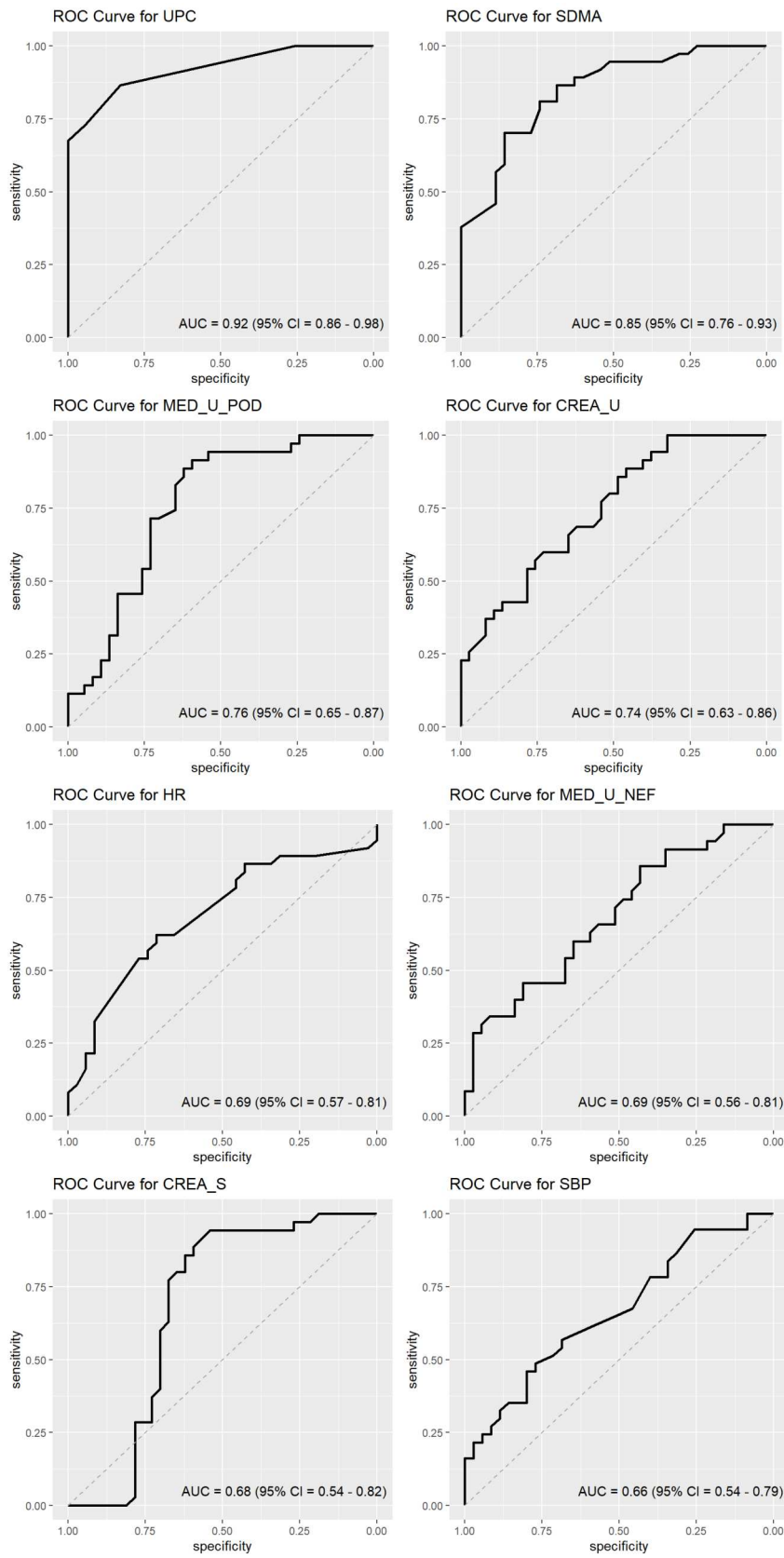


FIGURA 3.2: Curve ROC per le variabili quantitative con relativi AUC e intervallo di confidenza di livello 95% (Parte 1).

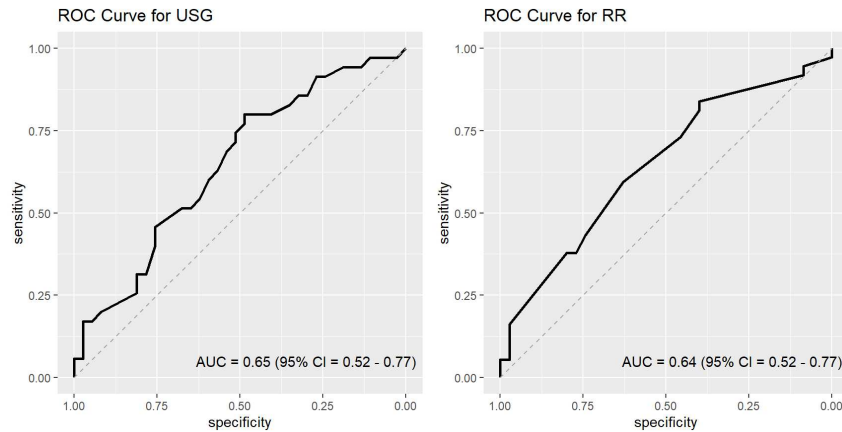


FIGURA 3.3: Curve ROC per le variabili quantitative con relativi AUC e intervallo di confidenza di livello 95% (Parte 2).

Per quanto riguarda la scelta della soglia di *cut-off*, si sono utilizzati il criterio della minima distanza dal punto ottimale (0,1):

$$D_{(0,1)} = \sqrt{(1 - Se)^2 + (1 + Sp)^2},$$

e il criterio di Youden (o della massima distanza dalla bisettrice):

$$D_{line} = Se + Sp - 1.$$

In caso di valori differenti indicati dal criterio di Youden e dal criterio della minima distanza, si è scelto il valore la cui somma di sensibilità e specificità risultava maggiore, preferendo un valore alto della sensibilità, ovvero la capacità di classificare correttamente i soggetti malati.

In Tabella 3.2 sono riportate le statistiche di sintesi delle analisi svolte, ordinando le variabili per valori di AUC decrescenti. Per l'interpretazione del valore stimato di AUC si segue la classificazione proposta da Swets (1998):

- $AUC = 0.5$: test non informativo;
- $0.5 < AUC \leq 0.7$: test poco accurato;
- $0.7 < AUC \leq 0.9$: test moderatamente accurato;
- $0.9 < AUC \leq 1.0$: test altamente accurato;
- $AUC = 1.0$: test perfetto.

TABELLA 3.2: Statistiche di sintesi dell'analisi delle curve ROC per le variabili quantitative, ordinate per valori di AUC decrescenti.

Variabile	AUC (i.c.)	p-value	Cut-off	Sensibilità	Specificità
UPC	0.92 (0.86 - 0.98)	< 0.001	0.25	86%	83%
SDMA	0.85 (0.76 - 0.93)	< 0.001	12.8	70%	86%
MED_U_POD	0.76 (0.65 - 0.87)	< 0.001	10.45	91%	59%
CREA_U	0.74 (0.63 - 0.86)	< 0.001	135.0	89%	46%
HR	0.69 (0.57 - 0.81)	0.003	104.5	62%	71%
MED_U_NEF	0.69 (0.56 - 0.81)	0.003	2.367	86%	43%
CREA_S	0.68 (0.54 - 0.82)	0.004	0.865	94%	54%
SBP	0.66 (0.54 - 0.79)	0.008	145.5	57%	69%
USG	0.65 (0.52 - 0.77)	0.016	1031.5	80%	49%
RR	0.64 (0.52 - 0.77)	0.017	24.5	84%	40%

Si osserva che tutti i p-value sono inferiori a 0.05 e, di conseguenza, si rifiuta l'ipotesi nulla $H_0 : AUC = 0.5$ per ogni variabile. I test sono quindi tutti almeno parzialmente informativi.

3.1.2 Considerazioni finali

Dallo studio delle curve ROC per le dieci variabili risulta che quelle che discriminano meglio tra sani e malati sono, nell'ordine:

- *UPC*, ovvero il rapporto tra proteine urinarie e creatinina urinaria, che presenta $AUC = 0.92$ e un'accuratezza dell'85%;
- *SDMA*, ovvero la quantità di dimetilarginina simmetrica, che presenta $AUC = 0.85$ e un'accuratezza del 78%;
- *MED_U_POD*, ovvero la quantità di podocina, che presenta $AUC = 0.76$ e un'accuratezza del 75%.

Tramite il criterio della minima distanza dal punto ottimale (0,1) e il criterio di Youden, per tali variabili sono stati individuati i valori ottimali per la soglia di *cut-off*:

- $k = 0.25$ per *UPC* (sensibilità = 86%, specificità = 83%),
- $k = 12.8$ per *SDMA* (sensibilità = 70%, specificità = 86%),
- $k = 10.45$ per *MED_U_POD* (sensibilità = 91%, specificità = 59%).

3.2 Classificazione per gravità della malattia

L'analisi della curva ROC tradizionalmente viene utilizzata per test con risposta dicotomica; spesso, tuttavia, in ambito medico è utile distinguere tra più di due alternative diagnostiche. In particolare, frequentemente si vuole identificare uno stadio intermedio di malattia, in aggiunta a quello dei sani e a quello dei malati.

Si definiscono i tre gruppi diagnostici in ordine di gravità di malattia crescente come D^- (sani), D^0 (intermedi) e D^+ (malati) e il test diagnostico come T , una variabile casuale che rappresenta il risultato del test per un soggetto. Si assume che T sia misurata su una scala continua e sia monotona crescente sulla scala di gravità; in caso sia monotona decrescente si utilizzerà il test $-T$.

Sia F_i , $i = -, 0, +$, la funzione di ripartizione di T nei tre gruppi e sia $G_i = 1 - F_i$. Sulla base di due valori di *cut-off*, t_- e t_+ con $t_- < t_+$, gli individui possono essere raggruppati in tre gruppi ordinali: i soggetti con $T \leq t_-$ saranno assegnati al gruppo D^- , quelli con $T \geq t_+$ saranno assegnati al gruppo D^+ e i rimanenti al gruppo D^0 .

Definiamo con x la probabilità che un individuo selezionato casualmente da D^- abbia un risultato del test minore di t_- , ovvero

$$x = Pr \{T \leq t_- | D^-\} = F_-(t_-),$$

e con z la probabilità che un individuo selezionato casualmente da D^+ abbia un risultato del test maggiore di t_+ , ovvero

$$z = Pr \{T \geq t_+ | D^+\} = G_+(t_+).$$

Di conseguenza, $t_- = F_-^{-1}(x)$ e $t_+ = G_+^{-1}(z)$. La probabilità che il risultato del test di un individuo selezionato casualmente da D^0 cada tra t_- e t_+ può essere espressa come

$$y = Pr \{t_- < T < t_+ | D^0\} = F_0(t_+) - F_0(t_-) = F_0(G_+^{-1}(z)) - F_0(F_-^{-1}(x)).$$

Per un test con esito tricotomico è possibile, quindi, rappresentare graficamente la superficie ROC su coordinate tridimensionali, al variare di tutti i possibili valori di *cut-off* (Mossman, 1999). Un esempio di superficie ROC è riportato in Figura 3.4.

La superficie ROC nello spazio tridimensionale (x, y, z) è costituita da $y = y(x, z)$ e il punto di coordinate $(1, 1, 1)$ corrisponde a un test diagnostico che classifica perfettamente gli individui nei tre gruppi.

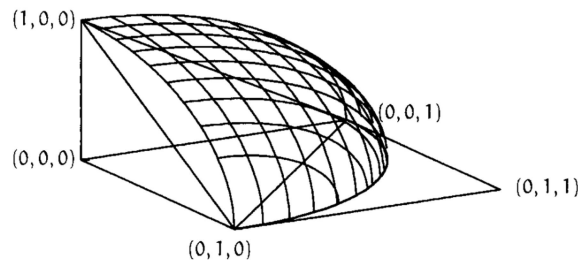


FIGURA 3.4: Esempio di superficie ROC.

La forma funzionale della superficie ROC può essere riscritta come:

$$ROC(x, z) = \begin{cases} F_0(G_+^{-1}(z)) - F_0(F_-^{-1}(x)) & \text{se } F_-^{-1}(x) \leq G_+^{-1}(z), \\ 0 & \text{altrimenti.} \end{cases}$$

La stima non parametrica della superficie ROC si ottiene sostituendo alle funzioni di ripartizione le rispettive funzioni di ripartizione empiriche \hat{F}_- , \hat{F}_0 e \hat{G}_+ (Li and Zhou, 2009).

Volume sotto la superficie ROC (VUS)

Così come l'area sottesa alla curva ROC (*AUC*) sintetizza l'accuratezza complessiva di un test diagnostico dicotomico, si può utilizzare il volume sotto la superficie ROC (*VUS*) come misura di sintesi dell'accuratezza per i test tricotomici.

Quest'ultimo equivale alla probabilità che i valori del test (X, Y, Z) per tre soggetti selezionati in modo casuale, uno per ogni gruppo, siano classificati correttamente, ovvero

$$V_{00} = Pr(X < Y < Z).$$

Il VUS può essere calcolato come

$$V_{00} = \iint_{D_{00}} \{F_0[G_+^{-1}(z)] - F_0[F_-^{-1}(x)]\} dx dz,$$

dove il dominio di integrazione è $D_{00} = \{(x, z) | 0 \leq x \leq 1, 0 \leq z < G_+[F_-^{-1}(x)]\}$ (Xiong *et al.*, 2006).

In presenza di un test diagnostico tricotomico non informativo si avrà $VUS = 1/6$, mentre per un test che classifica perfettamente nei tre gruppi si avrà $VUS = 1$. Inoltre, il VUS aumenta più lentamente dell'*AUC*, come mostrato dal grafico in Figura 3.5. Per esempio, $AUC = 0.8$ corrisponde a $VUS = 0.55$ (Dreiseitl *et al.*, 2000).

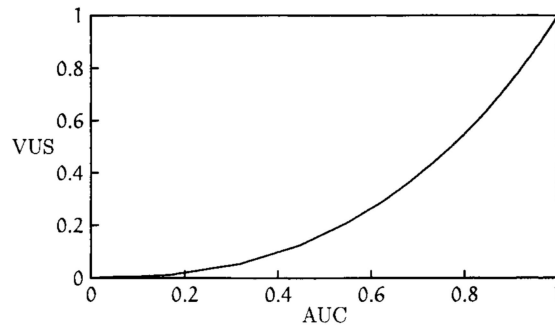


FIGURA 3.5: Grafico dei valori del VUS contro i valori dell'AUC.

Di conseguenza, si modifica la scala di interpretazione. Seguendo la classificazione per l'AUC proposta da Swets (1998), si avrà un test poco accurato per $0.17 < VUS \leq 0.42$, un test moderatamente accurato per $0.42 < VUS \leq 0.74$ e un test altamente accurato per $0.74 < VUS \leq 1.00$.

Uno stimatore non-parametrico non distorto del VUS è dato da

$$\widehat{VUS} = \frac{1}{n_x n_y n_z} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_z} I(X_i, Y_j, Z_k),$$

dove n_x, n_y, n_z indicano le numerosità campionarie dei tre gruppi e $I(X, Y, Z)$ è uguale a uno se (X, Y, Z) sono nell'ordine corretto e zero altrimenti.

In presenza di valori uguali di misurazione (*ties*), si definisce $I(X, Y, Z) = 1/2$ se $(X = Y < Z)$ o se $(X < Y = Z)$ e $I(X, Y, Z) = 1/6$ se $(X = Y = Z)$ (Nakas and Yiannoutsos, 2004). Il valore atteso del VUS è

$$Pr(X < Y < Z) + \frac{1}{2}Pr(X = Y < Z) + \frac{1}{2}Pr(X < Y = Z) + \frac{1}{6}Pr(X = Y = Z).$$

Per valutare se un test diagnostico è almeno parzialmente informativo si testa l'ipotesi nulla $H_0 : VUS = 1/6$. La statistica test è

$$Z_1 = \frac{\widehat{VUS} - 1/6}{\sqrt{\text{var}(\widehat{VUS})}},$$

la cui distribuzione nulla approssimata è $N(0,1)$ e la varianza può essere calcolata usando la metodologia delle statistiche di Mann-Whitney oppure il bootstrap (Nakas, 2014). Tramite questa statistica test si possono anche costruire intervalli di confidenza per il VUS.

Infine, per confrontare i VUS di due diversi test diagnostici si utilizza la statistica test proposta da Dreiseitl *et al.* (2000), data da

$$Z_2 = \frac{\widehat{VUS}^A - \widehat{VUS}^B}{\sqrt{\widehat{var}(\widehat{VUS}^A) + \widehat{var}(\widehat{VUS}^B) - 2\widehat{cov}(\widehat{VUS}^A, \widehat{VUS}^B)}},$$

la cui distribuzione nulla approssimata è $N(0,1)$.

Indice di Youden

Le probabilità di classificare correttamente i soggetti nei tre gruppi sono definite individualmente come $Sp(t_-) = P_-(T \leq t_-) = F_-(t_-)$, $Se(t_+) = P_+(T \geq t_+) = 1 - F_+(t_+)$ e $Sm(t_-, t_+) = P_0(t_- \leq T \leq t_+) = F_0(t_+) - F_0(t_-)$. L'indice di Youden che valuta l'accuratezza diagnostica di un test per tre gruppi ordinali può essere definito a partire dalla somma delle tre probabilità di corretta classificazione, ovvero

$$J(t_-, t_+) = \frac{1}{2}[Sp(t_-) + Sm(t_-, t_+) + Se(t_+) - 1] = \frac{1}{2}[F_-(t_-) - F_0(t_-) + F_0(t_+) - F_+(t_+)],$$

i cui valori cadono nell'intervallo $[0; 1]$, dove 0 indica un'assegnazione casuale ai gruppi e 1 una perfetta separazione tra i gruppi.

Una coppia ottimale di valori di *cut-off* può essere ricavata massimizzando J su tutte le possibili coppie. La stima non parametrica dell'indice di Youden si ottiene sostituendo alle funzioni di ripartizione le rispettive funzioni di ripartizione empiriche; i valori ottimali di *cut-off* si individuano tramite algoritmi di ottimizzazione (Luo and Xiong, 2013).

3.2.1 Test diagnostici analizzati

Si valuta ora l'accuratezza dei test diagnostici nel discriminare tra i livelli di gravità della malattia, determinati dalle variabili tricotomiche *Leishvet_2* e *IRIS_2*, tramite l'analisi delle superfici ROC.

La variabile *Leishvet_2* suddivide gli individui nel gruppo *H* che rappresenta i sani, nel gruppo *Iia* che rappresenta il livello intermedio di malattia e nel gruppo *Iib+III+IV* che rappresenta gli stadi più gravi. Si considerano soltanto le variabili quantitative per le quali i confronti tra gruppi sono risultati significativi.

In primo luogo, si verifica che i valori osservati delle variabili seguano un andamento monotono crescente (o decrescente) secondo la scala di gravità della malattia. In Tabella

3.3 si riportano le medie e in Tabella 3.4 le mediane nei tre gruppi per le variabili considerate.

TABELLA 3.3: Medie (Std.Dev.) delle variabili quantitative considerate nei tre gruppi definiti da *Leishvet_2*.

Variabile	Medie (Std.Dev.)		
	H	IIa	IIb+III+IV
HR	101.0 (19.9)	119.5 (31.0)	118.1 (29.1)
MED_U_POD	17.06 (9.4)	12.96 (4.6)	9.66 (5.5)
CREA_U	245.4 (105.5)	185.1 (65.3)	142.7 (81.6)
MED_U_NEF	3.96 (1.9)	3.67 (1.2)	2.44 (0.9)
UPC	0.2 (0.08)	0.3 (0.08)	5.1 (6.66)
USG	1041.3 (13.3)	1038.7 (12.0)	1032.0 (13.8)
CREA_S	1.11 (0.17)	0.92 (0.15)	1.42 (1.37)
SDMA	9.6 (3.7)	12.5 (4.3)	17.7 (5.2)

TABELLA 3.4: Mediane (IQR) delle variabili quantitative considerate nei tre gruppi definiti da *Leishvet_2*.

Variabile	Mediane (IQR)		
	H	IIa	IIb+III+IV
HR	100.0 (25)	120.0 (20)	120.0 (40)
MED_U_POD	15.06 (6.1)	11.77 (5.6)	7.93 (5.1)
CREA_U	233.0 (143.5)	177.0 (96.5)	131.5 (130.5)
MED_U_NEF	3.20 (2.8)	3.41 (1.5)	2.25 (1.0)
UPC	0.2 (0.05)	0.3 (0.1)	1.7 (6.15)
USG	1042.0 (16.5)	1041.0 (15.0)	1030.0 (20.5)
CREA_S	1.11 (0.23)	0.89 (0.14)	0.83 (0.96)
SDMA	8.4 (4.1)	11.0 (7.5)	17.5 (6.9)

L'andamento monotono crescente viene rispettato per *UPC* e *SDMA*, mentre l'andamento monotono decrescente per *MED_U_POD*, *CREA_U* e *USG*; le altre variabili vengono escluse dall'analisi.

Tramite il pacchetto R *DiagTest3Grp* (Luo and Xiong, 2012), si sono stimate in modo non-parametrico le due misure di sintesi dell'accuratezza per i test tricotomici precedentemente espone, ovvero il VUS e l'indice di Youden.

Per entrambe si sono calcolati i valori ottimali di *cut-off*; essi corrispondono, rispettivamente, alla coppia di coordinate sulla superficie ROC con la minima distanza dal punto ottimale (1,1,1) e alla coppia di valori che massimizzano l'indice di Youden. Nel caso in cui le coppie ottenute fossero differenti, si sono scelte quelle con valori più elevati delle probabilità di classificare correttamente i soggetti malati, $Se(t_+)$, e delle probabilità di classificare correttamente i soggetti con livello intermedio, $Sm(t_-, t_+)$.

Ulteriori informazioni sul pacchetto e sulle funzioni utilizzate sono riportate in Appendice.

In Tabella 3.5 si riportano le statistiche di sintesi relative al VUS, con le variabili ordinate per valori di VUS decrescenti, e in Tabella 3.6 quelle relative all'indice di Youden (J), con le variabili ordinate per valori dell'indice decrescenti.

TABELLA 3.5: Statistiche di sintesi relative al VUS per le variabili quantitative suddivise secondo *Leishvet_2*, ordinate per valori di VUS decrescenti.

Variabile	VUS (i.c.)	t_-	t_+	$Sp(t_-)$	$Sm(t_-, t_+)$	$Se(t_+)$
UPC	0.733 (0.591 - 0.849)	0.2	0.5	83%	54%	96%
SDMA	0.501 (0.350 - 0.634)	8.4	18.0	51%	73%	50%
MED_U_POD	0.444 (0.293 - 0.579)	8.63	12.07	71%	45%	65%
CREA_U	0.393 (0.260 - 0.485)	144	274	40%	64%	62%
USG	0.299 (0.191 - 0.402)	1035	1043	46%	36%	61%

TABELLA 3.6: Statistiche di sintesi relative all'indice di Youden per le variabili quantitative suddivise secondo *Leishvet_2*, ordinate per valori decrescenti dell'indice.

Variabile	J (i.c.)	t_-	t_+	$Sp(t_-)$	$Sm(t_-, t_+)$	$Se(t_+)$
UPC	0.591 (0.463 - 0.773)	0.23	0.51	83%	54%	81%
MED_U_POD	0.279 (0.121 - 0.363)	10.2	18.5	23%	64%	69%
CREA_U	0.261 (0.113 - 0.449)	141.4	249.1	40%	54%	58%
SDMA	0.244 (0.162 - 0.390)	11.7	15.6	74%	9%	65%
USG	0.200 (0.142 - 0.273)	1036	1062	6%	73%	61%

Seguendo il metodo esposto in precedenza, si scelgono i valori ottimali di *cut-off* ottenuti tramite il criterio della minima distanza dal punto ottimale (1,1,1) per le variabili *UPC*, *SDMA* e *CREA_U*; si scelgono i valori ottenuti tramite il criterio di Youden per *MED_U_POD* e *USG*.

Si osserva che nessuno degli intervalli di confidenza per il VUS (Tabella 3.5) contiene il valore 1/6, quindi i test diagnostici analizzati sono tutti almeno parzialmente informativi.

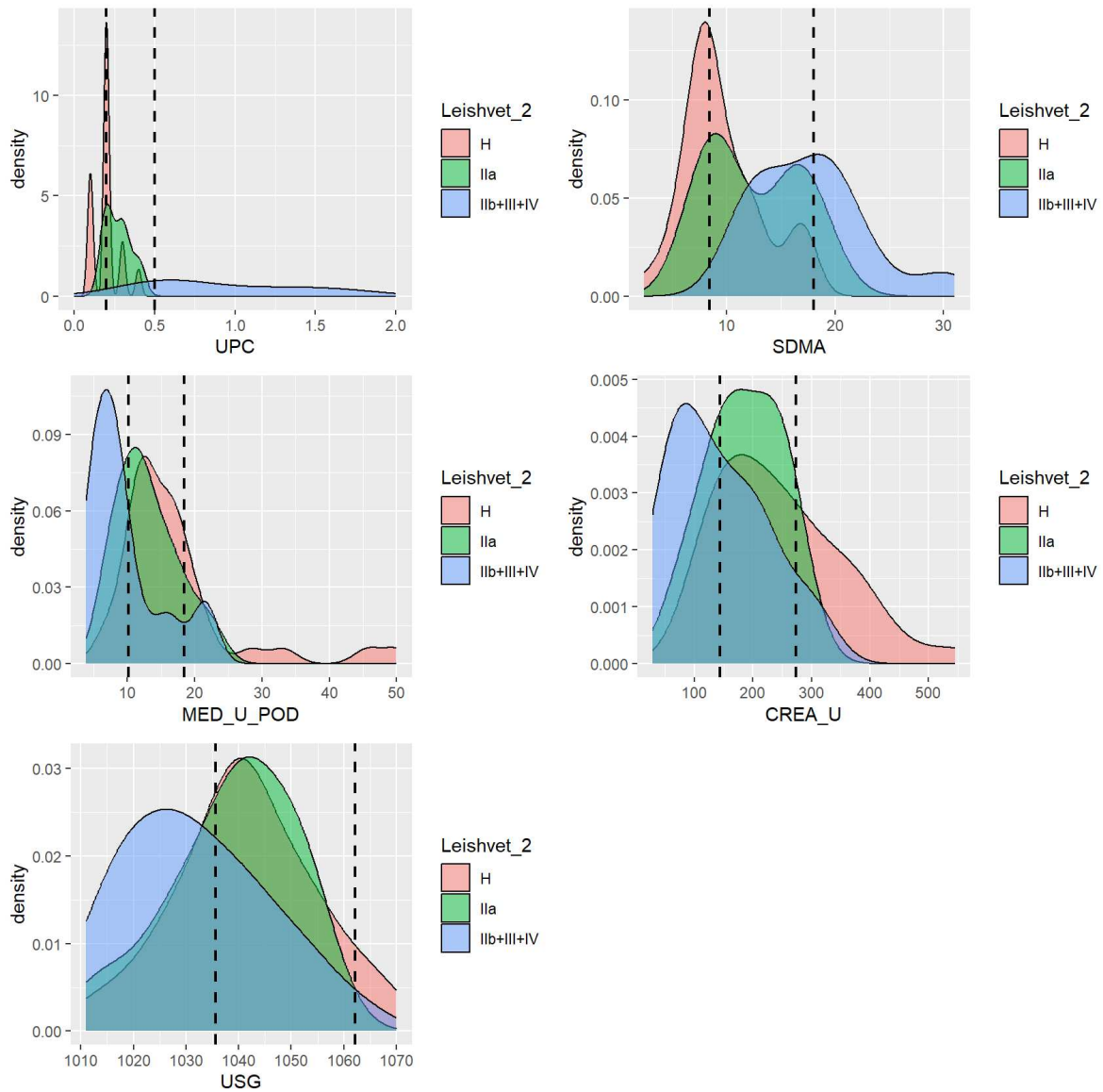


FIGURA 3.6: Densità stimate per i test diagnostici analizzati con relative soglie di *cut-off*.

In Figura 3.6 sono rappresentate le densità stimate per le variabili considerate, suddivise nei tre gruppi identificati da *Leishvet_2*, con le relative soglie di *cut-off* individuate.

La rappresentazione tridimensionale delle superfici ROC, riportata in Figura 3.7, è stata realizzata utilizzando il pacchetto R *bcROCsurface* (To Duc, 2017); per ulteriori informazioni si rimanda all'Appendice.

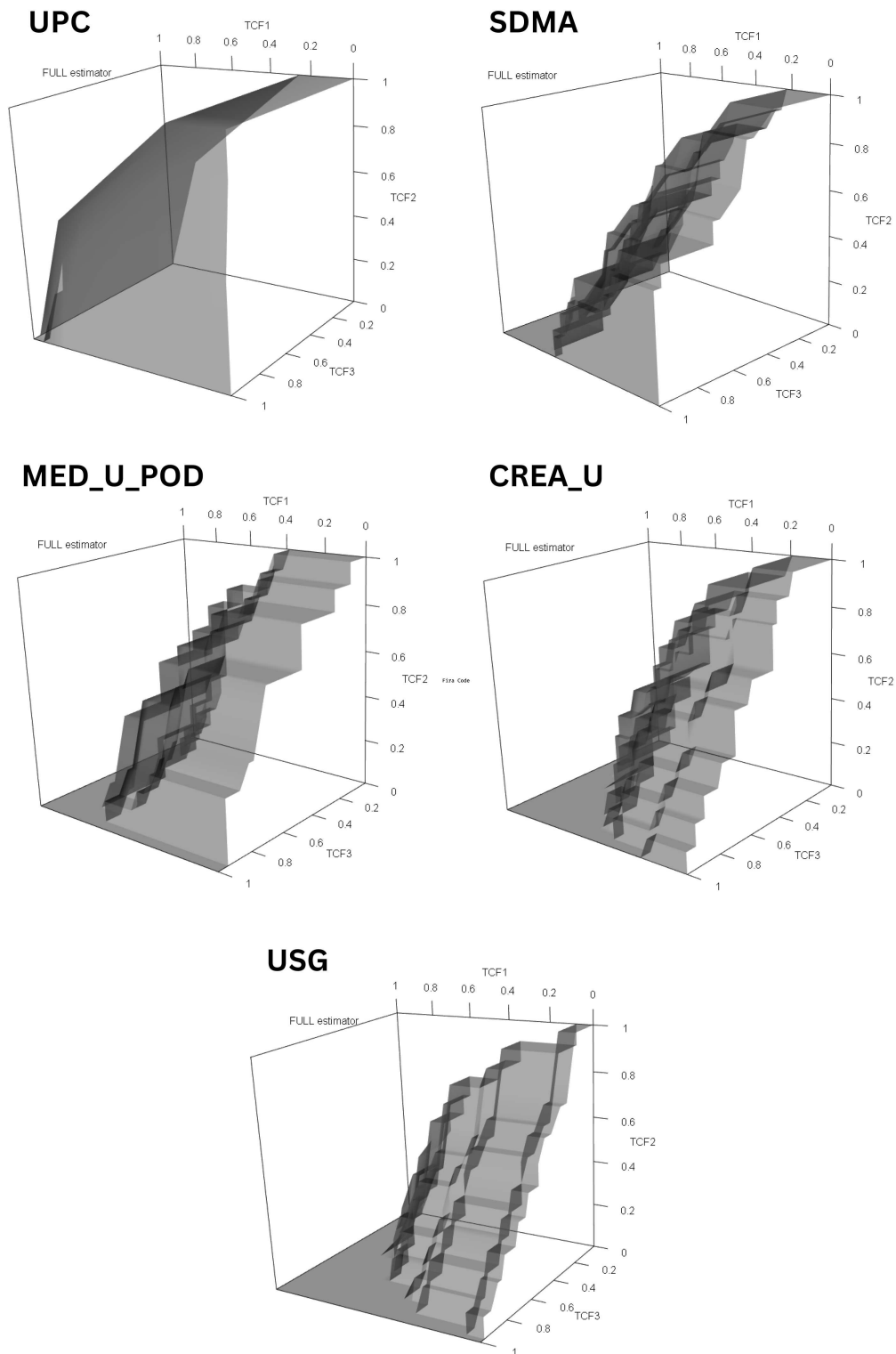


FIGURA 3.7: Superfici ROC per i test diagnostici analizzati, relative alla variabile *Leishvet_2*.

Si ripete l'analisi svolta anche per la variabile *IRIS_2*. Essa suddivide gli individui nel gruppo *H* che rappresenta i sani, nel gruppo *I* che rappresenta il livello intermedio di malattia e nel gruppo *II+III+IV* che rappresenta gli stadi più gravi. Si considerano soltanto le variabili quantitative per le quali i confronti tra gruppi sono risultati significativi.

Si verifica che i valori osservati delle variabili seguano un andamento monotono crescente (o decrescente) secondo la scala di gravità della malattia. In Tabella 3.7 si riportano le medie e in Tabella 3.8 le mediane nei tre gruppi per le variabili considerate.

TABELLA 3.7: Medie (Std.Dev.) delle variabili quantitative considerate nei tre gruppi definiti da *IRIS_2*.

Variabile	Medie (Std.Dev.)		
	H	I	II+III+IV
HR	101.0 (19.9)	120.1 (31.4)	111.7 (17.5)
SBP	142.4 (15.9)	153.4 (25.5)	174.0 (28.6)
MED_U_POD	17.1 (9.4)	11.5 (5.2)	6.9 (4.6)
CREA_U	245.4 (105.4)	156.6 (75.6)	149.9 (97.4)
MED_U_NEF	3.96 (1.9)	3.09 (1.0)	1.60 (0.4)
UPC	0.2 (0.1)	2.0 (3.5)	10.8 (9.1)
BUN	32.4 (7.8)	30.6 (19.0)	147.1 (97.5)
CREA_S	1.11 (0.2)	0.83 (0.2)	3.16 (1.6)
SDMA	9.6 (3.7)	14.5 (3.9)	23.4 (5.2)

TABELLA 3.8: Mediane (IQR) delle variabili quantitative considerate nei tre gruppi definiti da *IRIS_2*.

Variabile	Mediane (IQR)		
	H	I	II+III+IV
HR	100.0 (25.0)	120.0 (43.7)	100.0 (25.0)
SBP	142.0 (20.5)	145.5 (21.5)	179.0 (33.5)
MED_U_POD	15.1 (6.1)	10.2 (7.2)	4.8 (1.9)
CREA_U	233.0 (143.5)	147.5 (124.5)	127.0 (141.0)
MED_U_NEF	3.20 (2.8)	2.89 (1.2)	1.55 (0.6)
UPC	0.2 (0.1)	0.6 (1.3)	9.7 (9.2)
BUN	33.0 (11.5)	25.5 (16.5)	129.0 (95.5)
CREA_S	1.11 (0.2)	0.83 (0.2)	2.71 (2.2)
SDMA	8.4 (4.1)	14.9 (5.8)	22.0 (5.5)

L'andamento monotono crescente viene rispettato per *SBP*, *UPC* e *SDMA*, mentre l'andamento monotono decrescente per *MED_U_POD*, *CREA_U* e *MED_U_NEF*; le altre variabili vengono escluse dall'analisi.

In Tabella 3.9 si osservano le statistiche di sintesi relative al VUS, con le variabili ordinate per valori di VUS decrescenti, e in Tabella 3.10 quelle relative all'indice di Youden (J), con le variabili ordinate per valori dell'indice decrescenti.

TABELLA 3.9: Statistiche di sintesi relative al VUS per le variabili quantitative suddivise secondo *IRIS_2*, ordinate per valori di VUS decrescenti.

Variabile	VUS (i.c.)	t_-	t_+	$Sp(t_-)$	$Sm(t_-, t_+)$	$Se(t_+)$
SDMA	0.754 (0.614 - 0.875)	10.1	20.0	69%	77%	86%
UPC	0.747 (0.556 - 0.895)	0.2	7.3	83%	77%	71%
MED_U_POD	0.569 (0.351 - 0.771)	5.0	12.1	71%	67%	71%
MED_U_NEF	0.567 (0.427 - 0.665)	2.0	3.6	46%	67%	86%
SBP	0.400 (0.239 - 0.542)	145	179	69%	40%	57%
CREA_U	0.355 (0.135 - 0.557)	128	220	54%	47%	57%

TABELLA 3.10: Statistiche di sintesi relative all'indice di Youden per le variabili quantitative suddivise secondo *IRIS_2*, ordinate per valori decrescenti dell'indice.

Variabile	J (i.c.)	t_-	t_+	$Sp(t_-)$	$Sm(t_-, t_+)$	$Se(t_+)$
UPC	0.640 (0.485 - 0.755)	0.4	7.1	100%	57%	71%
SDMA	0.514 (0.440 - 0.755)	12.1	18.9	77%	40%	86%
MED_U_NEF	0.455 (0.381 - 0.584)	2.3	4.2	34%	57%	100%
MED_U_POD	0.269 (0.187 - 0.502)	9.7	17.2	31%	37%	86%
SBP	0.233 (0.060 - 0.429)	158.2	167.0	83%	7%	57%
CREA_U	0.167 (0.120 - 0.295)	80.4	220.7	51%	53%	29%

Seguendo il metodo esposto in precedenza, si scelgono i valori ottimali di *cut-off* ottenuti tramite il criterio della minima distanza dal punto ottimale (1,1,1) per le variabili *SDMA*, *UPC*, *MED_U_POD*, *SBP* e *CREA_U*; si scelgono i valori ottenuti tramite il criterio di Youden per *MED_U_NEF*.

Si osserva che l'intervallo di confidenza per il VUS relativo alla variabile *CREA_U* (Tabella 3.9) contiene il valore 1/6 e quindi non vi è evidenza per affermare che tale test sia informativo, al contrario degli altri.

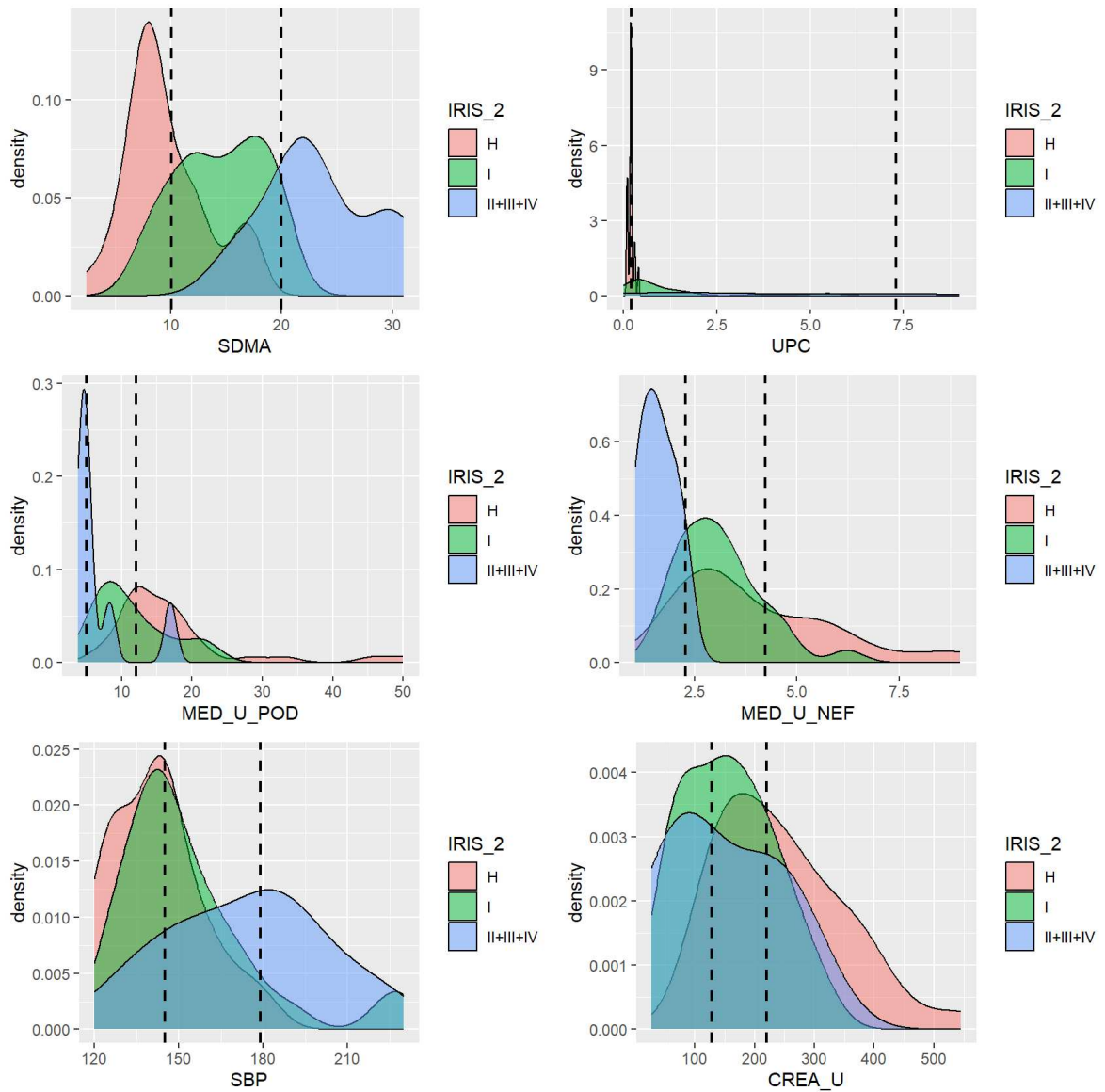


FIGURA 3.8: Densità stimate per i test diagnostici analizzati con relative soglie di *cut-off*.

In Figura 3.8 sono rappresentate le densità stimate per le variabili considerate, suddivise nei tre gruppi identificati da *IRIS_2*, con le relative soglie di *cut-off* individuate; in Figura 3.9 si possono osservare le rappresentazioni tridimensionali delle superfici ROC.

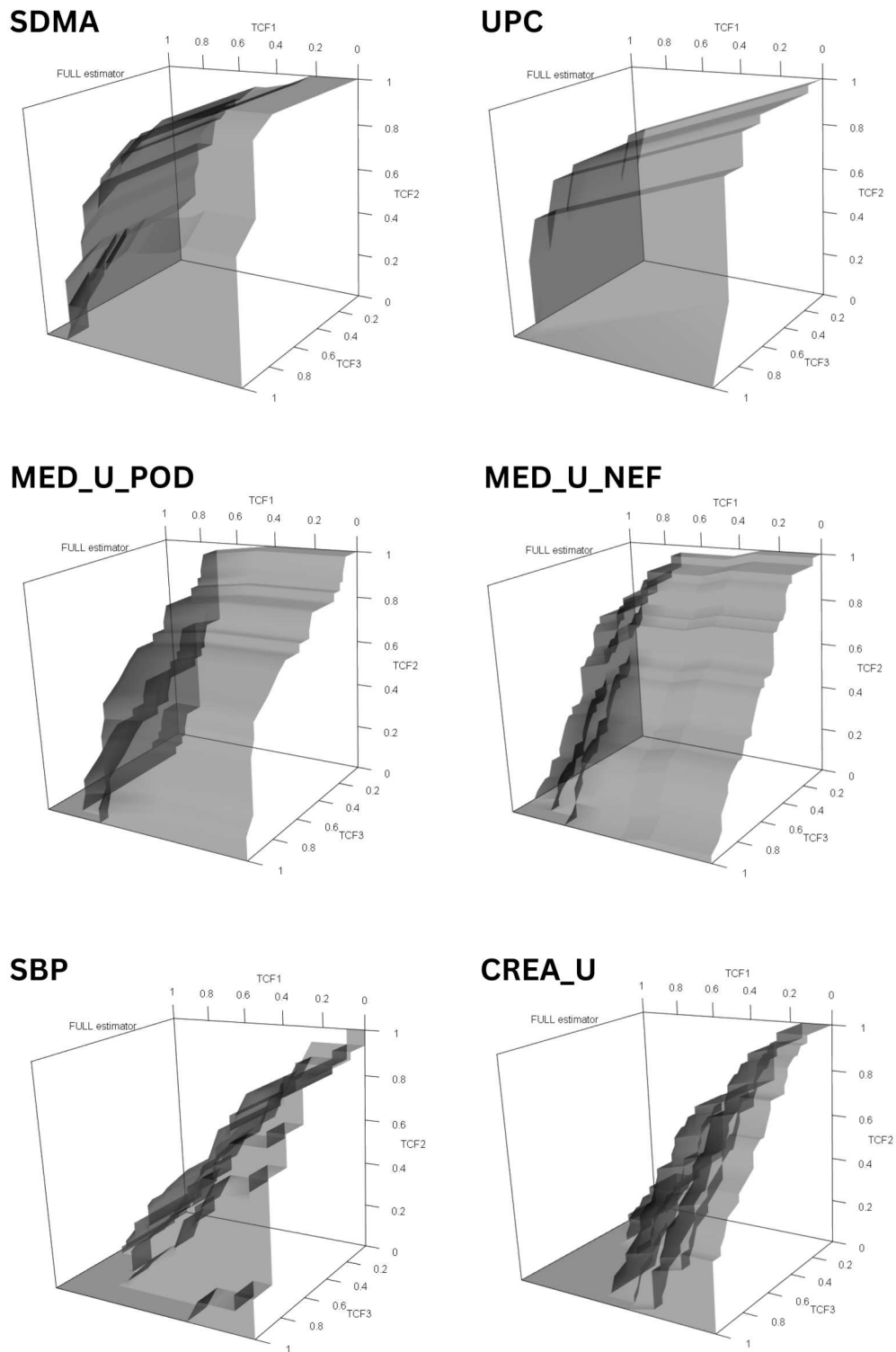


FIGURA 3.9: Superfici ROC per i test diagnostici analizzati, relative alla variabile *IRIS_2*.

3.2.2 Considerazioni finali

Dall'analisi delle superfici ROC sulle variabili quantitative risulta che, per la variabile *Leishvet_2*, i test che discriminano meglio tra i tre livelli di malattia sono, nell'ordine:

- **UPC**, ovvero il rapporto tra proteine urinarie e creatinina urinaria, che presenta $VUS = 0.733$ ($0.591 - 0.849$) e per cui i valori ottimali di *cut-off* sono $t_- = 0.2$ e $t_+ = 0.5$ ($Sp(t_-) = 83\%$, $Sm(t_-, t_+) = 54\%$, $Se(t_+) = 96\%$);
- **SDMA**, ovvero la quantità di dimetilarginina simmetrica, che presenta $VUS = 0.501$ ($0.350 - 0.634$) e per cui i valori ottimali di *cut-off* sono $t_- = 8.4$ e $t_+ = 18.0$ ($Sp(t_-) = 51\%$, $Sm(t_-, t_+) = 73\%$, $Se(t_+) = 50\%$);
- **MED_U_POD**, ovvero la quantità di podocina, che presenta $VUS = 0.444$ ($0.293 - 0.579$) e per cui i valori ottimali di *cut-off* sono $t_- = 10.2$ e $t_+ = 18.5$ ($Sp(t_-) = 23\%$, $Sm(t_-, t_+) = 64\%$, $Se(t_+) = 69\%$).

Secondo la scala di interpretazione di Swets (1998), tutti e tre i test sono moderatamente accurati.

Per la variabile *IRIS_2*, i test che discriminano meglio tra i tre livelli di malattia sono, nell'ordine:

- **SDMA**, ovvero la quantità di dimetilarginina simmetrica, che presenta $VUS = 0.754$ ($0.614 - 0.875$) e per cui i valori ottimali di *cut-off* sono $t_- = 10.1$ e $t_+ = 20.0$ ($Sp(t_-) = 69\%$, $Sm(t_-, t_+) = 77\%$, $Se(t_+) = 86\%$);
- **UPC**, ovvero il rapporto tra proteine urinarie e creatinina urinaria, che presenta $VUS = 0.747$ ($0.556 - 0.895$) e per cui i valori ottimali di *cut-off* sono $t_- = 0.2$ e $t_+ = 7.3$ ($Sp(t_-) = 83\%$, $Sm(t_-, t_+) = 77\%$, $Se(t_+) = 71\%$);
- **MED_U_POD**, ovvero la quantità di podocina, che presenta $VUS = 0.569$ ($0.351 - 0.771$) e per cui i valori ottimali di *cut-off* sono $t_- = 5.0$ e $t_+ = 12.1$ ($Sp(t_-) = 71\%$, $Sm(t_-, t_+) = 67\%$, $Se(t_+) = 71\%$);
- **MED_U_NEF**, ovvero la quantità di nefrina, che presenta $VUS = 0.567$ ($0.427 - 0.665$) e per cui i valori ottimali di *cut-off* sono $t_- = 2.3$ e $t_+ = 4.2$ ($Sp(t_-) = 34\%$, $Sm(t_-, t_+) = 57\%$, $Se(t_+) = 100\%$).

Secondo la scala di interpretazione di Swets (1998), i primi due test sono altamente accurati e gli ultimi due moderatamente accurati.

Capitolo 4

Regressione logistica

In questo capitolo l'obiettivo è modellare lo stato del paziente, determinato dalla variabile *GROUP*, in funzione delle altre variabili. Poiché la variabile risposta è dicotomica (sano/malato) si utilizza il modello di regressione logistica (Ventura and Racugno, 2017).

La regressione logistica appartiene alla classe dei modelli lineari generalizzati (GLM), nei quali la variabile risposta Y è non gaussiana e la relazione tra la media della variabile e il predittore lineare $\eta_i = \beta^T x_i$ è espressa tramite una funzione di legame $g(\cdot)$.

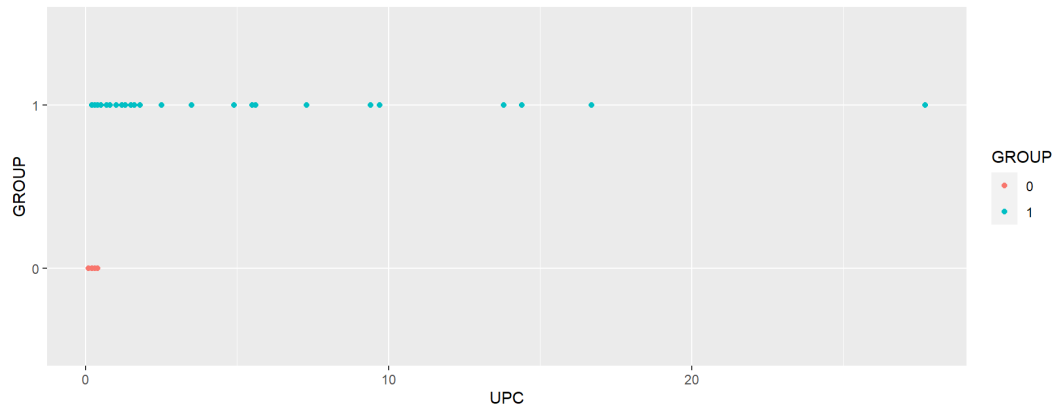
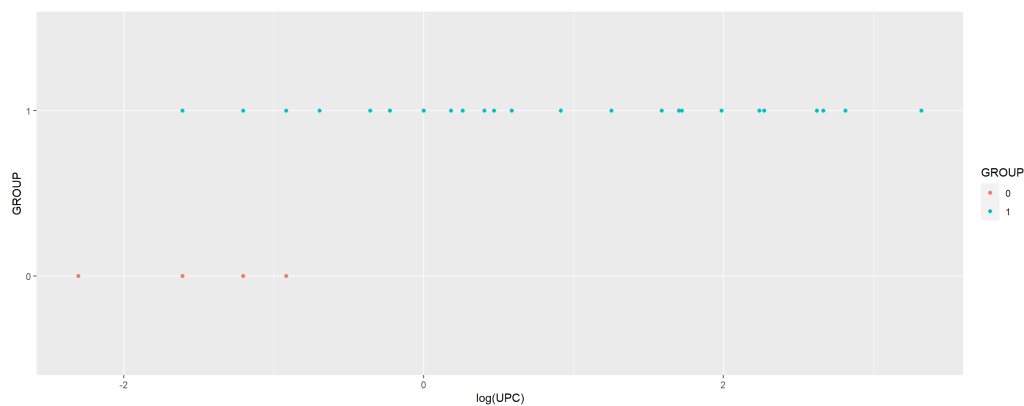
In particolare, il modello logistico è specificato dalle seguenti ipotesi:

- Y_1, \dots, Y_n sono variabili indipendenti, con $Y_i \sim Ber(\pi_i)$;
- il predittore lineare è $\eta_i = \beta^T x_i$, con $x_i = (x_{i1}, \dots, x_{ip})$ riga i -esima della matrice X di dimensioni $(n \times p)$, non stocastica e con rango pieno, e con $\beta = (\beta_1, \dots, \beta_p)$ vettore dei parametri;
- la funzione di legame è la funzione *logit*, espressa da

$$g(\pi_i) = \log \frac{\pi_i}{1 - \pi_i} = \beta^T x_i.$$

Questo modello presenta il vantaggio che i suoi coefficienti possono essere interpretati in termini di logaritmi dei rapporti tra quote (*log-odds ratio*).

Se la variabile esplicativa x è dicotomica il coefficiente β_j rappresenta il logaritmo del rapporto della quota di $Y = 1$ se $x = 1$ contro la quota di $Y = 1$ se $x = 0$, mentre se x è quantitativa β_j rappresenta il logaritmo della variazione percentuale della quota di $Y = 1$ corrispondente a un aumento unitario di x .

FIGURA 4.1: Scatterplot della variabile UPC rispetto a $GROUP$.FIGURA 4.2: Scatterplot della trasformata logaritmica di UPC rispetto a $GROUP$.

4.1 Stima del modello

Per la selezione del modello, dato il numero elevato di variabili, si procede con un approccio *forward*. Si stima il modello con la sola intercetta e poi si valuta di volta in volta se aggiungere la variabile che presenta il minimo p -value relativo al test del log-rapporto di verosimiglianza.

Seguendo questo metodo, la prima variabile da aggiungere è UPC . Tuttavia, l'*output* di R presenta un *warning*; si osserva, infatti, che vi sono numerose probabilità stimate pari a 1.

Per identificare la causa di ciò si osserva il grafico della variabile UPC rispetto a $GROUP$ (Figura 4.1). Nel gruppo dei malati ($GROUP=1$) la variabile presenta un range di valori molto ampio, il che porta alla stima di probabilità pari a 1 per i valori più elevati.

Per ridurre il range di valori, una soluzione è applicare una trasformazione logaritmica alla variabile UPC (Figura 4.2).

Riprendendo il processo di selezione *forward* con la variabile *UPC* trasformata, al modello con sola intercetta si aggiungono, nell'ordine, le variabili $\log(UPC)$ (p -value < 0.001), *SDMA* (p -value = 0.003), *RR* (p -value = 0.03) e *BUN* (p -value = 0.04).

L'adattamento del modello finale stimato è riportato in Tabella 4.1.

TABELLA 4.1: Adattamento del modello logistico.

Variabile	Stima	Std. Error	z-value	p-value
Intercetta	2.767	2.818	0.982	0.326
$\log(UPC)$	4.867	1.679	2.899	0.004
<i>SDMA</i>	0.344	0.124	2.766	0.006
<i>RR</i>	0.043	0.020	2.204	0.027
<i>BUN</i>	-0.075	0.028	-2.658	0.008

Il modello finale presenta tutti i coefficienti significativi al livello 5% e $AIC = 37.98$. L'ipotesi di omogeneità contro il modello corrente è rifiutata ($W_{olc}^{oss} = 71.78$, p -value < 0.001) e il modello corrente contro il modello saturo risulta soddisfacente ($W_{cls}^{oss} = 27.98 < 67$ g.d.l.).

Il coefficiente stimato relativo a $\log(UPC)$ è $\hat{\beta}_1 = 4.867$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_1} = 129.93$; esso indica che per ogni aumento unitario del logaritmo di *UPC* la quota di malattia aumenta di quasi 130 volte, a parità delle altre variabili.

Il coefficiente stimato relativo alla variabile *SDMA* è $\hat{\beta}_2 = 0.344$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_2} = 1.41$; esso indica che per ogni aumento unitario di *SDMA* la quota di malattia aumenta del 41%, a parità delle altre variabili.

Il coefficiente stimato relativo alla variabile *RR* è $\hat{\beta}_3 = 0.043$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_3} = 1.04$; esso indica che per ogni aumento unitario di *RR* la quota di malattia aumenta del 4%, a parità delle altre variabili.

Infine, il coefficiente stimato relativo alla variabile *BUN* è $\hat{\beta}_4 = -0.075$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_4} = 0.93$; esso indica che per ogni aumento unitario di *BUN* la quota di malattia diminuisce del 7%, a parità delle altre variabili.

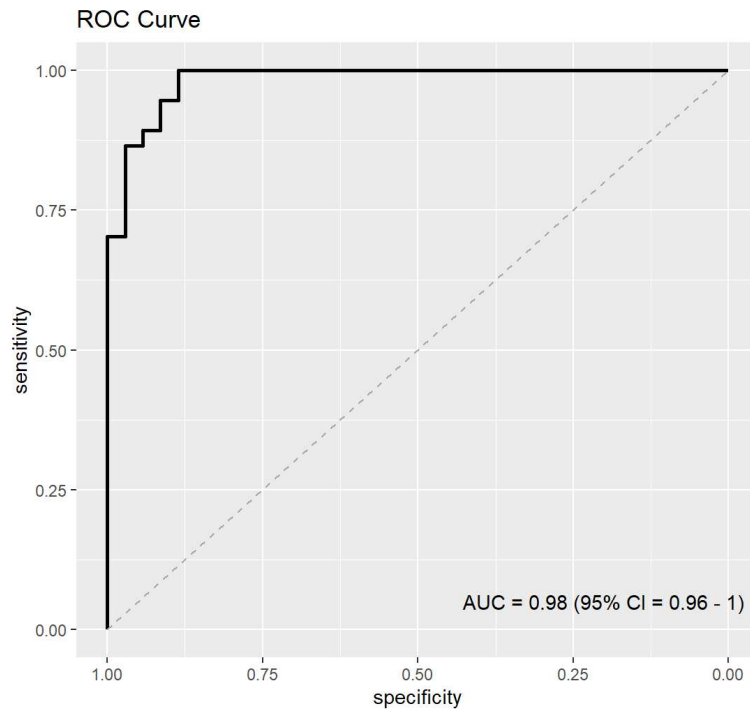


FIGURA 4.3: Curva ROC e AUC del modello logistico stimato.

Bontà di adattamento del modello

Per valutare la bontà del modello si costruisce la tabella di corretta classificazione, attribuendo le unità statistiche al gruppo dei malati o al gruppo dei sani, a seconda che la probabilità stimata sia, rispettivamente, superiore o inferiore a 0.5 (Tabella 4.2).

TABELLA 4.2: Tabella di corretta classificazione per il modello logistico stimato.

Osservati\Previsti	Sani	Malati
Sani	32	3
Malati	4	33

Si osserva che vi sono 65 soggetti classificati correttamente e soltanto 3 falsi positivi e 4 falsi negativi; l'accuratezza è pari al 90%, la sensibilità al 89% e la specificità al 91%.

Per valutare la capacità predittiva del modello si utilizzano la curva ROC e la sua misura di sintesi, AUC, precedentemente definite nella Sezione 3.1 (Figura 4.3). Il modello stimato risulta avere un'ottima capacità di classificazione ($AUC = 0.98$).

Infine, si conduce il test di Hosmer-Lemeshow (Hosmer and Lemeshow, 1989), il quale valuta l'adattamento tra frequenze attese e stimate. Il valore osservato della statistica test è $\chi^2 = 6.94$ con relativo p -value = 0.543, che indica un adattamento soddisfacente.

4.2 Metodi alternativi per la selezione delle variabili

Si valutano ora due metodi alternativi per la selezione delle variabili nel modello logistico, particolarmente utili quando vi è un numero elevato di variabili e un campione con numerosità ridotta: la selezione *stepwise* basata sull'AIC e il metodo *Lasso*.

4.2.1 Selezione *stepwise* basata sull'AIC

Questo metodo combina un approccio iterativo di inclusione ed esclusione di variabili per trovare il modello che ottimizza il criterio di informazione di Akaike (AIC).

L'AIC (*Akaike Information Criterion*) è una misura della qualità di un modello statistico che prende in considerazione sia la bontà di adattamento sia la complessità del modello. L'AIC è definito come:

$$AIC = 2k - 2l(\hat{\theta}^{(k)}; y),$$

dove k è il numero di parametri stimati nel modello e $l(\hat{\theta}^{(k)}; y)$ è la log-verosimiglianza massima (Salvan *et al.*, 2020). L'AIC penalizza i modelli con più parametri, quindi bilancia la bontà di adattamento con la semplicità del modello. Un valore di AIC più basso indica un modello migliore.

La selezione *stepwise* bidirezionale combina elementi delle selezioni *forward* e *backward*. Si parte da un modello, che può essere il modello con la sola intercetta oppure quello con tutte le variabili, e a ogni passo si considera sia l'aggiunta che la rimozione di variabili, scegliendo l'operazione che porta alla maggiore riduzione dell'AIC.

Il processo continua fino a quando non si può più diminuire l'AIC né aggiungendo né rimuovendo variabili.

Modello stimato

Per la stima del modello con selezione *stepwise* si è utilizzata la funzione `stepAIC()` della libreria `MASS`.

Partendo dal modello con la sola intercetta, che presenta $AIC = 101.76$, la prima variabile che viene inserita nel modello è $\log(UPC)$. La sua aggiunta porta a una notevole diminuzione dell'AIC, che si riduce a $AIC = 49.82$. Successivamente entrano nel modello, nell'ordine, le variabili $SDMA$ ($AIC = 43.12$), RR ($AIC = 40.19$) e BUN ($AIC = 37.98$).

Si è così raggiunto il modello stimato precedentemente con l'approccio *forward*. Secondo il principio del minimo AIC, dovrebbero essere inserite nel modello anche le variabili *SEX* ($AIC = 37.86$), *MED_U_NEF* ($AIC = 35.2$) e *HR* ($AIC = 33.45$). L'iterazione successiva prevedrebbe l'ingresso della variabile *AGE*, ma l'algoritmo non raggiunge la convergenza; termina quindi il processo di selezione.

Il modello così stimato, sebbene abbia un valore di AIC ridotto, presenta problemi di affidabilità, poiché vi sono numerose probabilità stimate pari a 1, e viene quindi scartato.

4.2.2 Metodo Lasso

Il *Lasso* (*Least Absolute Shrinkage and Selection Operator*), proposto da Tibshirani (1996), è un metodo di regolarizzazione che “contrae” (*shrink*) i coefficienti β verso lo zero. Il Lasso risolve il problema di minimo

$$\min_{\beta} (y - X\beta)^T (y - X\beta)$$

soggetto al vincolo $\sum |\beta_j| \leq t$, dove t è il parametro di regolarizzazione, o equivalentemente, in forma di Lagrange

$$\min_{\beta} (y - X\beta)^T (y - X\beta) + \lambda |\beta|.$$

Viene quindi introdotta una penalità $\lambda \sum_j |\beta_j|$ nel processo di stima dei coefficienti, dove λ è il parametro di regolarizzazione e β_j sono i coefficienti delle variabili indipendenti. Questo metodo ha l'effetto di ridurre alcuni coefficienti a zero, selezionando automaticamente un sottoinsieme di variabili.

Il parametro t , o analogamente λ , deve essere scelto in modo da bilanciare varianza e distorsione. Il metodo più comune per la scelta del valore ottimale di λ è la validazione incrociata (*Cross-Validation*) (Hastie *et al.*, 2001).

Si tratta di una tecnica di validazione che suddivide il dataset in diversi sottoinsiemi. Il modello viene addestrato su alcuni di questi sottoinsiemi e testato sugli altri, ripetendo il processo più volte. Il valore di λ che minimizza l'errore di previsione medio sui dati di test viene selezionato come ottimale.

Interpretazione dei coefficienti

Un coefficiente pari a zero indica che la variabile associata è stata esclusa dal modello. Questo implica che, data la presenza di altre variabili, questa particolare variabile non fornisce un contributo alla predizione della variabile dipendente.

Un coefficiente diverso da zero indica che la variabile associata è inclusa nel modello. Il segno del coefficiente (positivo o negativo) indica la direzione dell'effetto della variabile indipendente sulla probabilità di osservare l'evento di interesse e il valore del coefficiente indica la forza dell'effetto.

Tuttavia, a causa della penalizzazione, i coefficienti nel modello Lasso tendono ad essere più piccoli rispetto a quelli di un modello di regressione logistica non penalizzato.

Modello stimato

Si è applicato ai dati il metodo Lasso, tramite l'utilizzo della libreria di R `glmnet`. La validazione incrociata seleziona come valore ottimale per il parametro di regolarizzazione $\lambda = 0.035$ e in Tabella 4.3 si riportano i coefficienti stimati. I coefficienti relativi alle variabili non riportate in tabella sono stati stimati pari a zero e vengono, quindi, esclusi dal modello.

TABELLA 4.3: Coefficienti del modello Lasso.

Variabile	Stima
Intercetta	-1.350
HR	0.005
RR	0.017
MED_U_POD	-0.008
log(UPC)	1.507
SDMA	0.163

Allo scopo di confrontare il modello logistico stimato nella Sezione 4.1 con quello selezionato nella Tabella 4.3, si stima il relativo modello logistico con le variabili selezionate dal Lasso come covariate; il suo adattamento è riportato in Tabella 4.4.

I coefficienti relativi a *HR*, *RR* e *MED_U_POD* non sono significativi al livello 5% e il valore dell'AIC è pari a 42.32. L'ipotesi di omogeneità contro il modello corrente è rifiutata ($W_{o|c}^{oss} = 69.43$, $p - value < 0.001$) e il modello corrente contro il modello saturo risulta soddisfacente ($W_{c|s}^{oss} = 30.32 < 66$ *g.d.l.*).

TABELLA 4.4: Adattamento del modello logistico con le variabili selezionate dal Lasso.

Variabile	Stima	Std. Error	z-value	p-value
Intercetta	-0.483	3.721	-0.130	0.897
HR	0.025	0.022	1.101	0.271
RR	0.031	0.020	1.502	0.133
MED_U_POD	-0.077	0.092	-0.835	0.403
log(UPC)	4.246	1.502	2.827	0.005
SDMA	0.266	0.111	2.400	0.016

Il coefficiente stimato relativo alla variabile HR è $\hat{\beta}_1 = 0.025$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_1} = 1.03$; esso indica che per ogni aumento unitario di HR la quota di malattia aumenta del 3%, a parità delle altre variabili.

Il coefficiente stimato relativo alla variabile RR è $\hat{\beta}_2 = 0.031$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_3} = 1.03$; esso indica che per ogni aumento unitario di RR la quota di malattia aumenta del 3%, a parità delle altre variabili.

Il coefficiente stimato relativo alla variabile MED_U_POD è $\hat{\beta}_3 = -0.077$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_4} = 0.93$; esso indica che per ogni aumento unitario di MED_U_POD la quota di malattia diminuisce del 7%, a parità delle altre variabili.

Il coefficiente stimato relativo a $\log(UPC)$ è $\hat{\beta}_4 = 4.246$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_4} = 69.83$; esso indica che per ogni aumento unitario del logaritmo di UPC , la quota di malattia aumenta di quasi 70 volte, a parità delle altre variabili.

Infine, il coefficiente stimato relativo alla variabile $SDMA$ è $\hat{\beta}_5 = 0.266$, per cui il relativo *odds ratio* è $e^{\hat{\beta}_5} = 1.30$; esso indica che per ogni aumento unitario di $SDMA$ la quota di malattia aumenta del 30%, a parità delle altre variabili.

Per valutare la bontà di adattamento del modello si costruiscono la tabella di corretta classificazione (Tabella 4.5) e la curva ROC (Figura 4.4).

TABELLA 4.5: Tabella di corretta classificazione per il modello logistico stimato.

Osservati\Previsti	Sani	Malati
Sani	33	2
Malati	4	33

Si osserva che vi sono 66 soggetti classificati correttamente e soltanto 2 falsi positivi e 4 falsi negativi; l'accuratezza è pari al 92%, la sensibilità al 89% e la specificità al 94%.

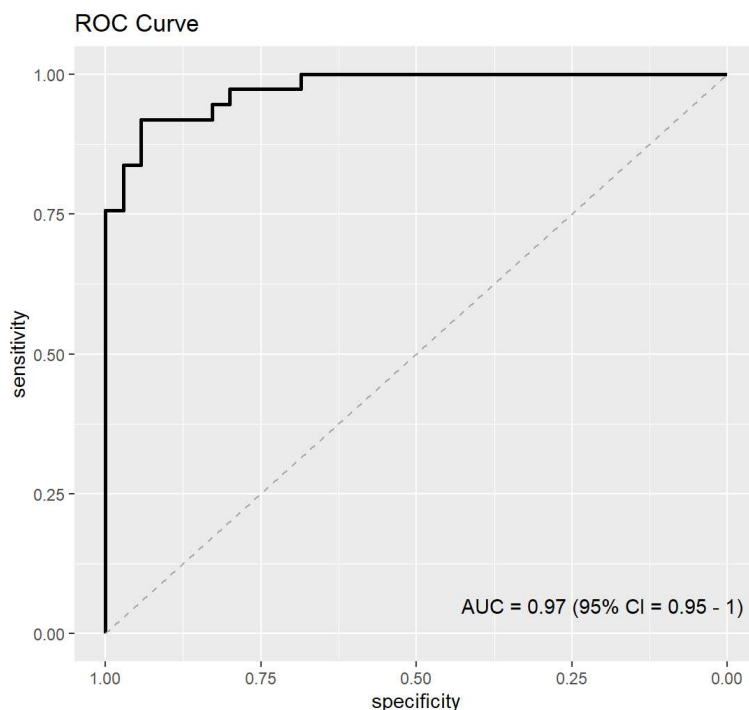


FIGURA 4.4: Curva ROC e AUC del modello logistico stimato.

Il modello stimato presenta $AUC = 0.97$ e risulta, quindi, avere un'ottima capacità di classificazione.

Il valore osservato della statistica test del test di Hosmer-Lemeshow è $\chi^2 = 2.19$ con relativo $p - value = 0.975$, che indica un adattamento soddisfacente.

4.3 Considerazioni finali

Tra i metodi considerati per la selezione delle variabili, quello che sembra produrre il miglior modello stimato è l'approccio *forward* basato sul test del log-rapporto di verosimiglianza.

Tale modello presenta tutti i coefficienti significativi al livello 5% e $AIC = 37.98$; l'accuratezza è pari al 90%, la curva ROC indica un'ottima capacità di classificazione ($AUC = 0.98$) e il test di Hosmer-Lemeshow mostra un adattamento soddisfacente tra frequenze attese e stimate.

I coefficienti stimati indicano un'elevata influenza della variabile *UPC* nell'identificare lo status del paziente; nello specifico, un incremento del valore di tale variabile porta ad un aumento notevole della quota di malattia. Un importante contributo è dato anche da *SDMA*, *RR* e *BUN*.

Il modello stimato tramite la selezione *stepwise* basata sull'AIC, sebbene presenti un valore di AIC inferiore al precedente, non può ritenersi affidabile poiché vi sono numerose probabilità stimate pari a 1.

Infine, il modello con le variabili selezionate dal Lasso presenta un buon adattamento in termini di accuratezza (92%), capacità di classificazione ($AUC = 0.97$) e confronto tra frequenze attese e stimate. Tuttavia, sono presenti tre coefficienti non significativi relativi alle variabili *HR*, *RR* e *MED_U_POD* e l'AIC ha un valore più elevato, pari a 42.32.

Capitolo 5

Risultati e conclusioni

La leishmaniosi è una malattia cronica che colpisce principalmente il cane e raramente anche l'uomo. Si trasmette attraverso la puntura dei flebotomi e, se non curata, porta spesso alla morte del cane.

I sintomi possono comparire in tempi diversi e presentarsi in combinazioni differenti a seconda dei soggetti. Esiste una cura per la leishmaniosi e le probabilità di successo sono tanto maggiori quanto più il cane viene curato precocemente.

Di conseguenza, è di primaria importanza poter far ricorso a test diagnostici che stabiliscano in tempi brevi la presenza o meno della malattia nel soggetto.

In questa relazione sono state valutate la performance e l'accuratezza di una serie di test diagnostici, al fine di individuare quelli che discriminano meglio tra sani e malati.

In primo luogo si sono svolte le analisi esplorative, univariate e bivariate, al fine di osservare la natura delle variabili presenti nel dataset e verificare la presenza di associazioni tra esse. È risultato che la quasi totalità delle variabili quantitative si discosta dalla distribuzione normale.

Si sono osservate differenze significative tra il gruppo dei sani e il gruppo dei malati per le variabili *HR*, *RR*, *SBP*, *MED_U_POD*, *CREA_U*, *MED_U_NEF*, *UPC*, *USG*, *CREA_S* e *SDMA*. In particolare, la mediana è risultata minore nel gruppo dei sani rispetto a quello dei malati per le variabili riferite alla frequenza cardiaca, alla frequenza respiratoria, alla pressione sistolica, al rapporto tra proteine urinarie e creatinina urinaria e alla dimetilarginina simmetrica.

Per queste variabili si è notato, infatti, un andamento crescente all'aumentare della gravità della malattia, codificata dalle variabili *Leishvet_2* e *IRIS_2*.

Al contrario, la mediana è risultata maggiore nel gruppo dei sani rispetto a quello dei malati per le variabili riferite alla podocina, alla creatinina nelle urine, alla nefrina

e alla creatinina nel sangue, e la media è risultata maggiore nei sani per la variabile riferita al peso specifico.

Per queste variabili si è notato, infatti, un andamento decrescente all'aumentare della gravità della malattia, codificata dalle variabili *Leishvet_2* e *IRIS_2*, ad esclusione della creatinina nel sangue che decresce nel gruppo I ma è notevolmente più alta nel gruppo II+III+IV.

Infine, si è osservato che le correlazioni positive più alte si hanno tra la creatinina urinaria e il peso specifico *USG*, tra la podocina e il peso specifico *USG*, tra la podocina e la creatinina urinaria e tra il rapporto nefrina/creatinina urinaria e il rapporto podocina/creatinina urinaria. Le correlazioni negative più alte si hanno tra la creatinina urinaria e il rapporto nefrina/creatinina urinaria, tra la creatinina urinaria e il rapporto proteine urinarie/creatinina urinaria e tra la podocina e la dimetilarginina simmetrica.

Successivamente, lo studio è proseguito con l'analisi delle curve e superfici ROC, al fine di valutare la performance dei test diagnostici considerati. Le variabili che discriminano meglio tra sani e malati sono risultate essere, nell'ordine:

- *UPC*, ovvero il rapporto tra proteine urinarie e creatinina urinaria, con $AUC = 0.92$ e valore ottimale di *cut-off* $k = 0.25$;
- *SDMA*, ovvero la quantità di dimetilarginina simmetrica, con $AUC = 0.85$ e valore ottimale di *cut-off* $k = 12.8$;
- *MED_U_POD*, ovvero la quantità di podocina, con $AUC = 0.76$ e valore ottimale di *cut-off* $k = 10.45$.

Per la variabile *Leishvet_2*, i test che discriminano meglio tra i tre livelli di gravità della malattia sono risultati essere:

- *UPC*, che presenta $VUS = 0.733$ e valori ottimali di *cut-off* $t_- = 0.2$ e $t_+ = 0.5$;
- *SDMA*, che presenta $VUS = 0.501$ e valori ottimali di *cut-off* $t_- = 8.4$ e $t_+ = 18.0$;
- *MED_U_POD*, che presenta $VUS = 0.444$ e valori ottimali di *cut-off* $t_- = 10.2$ e $t_+ = 18.5$.

Per la variabile *IRIS_2*, i test che discriminano meglio tra i tre livelli di gravità della malattia sono risultati essere:

- *SDMA*, che presenta $VUS = 0.754$ e valori ottimali di *cut-off* $t_- = 10.1$ e $t_+ = 20.0$;
- *UPC*, che presenta $VUS = 0.747$ e valori ottimali di *cut-off* $t_- = 0.2$ e $t_+ = 7.3$;
- *MED_U_POD*, che presenta $VUS = 0.569$ e valori ottimali di *cut-off* $t_- = 5.0$ e $t_+ = 12.1$;
- *MED_U_NEF*, che presenta $VUS = 0.567$ e valori ottimali di *cut-off* $t_- = 2.3$ e $t_+ = 4.2$.

I test diagnostici sopra riportati sono tutti moderatamente o altamente accurati.

Infine, si è modellato lo status del paziente, determinato dalla variabile *GROUP*, in funzione delle altre variabili, tramite il modello di regressione logistica. Si è applicata una trasformazione logaritmica a *UPC*, poiché presentava un range di valori molto ampio e causava problemi di stima.

Per la selezione delle variabili nel modello si sono utilizzati tre diversi approcci: la selezione *forward* basata sul test del log-rapporto di verosimiglianza, la selezione *stepwise* basata sull'AIC e il metodo Lasso.

Il modello migliore è risultato essere quello stimato tramite l'approccio *forward*. Esso ha come covariate le variabili $\log(UPC)$, *SDMA*, *RR* e *BUN*; tutti i coefficienti stimati sono significativi al livello 5% , l'AIC è pari a 37.98 e il modello presenta un buon adattamento (accuratezza = 90%, $AUC = 0.98$).

Il contributo maggiore nell'identificare lo status del paziente è dato da *UPC*; in particolare, un incremento di tale variabile porta ad un aumento notevole della quota di malattia.

Tramite il presente studio si è quindi stabilito che i migliori test diagnostici per identificare i soggetti affetti da Leishmaniosi canina, e i relativi livelli di gravità della malattia, sono il rapporto tra proteine urinarie e creatinina urinaria, la quantità di dimetilarginina simmetrica e la quantità di podocina.

Per ottenere una classificazione ancora più accurata è possibile combinare i risultati di quattro test diagnostici, ovvero il rapporto tra proteine urinarie e creatinina urinaria, la quantità di dimetilarginina simmetrica, la frequenza respiratoria e la quantità di urea nel sangue, come indicato dal modello di regressione logistica stimato.

Appendice

Comandi R per l'analisi delle superfici ROC

Per svolgere l'analisi delle superfici ROC si sono utilizzati i pacchetti R `DiagTest3Grp` (Luo and Xiong, 2012) e `bcROCsurface` (To Duc, 2017).

Il pacchetto `DiagTest3Grp` permette di stimare, sia in modo parametrico che non-parametrico, le due misure di sintesi dell'accuratezza per test diagnostici tricotomici, ovvero il VUS e l'indice di Youden, insieme ai relativi valori ottimali di *cut-off*. Nel presente studio si sono applicati i metodi non-parametrici.

Per calcolare il VUS si è utilizzata la funzione

$$\text{VUS}(x, y, z, \text{method}=\text{"NonPar"}),$$

dove x, y, z indicano le misurazioni del test diagnostico analizzato nei tre gruppi D^- , D^0 , D^+ , rispettivamente. La specificazione di `method="NonPar"` permette la stima non-parametrica del VUS, mentre `method="Normal"` la stima parametrica sotto assunzioni di normalità.

Per l'indice di Youden, invece, si utilizza

$$\text{Youden3Grp}(x, y, z, \text{method}=\text{"EMP"}),$$

dove x, y, z come sopra e `method="EMP"` indica di utilizzare le funzioni di ripartizione empiriche nella stima. Altri metodi possibili sono `"Normal"` per la stima parametrica, `"TN"` che applica la trasformazione di Box-Cox, `"KS"` per la stima con densità *kernel* e `"KS-SJ"` per la stima con densità *kernel* applicando il metodo *plug-in* di Sheather-Jones.

L'output di queste due funzioni fornisce la stima con il relativo intervallo di confidenza di livello 95%, i valori ottimali di *cut-off* e le probabilità di corretta classificazione per i tre gruppi per tali valori.

Per la rappresentazione tridimensionale delle superfici ROC si è utilizzato il pacchetto `bcROCsurface`, realizzato per correggere le stime delle superfici ROC in caso di *verification bias*.

In primo luogo si applica la funzione `preDATA()`, che stabilisce se il pacchetto è utilizzabile; essa prende come argomenti la variabile che distingue nei tre gruppi e la variabile che indica le misurazioni del test diagnostico.

Per il grafico si utilizza la funzione

```
ROCs(method="full", T, Dvec, cps),
```

dove `method="full"` indica l'assenza di *verification bias*, `T` rappresenta il test diagnostico, `Dvec` la matrice per i gruppi creata da `preDATA()` e `cps` i valori di *cut-off* individuati precedentemente.

Bibliografia

- Capelli, G., Mantovani, C., and Contalbrigo, L. (2020). *Leishmaniosi canina. Come evitarla, come difenderci*. Istituto Zooprofilattico Sperimentale delle Venezie.
- Dreiseitl, S., Ohno-Machado, L., and Binder, M. (2000). Comparing Three-class Diagnostic Tests by Three-way ROC Analysis. *Medical Decision Making*, 20:323–331.
- Gruppo di studio sulla Leishmaniosi canina, G. S. L. C. (2007). Leishmaniosi canina: Approccio diagnostico e classificazione del paziente leishmaniotico e gestione del paziente proteinurico. Parte I. *Veterinaria*.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley.
- Li, J. and Zhou, X. H. (2009). Nonparametric and Semiparametric Estimation of the Three-Way Receiver Operating Characteristic Surface. *Journal of Statistical Planning and Inference*, 139:4133–4142.
- Luo, J. and Xiong, C. (2012). Diagtest3grp: An R package for analyzing diagnostic tests with three ordinal groups. *Journal of Statistical Software*, 53:1–24.
- Luo, J. and Xiong, C. (2013). Youden index and associated cut-points for three ordinal diagnostic groups. *Communications in Statistics — Simulation and Computation*, 42:1213–1234.
- Mossman, D. (1999). Three-way ROCs. *Medical Decision Making*, 19:78–89.
- Nakas, C. T. (2014). Developments in ROC Surface Analysis and Assessment of Diagnostic Markers in Three-Class Classification Problems. *REVSTAT – Statistical Journal*, 12:43–65.
- Nakas, C. T. and Yiannoutsos, C. T. (2004). Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*, 23:3437–3449.

- Salvan, A., Sartori, N., and Pace, L. (2020). *Modelli Lineari Generalizzati*. Springer.
- Swets, J. A. (1998). Measuring the accuracy of diagnostic systems. *Science*, 240:1285–1293.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58:267–288.
- To Duc, K. (2017). bcROCSurface: An R package for correcting verification bias in estimation of the ROC surface and its volume for continuous diagnostic tests. *BMC Bioinformatics*, 18:503.
- Ventura, L. and Racugno, W. (2017). *Biostatistica. Casi di studio in R*. Egea.
- Xiong, C., van Belle, G., Miller, J. P., and Morris, J. C. (2006). Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*, 25:1251–1273.