

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in

Statistica per l'Economia e l'Impresa



RELAZIONE FINALE

**Test di normalità basati sulla funzione di verosimiglianza empirica**

Relatore Professor Gianfranco Adimari

Dipartimento di Scienze Statistiche

Laureando: Andrea Savio  
Matricola N 1148877

Anno Accademico 2018/2019



## SOMMARIO

Introduzione .....	3
La verosimiglianza empirica .....	5
1.1 Definizione di verosimiglianza empirica.....	5
1.2 Inferenza con la verosimiglianza empirica .....	7
Alcuni test di normalità .....	8
2.1 Test basati sui momenti.....	8
2.1.1 Test di Jarque-Bera.....	9
2.1.2 Test di D'Agostino e Pearson .....	10
2.2 Test di Shapiro-Wilk .....	10
2.3 Test basati sulla funzione di distribuzione empirica.....	11
2.3.1 Test di Kolmogorov-Smirnov .....	12
2.3.2 Test di Lilliefors.....	13
2.3.3 Test di Cramer Von Mises .....	13
2.3.4 Test di Anderson Darling .....	14
I test $ELR_z$ e SEELR.....	16
3.1 Teoria dei test SEELR e $ELR_z$ .....	16
3.1.1 Teoria del test SEELR .....	17
3.1.2 Teoria del test $ELR_z$ .....	19
3.2 Sviluppo dei test SEELR e $ELR_z$ .....	20
3.3 Risultati e valutazioni sul test $ELR_z$ .....	23
3.4 Aspetti computazionali .....	29
3.5 Commenti e conclusioni .....	30
Applicazione dei test SEELR e $ELR_z$ ad un caso di studio reale .....	31
4.1 La curva ROC .....	31
4.2 Introduzione al caso di studio .....	32
4.3 Applicazione dei due test tramite il software R .....	33
4.4 Conclusioni e commenti .....	37
Appendice.....	40
A.1 Valori tabulati del test SEELR .....	40
A.2 Comandi R .....	42

## INTRODUZIONE

Uno degli strumenti più importanti nell'ambito parametrico è la funzione di verosimiglianza, nota per il suo utilizzo nel risolvere problemi di stima o di verifica di ipotesi. I metodi di verosimiglianza sono davvero importanti. Possono essere applicati per trovare stimatori efficienti e per costruire test e intervalli o regioni di confidenza con buone proprietà. La verosimiglianza è anche molto flessibile, infatti si può notare che quando i dati non sono stati completamente osservati, o ci sono stati errori nel campionamento, i metodi di verosimiglianza possono essere usati per compensare o correggere questi problemi. Una condizione fondamentale nei metodi di verosimiglianza parametrica è che la distribuzione comune di tutti i dati disponibili sia conosciuta a priori. Nel caso in cui i dati non provenissero dalla famiglia parametrica fissata, l'applicazione di questi metodi potrebbe essere problematica e causare, per esempio, anche nei casi più gravi, la non consistenza degli stimatori basati sulla verosimiglianza.

Per evitare di dover specificare una famiglia parametrica per i dati, diversi statistici si sono rivolti verso il mondo dell'inferenza non parametrica. I metodi non parametrici consentono di costruire intervalli di confidenza e test validi che non dipendono dall'assunzione sulla distribuzione.

Un metodo di inferenza non parametrica è la verosimiglianza empirica, definita da *Owen*<sup>[1],[2]</sup>, che fornisce lo spunto dal quale diversi studiosi sono partiti per affrontare un altro argomento importante nella teoria e nella pratica statistica: il test di normalità. Nei metodi di inferenza classica, diverse procedure e applicazioni statistiche hanno come assunzione un modello normale o in termini di modello per i dati osservati o come modello per la distribuzione degli errori. Perciò, prima di applicare i metodi statistici basati sull'assunzione di normalità, è necessario compiere un test per verificare tale assunzione. In letteratura si trovano diversi test per verificare la normalità distributiva, ma l'obiettivo di questo studio è quello di discutere lo sviluppo e le potenzialità di due particolari test, *SEELR* e *ELR<sub>Z</sub>*, che sono stati sviluppati sulla base della metodologia della verosimiglianza empirica, e il loro confronto con i test di normalità esistenti.

Sarà presentata anche un'applicazione ad un caso di studio reale per poter valutare le loro proprietà. Il dataset, su cui il lavoro viene svolto, proviene da uno studio riguardante la malattia coronarica (coronaropatia), condotto su 141 pazienti, del dipartimento di cardiologia del "Teaching Hospital" in Galizia, comunità autonoma nel nord-ovest della Spagna. Andremo a valutare il comportamento dei due test SEELR e  $ELR_Z$  attraverso una verifica di ipotesi di normalità relativa alla distribuzione di un biomarcatore basandoci su alcuni campioni osservati e su alcuni campioni di dati trasformati tramite una trasformazione di Box-Cox. Confronteremo i due test con il test di Shapiro-Wilk, che viene preso come riferimento.

## CAPITOLO 1

### LA VEROSIMIGLIANZA EMPIRICA

La verosimiglianza empirica è un metodo non parametrico utilizzato nell'inferenza statistica che permette l'analisi dei dati attraverso i metodi di tipo verosimiglianza, senza dover assumere che i dati appartengono ad una qualche famiglia di distribuzioni. Tra gli anni 1988 e 1990 Owen<sup>[1],[2]</sup> presenta alcuni lavori in cui utilizza la verosimiglianza empirica come uno strumento alternativo ai famosi metodi bootstrap per costruire intervalli di confidenza. Una novità della nuova metodologia consiste nel non dover imporre una forma predeterminata all'intervallo di confidenza, in quanto la forma rifletterà in maniera naturale i dati impiegati, donando più peso alle informazioni che attribuiscono maggiore "verosimiglianza" al parametro di interesse.

Il vantaggio della verosimiglianza empirica è dunque quello di combinare l'affidabilità dei metodi non parametrici con la flessibilità e l'efficacia dell'approccio basato sulla verosimiglianza. Il nome "verosimiglianza empirica" è stato adottato perché la distribuzione empirica dei dati riveste un ruolo centrale.

#### 1.1 Definizione di verosimiglianza empirica

Nei suoi elaborati, Owen parte dalla definizione di funzione di distribuzione empirica quale stima di massima verosimiglianza non parametrica della distribuzione  $F$  dalla quale il campione è estratto, per poi costruire un rapporto di massima verosimiglianza. Per una variabile casuale  $X \in \mathbb{R}$ , la funzione di distribuzione cumulata (CDF) è definita come:  $F(x) = \Pr ( X \leq x )$  per  $-\infty < x < \infty$ . Per denotare  $\Pr ( X < x )$  utilizziamo  $F(x-)$  e, quindi, possiamo definire  $\Pr ( X = x ) = F(x) - F(x-)$ . La notazione  $1_{A(x)}$  è usata per la funzione indicatrice, che assume valore pari a 1 se l'asserzione  $A(x)$  è verificata, e a 0

altrimenti. Allora, dato il campione  $X_1, X_2, \dots, X_n$ , si definisce la funzione di distribuzione empirica (ECDF) di  $X_1, X_2, \dots, X_n$  come:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x} \quad \text{per } -\infty < x < \infty.$$

Dati  $X_1, X_2, \dots, X_n$ , assunti indipendenti, con funzione di distribuzione cumulata (CDF) comune  $F_0$ , definiamo la verosimiglianza non parametrica per la CDF  $F$  come:

$$L(F) = \prod_{i=1}^n (F(X_i) - F(X_i -)).$$

Il valore  $L(F)$  è la probabilità di prendere esattamente il campione di valori osservati  $X_1, X_2, \dots, X_n$  dalla CFD  $F$  nel caso discreto. Una conseguenza è che  $L(F)=0$  se  $F$  è una distribuzione continua. Per avere una verosimiglianza non parametrica positiva, la distribuzione  $F$  deve avere una probabilità positiva su ognuno dei valori dei dati osservati.

Owen decide di associare ad ogni osservazione campionaria la funzione di verosimiglianza empirica. Si considerino  $X_1, X_2, \dots, X_n$  come  $n$  osservazioni indipendenti ed identicamente distribuite con distribuzione  $F$ , caratterizzata dal vettore di probabilità  $p = (p_1, p_2, \dots, p_n)$  che soddisfa i seguenti vincoli:  $p_i \geq 0$  e  $\sum_{i=1}^n p_i = 1$ , e con:  $L(F) = \prod_{i=1}^n p_i$ . Per un dato funzionale d'interesse  $\theta(\bullet)$ , la verosimiglianza non parametrica che Owen chiama verosimiglianza empirica, è dedotta come verosimiglianza profilo. Formalmente, cioè, la funzione di verosimiglianza per  $\theta(\bullet)$ , calcolata in un generico punto  $\theta_1$  è:

$$L_\theta(\theta_1) = \max_{(p: \theta(p) = \theta_1, \sum_{i=1}^n p_i = 1)} \prod_{i=1}^n p_i$$

Tale funzione è massima quando  $p_i = \frac{1}{n}$ , per  $1 \leq i \leq n$ , per cui  $L_\theta(\theta)$  avrà il suo punto di massimo in  $\hat{\theta} = \theta(\hat{F})$  e risulterà allora che  $L_\theta(\hat{\theta}) = n^{-n}$ . Qui  $\hat{F}$  indica la funzione di ripartizione empirica.

## 1.2 Inferenza con la verosimiglianza empirica

Nell'inferenza parametrica, verifiche di ipotesi e regioni di confidenza si basano sul rapporto di verosimiglianza. Se  $L(\eta)$  è molto più piccolo rispetto a  $L(\hat{\eta})$ , l'ipotesi  $\eta_0 = \eta$  viene rifiutata e si esclude  $\eta$  dalla regione di confidenza per  $\eta_0$ . Il teorema di Wilks stabilisce che  $-2\log(L(\eta_0)/L(\hat{\eta}))$  tende ad una distribuzione chi-quadro quando  $n$  tende all'infinito, sotto deboli condizioni di regolarità, consentendoci di decidere quanto piccolo deve essere il valore di  $L(\eta)$  affinché  $\eta$  sia rifiutato. I gradi di libertà per la distribuzione chi-quadro sono, solitamente, pari alla dimensione del parametro di interesse. Una delle più importanti proprietà che la verosimiglianza empirica assorbe dal caso parametrico è l'estensione del teorema di Wilks, perciò per verifiche di ipotesi e per costruire intervalli di confidenza si può anche utilizzare il rapporto di verosimiglianza non parametrica, procedendo in analogia con quanto si fa con la verosimiglianza parametrica. Owen sviluppò una versione non parametrica della statistica rapporto di verosimiglianza e dimostrò, proprio come nel caso parametrico che, se il parametro di interesse è definito mediante equazioni di stima non distorte, tale statistica ha distribuzione limite  $\chi^2$ , e diventa così uno strumento quasi parametrico, permettendone anche l'utilizzo per verifiche d'ipotesi e per costruire regioni di confidenza per parametri  $\theta(F)$  che sono caratteristiche di una funzione  $F$  ignota. D'altra parte però, si osserva che, avendo una natura non parametrica, i metodi che utilizzano la verosimiglianza empirica di solito sono inefficienti quando il campione ha una dimensione molto piccola. Per una distribuzione  $F$ , il rapporto di verosimiglianza empirica viene definito in tale modo:  $R(F) = L(F)/L(F_n)$ , dove  $L(F)$  è la verosimiglianza non parametrica. Per il parametro  $\theta$  invece, il rapporto di verosimiglianza empirica si definisce in tale modo:

$$R_{\theta}(\theta_1) = L_{\theta}(\theta_1)/L_{\theta}(\hat{\theta}) = \sup_{(p:\theta(p)=\theta_1, \sum_{i=1}^n p_i = 1)} \prod_{i=1}^n np_i.$$



## CAPITOLO 2

### ALCUNI TEST DI NORMALITÀ

In molte procedure statistiche come il t-test, i test sui coefficienti di un modello di regressione dei coefficienti, l'analisi della varianza, e il test F per l'omogeneità della varianza, una delle assunzioni fondamentali che permettono lo svolgimento della procedura è quella della normalità dei dati. Per verificare la normalità dei dati esistono diversi test, che possiamo classificare in base a diverse caratteristiche: test basati sulla regressione e correlazione (come il test di Shapiro-Wilk); test basati sulla funzione di distribuzione empirica (Kolmogorov-Smirnov, Lilliefors, Anderson-Darling, Cramer-Von Mises); test basati sui momenti (Jarque-Bera, D'Agostino-Pearson); altri particolari tipi di test. In questa sezione illustreremo dettagliatamente i test indicati.

#### 2.1 Test basati sui momenti

Karl Pearson è riconosciuto come il primo ad aver scoperto che le deviazioni da una distribuzione normale, per la maggior parte, sono caratterizzate dalle differenze dal terzo e dal quarto momento della normale standardizzata. Di conseguenza è stato naturale pensare che un test per verificare la normalità poteva essere formulato comparando i momenti calcolati sul campione con quelli teorici. Tentativi di approssimare le distribuzioni di asimmetria e curtosi empiriche per campioni normali, e di conseguenza la ricerca di valori critici per i test proposti, continuano ancora nei nostri giorni.

I test di normalità basati sui momenti includono sia test che considerano un solo momento come il test di asimmetria ( $S^2$ ) e il test di curtosi ( $K$ ), sia altri test che sono più precisi poiché ne utilizzano più di uno come il test di D'Agostino-Pearson (DP) e il test di Jarque-Bera (JB), che vediamo di seguito.

### 2.1.1 Test di Jarque-Bera (JB)

Il test di Jarque-Bera (1979) è un test statistico per la verifica dell'ipotesi di normalità che si basa sulla misura dell'asimmetria e della curtosi di una distribuzione. Questo test è definito come omnibus test in quanto si avvale di una statistica test che tiene conto contemporaneamente sia dell'indice di curtosi sia dell'indice di asimmetria per la verifica della normalità dei dati.

Una distribuzione normale è caratterizzata da un'asimmetria nulla e da un indice di curtosi pari a tre. Nella realtà, però, la maggior parte delle altre distribuzioni sono caratterizzate da asimmetrie diverse da zero e da curtosi maggiori di 3. Dunque, Jarque e Bera hanno proposto il seguente test statistico:

$$JB = \frac{n}{6} (S^2 + \frac{1}{4} (K - 3)^2)$$

Dove  $n$  è il numero delle osservazioni, mentre  $S^2$  e  $K$  (indici di curtosi e asimmetria) vengono calcolate così:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\hat{\mu}_3}{\hat{\sigma}^2 \frac{3}{2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^3}{(\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2)^{3/2}}$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\hat{\mu}_4}{\hat{\sigma}^2 \frac{3}{2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^4}{(\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2)^2}$$

Si tratta di un test di normalità che verifica, come ipotesi nulla, simultaneamente se l'asimmetria e la curtosi sono coerenti con i valori che dovrebbero assumere sotto l'ipotesi nulla di normalità, ossia rispettivamente 0 e 3. Sotto l'ipotesi nulla  $H_0$  il test si distribuisce asintoticamente come una chi-quadro con 2 gradi di libertà. Tale ipotesi viene rifiutata per valori di JB troppo grandi.

### 2.1.2 Test di D'Agostino e Pearson (DA)

Anche il test di D'Agostino e Pearson (1971) si può definire come un omnibus test in quanto si avvale di una statistica test che tiene conto contemporaneamente sia dell'indice di curtosi sia dell'indice di asimmetria per la verifica della normalità dei dati. La statistica che utilizzano è:

$$K^2 = Z_1(S^2) + Z_2(K)$$

Introducono, rispetto al test JB, delle trasformazioni delle statistiche S e K per accelerare la loro convergenza alle corrispondenti distribuzioni normali.<sup>[3]</sup> Come per la statistica JB, anche  $K^2$  segue approssimativamente una distribuzione chi quadro con due gradi di libertà sotto  $H_0$ . Perciò l'ipotesi nulla verrà rifiutata per valori troppo grandi di  $K^2$ .

### 2.2 Test di Shapiro-Wilk (SW)

Tale test fu introdotto nel 1965 da Shapiro e Wilk, ed è considerato ancora oggi uno dei più potenti test per la verifica dell'ipotesi di normalità, in particolare per campioni di piccole dimensioni ( $n \leq 50$ ). La statistica test utilizzata per verificare tale ipotesi si basa sul confronto di due stimatori alternativi della varianza  $\sigma^2$ , e rientra dunque nella categoria di test che utilizzano regressioni e combinazioni lineari.

Assumiamo  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  come un campione casuale ordinato di dimensione n estratto da una distribuzione normale standard ( $\mu = 0, \sigma^2 = 1$ ), e definiamo  $m' = (m_{(1)} \leq m_{(2)} \leq \dots \leq m_{(n)})$  il vettore dei valori attesi delle statistiche d'ordine e  $V = (v_{ij})$  come la matrice di covarianza delle suddette statistiche.

Inoltre, assumiamo  $y' = (y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)})$  come un vettore di osservazioni casuali ordinate prese da una popolazione arbitraria. Se le  $y_{(i)}$  fossero osservazioni ordinate prese da una distribuzione normale con media  $\mu$  e varianza  $\sigma^2$  sconosciute, allora si potrebbe scrivere le  $y_{(i)}$  come  $y_{(i)} = \mu + \sigma x_{(i)}$  ( $i=1,2,\dots,n$ ). La statistica test per verificare la normalità è:

$$SW = \frac{(\sum_{i=1}^n a_i y_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

dove  $a' = m'V^{-1}(m'V^{-1}V^{-1} m)^{-1/2}$  e  $a'a = 1$ . Gli elementi di  $a$  sono noti; però il limite di questo test è che gli elementi di  $V$  e di conseguenza anche gli  $a_i$  sono perfettamente conoscibili per campioni di massimo 20 unità. Per campioni tra le 21 e 50 unità Shapiro e Wilk forniscono una stima degli  $a_i$ .<sup>[4]</sup>

Osservando la statistica test, al numeratore troviamo uno stimatore non parametrico basato sulla combinazione lineare ottimale della statistica d'ordine di una variabile aleatoria normale e al denominatore il consueto stimatore parametrico, ossia la varianza campionaria. Il valore del test varia tra 0 e 1. La statistica test sotto l'ipotesi nulla segue una distribuzione non nota, per cui sono stati tabulati i valori critici di questa distribuzione.<sup>[2]</sup> Rifiuteremo dunque l'ipotesi di normalità dei dati se la statistica SW ha valori minore del valore critico  $w_\alpha$ . La procedura è basata sulla coda inferiore della distribuzione perciò valori piccoli del test portano a rifiutare l'ipotesi di normalità, mentre un valore molto vicino ad uno supporterà l'ipotesi di normalità dei dati.

### 2.3 Test basati sulla funzione di distribuzione empirica (EDF)

Siano  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  le osservazioni ordinate di un campione casuale di numerosità  $n$ , allora la funzione di distribuzione empirica di un campione, chiamata  $F_n(x)$ , è una funzione del tipo:

$$F_n(x) = \begin{cases} 0 & \text{se } x < x_{(1)} \\ \frac{i}{n} & \text{se } x_{(i)} \leq x < x_{(i+1)}, \text{ con } i = 1, \dots, n-1 \\ 1 & \text{se } x_{(n)} \leq x \end{cases}$$

Ci sono due tipi di test EDF: quelli che si basano sulla massima distanza della funzione di distribuzione empirica da quella nulla (come il test di Kolmogorov-Smirnov) e i test quadratici (come il test di Anderson-Darling).

I test EDF si basano sulla differenza tra la funzioni di distribuzione empirica e la distribuzione sotto l'ipotesi nulla; sono quindi tutti test a una coda. Questi test rifiuteranno l'ipotesi nulla se la discrepanza tra la EDF e la funzione di distribuzione cumulata sotto l'ipotesi nulla è molto ampia.

### 2.3.1 Test di Kolmogorov-Smirnov (KS)

La statistica KS appartiene alla classe delle statistiche EDF e questa classe si basa sulla maggiore differenza verticale tra la funzione di distribuzione empirica e quella ipotetica  $F_{H_0}(x)$ . Questo test richiede che la distribuzione  $F_{H_0}(x)$  sia completamente definita con parametri conosciuti. Per la verifica della normalità, utilizzando questo test,  $F_{H_0}(x)$  viene presa da una distribuzione normale con media  $\mu$  e deviazione standard  $\sigma$  note.

La statistica test è definita in tre maniere differenti a seconda del tipo di ipotesi alternativa che si vuole considerare. L'ipotesi nulla sarà sempre la seguente:  $H_0: F(x)=F_{H_0}(x)$ , mentre le tre ipotesi alternative che si possono considerare sono:

- a)  $H_1: F(x) < F_{H_0}(x)$  , dove la statistica  $KS^+ = \sup[F(x) - F_{H_0}(x)]$
- b)  $H_1: F(x) > F_{H_0}(x)$  , dove la statistica  $KS^- = \sup[F_{H_0}(x) - F(x)]$
- c)  $H_1: F(x) \neq F_{H_0}(x)$  , dove la statistica  $KS = \max(KS^+, KS^-)$

L'ipotesi nulla viene rifiutata se  $\sqrt{n} KS > K_\alpha$ , dove  $K_\alpha$  è il quantile della distribuzione di Kolmogorov, che è determinato da  $\Pr(\sqrt{n} KS > K_\alpha) = 1 - \alpha$ .

I quantili di tale distribuzione sono tabulati.<sup>[5]</sup>

### 2.3.2 Test di Lilliefors(LL)

Il test LL (1967) è una modifica del test KS e viene utilizzato quando è necessario stimare i parametri sconosciuti della distribuzione  $F_{H_0}(x)$  attraverso il campione di dati. Il test LL mette a confronto la distribuzione empirica con una distribuzione normale dove i parametri incogniti vengono stimati dal campione dato. La statistica LL è formulata riprendendo l'ipotesi alternativa bilaterale del test KS, ma, in questo caso, i valori di  $\mu$  e  $\sigma$  che vengono utilizzati sono la media e la varianza campionaria; infatti le statistiche  $KS^+$  e  $KS^-$  vengono cambiate in questo modo:

$$KS^+ = \sup\left[\sum_{i=1}^n \left[\frac{i}{n} - \Phi\left(\frac{x_{(i)} - \hat{\mu}}{\hat{\sigma}}\right)\right]\right]$$

$$KS^- = \sup\left[\sum_{i=1}^n \left[\frac{x_{(i)} - \hat{\mu}}{\hat{\sigma}} - \frac{i-1}{n}\right]\right]$$

Grazie alle stime basate sui dati, sotto l'ipotesi nulla la funzione di ripartizione si avvicina ai dati, per cui la distanza verticale massima è ridotta rispetto a quanto lo sarebbe se l'ipotesi nulla avesse individuato una predefinita distribuzione normale, come nel caso della statistica KS. Dunque, in questo modo la distribuzione della statistica test sotto l'ipotesi nulla è stocasticamente più piccola della distribuzione di Kolmogorov-Smirnov. Non esistono dei valori critici già definiti per questo test; infatti fino ad oggi, le tabelle dei valori critici per questa distribuzione sono state calcolate solo attraverso metodi Monte Carlo.<sup>[6]</sup>

### 2.3.3 Test Cramer-Von Mises (CVM)

Anche il test Cramer-Von Mises (1928-30) fa parte dei test basati sulla funzione di distribuzione empirica, ed è un test alternativo al test di Kolmogorov-Smirnov. Tale test è utilizzato per valutare la bontà di adattamento di una funzione di ripartizione teorica  $F_{H_0}(x)$  ai dati, nel nostro caso una normale, quindi ad una funzione di distribuzione empirica  $F(x)$ .

Tale test può essere anche utilizzato per confrontare due distribuzioni empiriche. La statistica che definisce il test è la seguente:

$$n\omega^2 = n \int_{-\infty}^{\infty} [F(x) - F_{H_0}(x)]^2 \psi(F_{H_0}(x)) dF_{H_0}(x)$$

dove  $\psi(F_{H_0}(x))$  è una funzione di ponderazione. In particolare, quando questa funzione è uguale ad uno ( $\psi(F_{H_0}(x) = 1)$ ), e consideriamo le osservazioni ordinate  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , possiamo vedere la statistica test come:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^n \left( \phi(x_i) - \frac{2i-1}{2n} \right)^2$$

Dunque, l'ipotesi che i dati provengano da una distribuzione  $F_{H_0}(x)$ , che nel nostro caso considereremo una normale, viene rifiutata se il valore che assume la statistica test è maggiore del valore critico. Come nel test LL, anche in questo test, non esistono dei valori critici già definiti e la statistica test ha una distribuzione nulla non nota; perciò i valori critici andranno calcolati tramite simulazione.<sup>[7]</sup>

### 2.3.4 Test di Anderson-Darling

Il test di Anderson-Darling (1952) è sostanzialmente una modifica del test CVM. Anche questo test può essere utilizzato per verificare bontà di adattamento dei dati ad una particolare ad una funzione di ripartizione teorica. Nel nostro caso verrà utilizzato per verificare la normalità dei dati.

La statistica test viene calcolata in maniera molto simile al test CVM:

$$A^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \omega(F^*(x)) dF^*(x)$$

dove rispetto al test CMV cambia la funzione di ponderazione, infatti abbiamo che:

$$\omega(F_{Ho}(x)) = [F_{Ho}(x)(1 - F_{Ho}(x))]^{-1}$$

Considerando le osservazioni ordinate  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  di un campione di  $n$  osservazioni, la statistica può essere calcolata come:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\log(\phi(x_i)) + \log(1 - \phi(x_{n-i+1}))]$$

Questo test, basato sulla coda superiore della distribuzione, ha un problema: a differenza dei due test precedenti, i valori critici devono essere calcolati per ogni specifica distribuzione; perciò il test AD utilizza una specifica distribuzione teorica nel calcolo dei valori critici, risultando più sensibile dei precedenti.<sup>[8]</sup>



## CAPITOLO 3

### I TEST $ELR_Z$ E SEELR

Negli ultimi tempi i ricercatori sono attratti dalla buona combinazione tra l'uso dei test basati sui momenti e la metodologia della verosimiglianza empirica per lo sviluppo di test per la normalità.

Tra i primi, *Dong e Giles*<sup>[9]</sup> propongono un test basato sul rapporto di verosimiglianza empirica (ELR), con l'ipotesi nulla che fissa un singolo elemento della classe normale. Si concentrano sui primi quattro momenti della distribuzione normale, e il loro test risulta avere prestazioni migliori dei comuni test esistenti. In seguito *G. Shan, A. Vexler, G. E. Wilding*<sup>[10]</sup> elaborano un semplice test ELR (il test SEELR) per normalità con ipotesi nulla non semplice, basato sulle relazioni tra i momenti, usando una variabile standardizzata. Il loro test dimostra di essere più potente dei test per la bontà di adattamento che erano conosciuti per campioni di piccole o medie dimensioni per diverse distribuzioni alternative. Ora viene illustrato come C. S. Marange e Y. Qin<sup>[11]</sup>, partendo dall'approccio usato da *Shan et al.*<sup>[10]</sup>, si focalizzano sulla costruzione di un semplice ELR test di normalità ( $ELR_Z$ ), basato sui vincoli dei momenti di una distribuzione mezza-normale.

#### 3.1 Teoria alla base dei test SEELR e $ELR_Z$

Si assume di avere delle variabili casuali non ordinate  $X_1, X_2, \dots, X_n$  indipendenti e identicamente distribuite. L'intento è quello di valutare se i dati osservati sono normalmente distribuiti, perciò viene considerata la seguente ipotesi nulla:  $H_0: X \sim N(\mu, \sigma^2)$ , dove  $\mu$  e  $\sigma^2$  sono parametri ignoti.

I dati di partenza vengono standardizzati secondo due trasformazioni adeguate. La prima trasformazione si avvale dell'utilizzo delle variabili casuali standardizzate della distribuzione normale usando la seguente trasformazione:

$$Z_i^{**} = (X_i - \hat{\mu})/SD, \quad i=1,2,\dots,n \quad (1)$$

dove  $\hat{\mu} = (1/n) \sum_{i=1}^n X_i = \bar{X}$  e SD è la deviazione standard che viene stimata partendo dalla quantità non distorta  $s^2 = S/(n-1)$ . Si può anche decidere di usare la stima di massima verosimiglianza (MLE)  $\hat{\sigma}^2 = S/n$  dove  $S = \sum_{i=1}^n (X_i - \bar{X})^2$ .

E' noto che entrambe le quantità  $s^2$  e  $\hat{\sigma}^2$  convergono a  $\sigma^2$  quando  $n$  tende all'infinito.

La seconda trasformazione che viene usata, proposta nel lavoro di *Lin e Mudholkar*<sup>[12]</sup>.

Le osservazioni saranno dunque trasformate in questo modo:

$$Z_i^* = \frac{\sqrt{n/(n-1)}((X_i - \bar{X}))}{SD_{-i}}, \quad i=1,2,\dots,n \quad (2)$$

dove:  $\bar{X} = \left(\frac{1}{n}\right) \sum_{j=1}^n X_j$ , e  $SD_{-i}^2 = (1/(n-2)) \sum_{j=1, j \neq i}^n (X_j - \bar{X}_{-i})^2$  e  $\bar{X}_{-i} = (1/(n-1)) \sum_{j=1, j \neq i}^n X_j$ .

E' da notare che il numeratore e il denominatore di  $Z_i^*$  sono indipendenti sotto l'ipotesi nulla  $H_0$ . In entrambi gli sviluppi dei due test (SEELR e  $ELR_Z$ ), sono risultate più efficienti le statistiche test costruite con la seconda trasformazione (2), per cui illustreremo in seguito solo i risultati ottenuti utilizzando quest'ultimo tipo di standardizzazione.

A questo punto *C.S. Marange e Y. Qin*<sup>[11]</sup> prendono una strada alternativa rispetto a quella di *Shan et al.*<sup>[10]</sup>, infatti pongono le loro attenzioni sulla variabile mezza-normale. Di seguito vengono riportati gli approcci ai due test.

### 3.1.1 Teoria alla base del test SEELR

Il test SEELR è un test per la normalità per campioni di ampie dimensioni. Infatti, per costruire la statistica test, sono stati utilizzati i momenti della distribuzione t di Student,

considerando che, quando il numero di gradi di libertà cresce, la  $t$  di Student si avvicina sempre di più alla normale standard.

Ora viene illustrata la teoria alla base di questo test.

Se consideriamo  $X_i \sim N(\mu, \sigma^2)$   $i=1,2,\dots,n$ , ne consegue che  $\sqrt{n/(n-1)} ((X_i - \mu))/\sigma \sim N(0,1)$  e, usando il teorema di Cochran, si può affermare che  $(n-2)SD_{-i}^2/\sigma^2 \sim \chi_{n-2}^2$ . Considerando ora la seconda trasformazione definita precedentemente per standardizzare i dati (2), si può notare che le variabili  $Z_1^*, Z_2^*, \dots, Z_n^*$ , grazie alla definizione della distribuzione  $t$  di Student, sotto l'ipotesi nulla, seguono la distribuzione  $t$  con  $(n-2)$  gradi di libertà. Quando la dimensione del campione ( $n$ ) è ampia, le variabili  $Z_1^*, Z_2^*, \dots, Z_n^*$ , sono asintoticamente indipendenti.

La famiglia delle distribuzioni  $t$  di Student è descritta dalla seguente funzione di densità:

$$f(t|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} (1 + \frac{t^2}{\nu})^{-(\nu+1)/2}, \text{ dove } \nu \text{ è il numero di gradi di libertà, e } \Gamma \text{ è la}$$

funzione gamma. La forma della distribuzione  $t$  di Student assomiglia alla forma unimodale della distribuzione normale standard, ossia con media pari a zero e varianza unitaria, anche se con code più pesanti. Quando il numero di gradi di libertà cresce, la  $t$  di Student si avvicina sempre di più alla normale standard. I momenti della distribuzione  $t$  con  $\nu$  gradi di libertà si definiscono come segue:

$$\mu_k = \begin{cases} 0 & k \text{ dispari}, 0 < k < \nu \\ \frac{\Gamma(\frac{k+1}{2})\Gamma(\frac{\nu-k}{2})\nu^{k/2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} & k \text{ pari}, 0 < k < \nu \\ \text{Non esiste} & k \text{ dispari}, \nu \leq k \\ \infty & k \text{ pari}, \nu \leq k \end{cases} \quad (3)$$

Le variabili trasformate,  $Z_i^*$ , mantengono le proprietà descritte dall'equazione (3). Seguendo la metodologia della verosimiglianza empirica (EL), al momento  $k$ -esimo di  $Z^*$ ,  $E(Z^{*k})$ , corrisponde il momento campionario  $\sum_{i=1}^n p_i Z_i^{*k}$ , dove  $Z_i^*$  è realizzazione di una

distribuzione t con  $n - 2$  gradi di libertà e i  $p_i$  sono le componenti della verosimiglianza empirica,  $\prod_{i=1}^n p_i$ , che la massimizzano sotto opportuni vincoli.

### 3.1.2 Teoria del test $ELR_Z$

Il test  $ELR_Z$  è un test per la normalità che utilizza i momenti di una variabile mezza-normale per costruire la statistica test. Nel loro lavoro, C. S. Marange e Y. Qin<sup>[11]</sup>, per sperimentare il loro test utilizzano entrambe le trasformazioni (1) e (2) descritte precedentemente per standardizzare i dati. Attraverso delle simulazioni Monte Carlo dimostrano che il test che usa i dati standardizzati attraverso la (2) risulta essere più efficiente rispetto all'altro. Per cui ora illustreremo solamente la metodologia che viene usata applicando la trasformazione (2) per standardizzare i dati.

Ora viene spiegata la teoria alla base di questo test.

All'aumentare di  $n$ , le variabili standardizzate  $Z_1^*, Z_2^*, \dots, Z_n^*$  diventano asintoticamente indipendenti. Se  $X \sim N(0,1)$ , allora il suo valore assoluto  $|X| \sim HN(0,1)$ . Ne segue che se  $X \sim N(\mu, \sigma^2)$ , allora il modulo della variabile casuale normale standardizzata  $Z^*$  segue una distribuzione di una variabile mezza-normale  $W$  con media uguale a  $\sqrt{2/\pi}$  e momento secondo pari a 1. La variabile casuale standardizzata mezza normale  $W$  ha funzione di densità pari a:

$$f(w) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-(1/2)w^2} & \text{per } w \geq 0 \\ 0 & \text{per } w < 0 \end{cases} \quad \text{e indichiamo ciò con } |Z^*| = W \sim HN(0,1).$$

Seguendo il lavoro di *Prudnikov e al.*<sup>[13]</sup>, il  $k$ -esimo momento della variabile mezza normale standardizzata per diversi valori interi di  $k > 0$  viene descritto dalla proposizione seguente:

**PROPOSIZIONE 3.1.** Data la variabile  $W \sim HN(0,1)$ , per  $k=1,2,\dots,n$ ; allora i  $k$  momenti sono

dati da: 
$$E(W^k) = \mu_k = \frac{1}{\sqrt{\pi}} 2^{\frac{k}{2}} \Gamma\left(\frac{k+1}{2}\right)$$

Usando questa formula si derivano i primi quattro momenti. Questi sono facilmente ottenuti come segue:

COROLLARIO 3.1. Data  $W \sim \text{HN}(0,1)$ . I primi due momenti di  $W$ , sono:

$$E(W) = \mu_1 = \sqrt{\frac{2}{\pi}} \Gamma(1) = \sqrt{\frac{2}{\pi}} \approx 0.7979 \quad ; \quad E(W^2) = \mu_2 = \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = 1$$

COROLLARIO 3.2. Data  $W \sim \text{HN}(0,1)$ . I coefficienti di asimmetria e di curtosi di  $W$  sono dati da:

$$E(W^3) = \mu_3 = \sqrt{\frac{2^3}{\pi}} \Gamma(2) = 2 = \sqrt{\frac{2}{\pi}} \approx 1.5958$$

$$E(W^4) = \mu_4 = \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right) = 3$$

### 3.2 Sviluppo dei test SEELR e ELR<sub>Z</sub>

Si prendano in considerazione le osservazioni non ordinate  $X_1, X_2, \dots, X_n$  indipendenti e identicamente distribuite, con  $\mu$  e  $\sigma^2$  ignoti. L'intenzione è di ottenere un test per verificare l'ipotesi di avere le  $X_1, X_2, \dots, X_n$  da distribuzione normale. Ora, si prendano in esame le variabili casuali  $Z_1^*, Z_2^*, \dots, Z_n^*$  che sono le variabili normali standardizzate delle variabili  $X_1, X_2, \dots, X_n$  di partenza. Per il test SEELR vengono utilizzate direttamente le variabili  $Z_1^*, Z_2^*, \dots, Z_n^*$ , mentre per il test ELR<sub>Z</sub> vengono considerate le variabili  $W_1, W_2, \dots, W_n$  definite precedentemente. Per semplicità di notazione, di seguito, verrà analizzato lo sviluppo del test SEELR, tenendo conto che lo stesso procedimento vale per il test ELR<sub>Z</sub>. Lavoreremo quindi sulle variabili  $Z_i^*$ , ma procedura analoga viene fatta per le variabili  $W$ .

Seguendo il metodo EL, vengono assegnati dei pesi  $p_i$ , ossia delle probabilità ignote per ciascuna osservazione trasformata  $Z_i^*$  e poi viene formulata la funzione EL, che è data da :

$L(F) = \prod_{i=1}^n p_i$  ; dove i  $p_i$  soddisfano la usuale proprietà delle probabilità, ovvero che:

$$0 \leq p_i \leq 1 \quad \text{e} \quad \sum_{i=1}^n p_i = 1 .$$

I parametri di probabilità sono scelti attraverso equazioni di stima non distorte basate sui momenti, in modo da massimizzare la funzione EL. In particolare,  $\hat{E}(Z^{*k})$  è il momento campionario  $\sum_{i=1}^n p_i Z_i^{*k}$  dove i pesi  $p_i$  sono gli elementi della funzione EL. Sotto  $H_0$ , le equazioni basate sui momenti empirici hanno tale forma:  $\sum_{i=1}^n p_i Z_i^{*k} - \mu_k = 0, k=1,2,\dots$  . Si considera un momento per volta.

Il sistema di ipotesi per i test SEELR si può scrivere (con riferimento al momento k-esimo) come:

$$H_0: E(Z^{*k}) = \mu_k$$

$$H_1: E(Z^{*k}) \neq \mu_k$$

Per il test ELR<sub>Z</sub>, il sistema di ipotesi è analogo utilizzando le equazioni basate sui momenti empirici della variabile  $W^k$ .

Sotto  $H_0$ , la funzione EL è massimizzata rispetto ai  $p_i$  sotto i vincoli:  $\sum_{i=1}^n p_i = 1$  e  $\sum_{i=1}^n p_i Z_i^{*k} = \mu_k$ . Più precisamente i pesi  $p_i$  sono indentificati da:  $p_1, p_2, \dots, p_n: \arg. \sup \prod_{i=1}^n a_i \mid \sum_{i=1}^n a_i = 1$

$$\sum_{i=1}^n a_i Z_i^{*k} = \mu_k$$

dove  $0 \leq a_j \leq 1$  con  $j=1,2,\dots,n$ . Se poi viene usata la tecnica dei moltiplicatori di Lagrange, possiamo mostrare che la funzione di EL massima sotto  $H_0$  può essere espressa dalla

$$\text{seguente formula: } L(F_{H_0}) = L(Z_1^*, Z_2^*, \dots, Z_n^* | \mu_k) = \prod_{i=1}^n \frac{1}{n(1+\lambda_k(Z_i^{*k} - \mu_k))}$$

dove  $\lambda_k$  è una radice dell'equazione  $\sum_{i=1}^n \frac{Z_i^{*k} - \mu_k}{1+\lambda_k(Z_i^{*k} - \mu_k)} = 0$ .

Sotto l'ipotesi alternativa l'equazione  $\sum_{i=1}^n p_i Z_i^{*k} = \mu_k$  non è richiesta per identificare i pesi  $p_i$ , ma è sufficiente l'equazione  $\sum_{i=1}^n p_i = 1$ . Perciò sotto l'ipotesi  $H_1$  la funzione EL non parametrica è data da :  $L(F_{H_1}) = L(Z_1^*, Z_2^*, \dots, Z_n^* | \mu_k) = \prod_{i=1}^n \frac{1}{n} = \left(\frac{1}{n}\right)^n$ .

Ora consideriamo il test log rapporto di verosimiglianza  $(-2LLR)_k$  per l'ipotesi

$H_0: E(Z^{*k}) = \mu_k$  contro  $H_1: E(Z^{*k}) \neq \mu_k$ .

Perciò considerando l'ipotesi nulla e l'ipotesi alternativa, il test statistico viene semplificato dalla trasformazione:

$$(-2LLR)_k = -2 \log \frac{L(F_{H_0})}{L(F_{H_1})} = -2 \log \frac{L(Z_1^*, Z_2^*, \dots, Z_n^* | \mu_k)}{L(Z_1^*, Z_2^*, \dots, Z_n^*)}$$

Con una semplice sostituzione si può semplificare in questo modo:

$$(-2LLR)_k = -2 \log \frac{\prod_{i=1}^n \frac{1}{n(1 + \lambda_k(Z_i^{*k} - \mu_k))}}{\prod_{i=1}^n \left(\frac{1}{n}\right)} = 2 \sum_{i=1}^n \log[1 + \lambda_k(Z_i^{*k} - \mu_k)]$$

dove  $\lambda$  è una radice dell'equazione  $\sum_{i=1}^n \frac{Z_i^{*k} - \mu_k}{1 + \lambda_k(Z_i^{*k} - \mu_k)} = 0$ .

Il test rapporto di verosimiglianza viene usato per classificare i valori critici aggiustati usati per decidere se rifiutare o meno  $H_0$ . Rifiuteremo l'ipotesi nulla se:

- 1- Per il test  $ELR_Z$ , si utilizza come statistica test il valore più grande del test  $(-2LLR)_k$  tra quelli calcolati per  $k \in \{1,2,3,4\}$ , ossia per i primi quattro momenti della variabile mezza-normale standardizzata. Rifiuteremo dunque l'ipotesi nulla se il valore calcolato della statistica test è maggiore di  $C_\alpha$ , che è il valore soglia ed è il percentile  $(1 - \alpha)$  della distribuzione  $\chi^2(1)$ . Questa è la distribuzione della statistica test sotto  $H_0$ , perché i  $W_i$  sono asintoticamente indipendenti.
- 2- Per il test SEELR, si utilizza come statistica test il valore più grande di  $(-2LLR)_k$  tra quelli calcolati per  $k \in \{3,4,5,6,7,8,9,10,11\}$ , ossia per i momenti della variabile t di Student. Il livello di significatività osservato di questo test può essere ottenuto da valori tabulati della statistica test sotto  $H_0$ . Per tabulare i valori critici di questo test è

stato condotto uno studio Monte Carlo usando 50000 iterazioni per diverse dimensioni dei campioni e livelli di significatività nominali. Si possono osservare nella tabella 1, che si trova nell'appendice, i valori critici del test statistico SEELR per campioni di dimensioni finite.

### 3.3 Risultati e valutazioni sul test $ELR_Z$

Nel loro studio, *Marange e Qin*<sup>[11]</sup> comparano la potenza del test proposto  $ELR_Z$  con alcuni dei comuni test per bontà di adattamento che includono il test di *Anderson-Darling* (AD), il test di *Kolmogorov-Smirnov modificato* (KS), il test di *Cramer-von Mises* (CVM), il test di *Jarque-Bera* (JB), il test di *Shapiro-Wilk* (SW), il test rapporto di verosimiglianza empirica basata sulla densità (DB), e il test di verosimiglianza empirica basato sulle relazioni tra i momenti (SEELR), ad un livello di significatività del 5%. Lo studio è eseguito attraverso una simulazione Montecarlo usando 5000 replicazioni per tutti i test con campioni di dimensioni variabili ( $n=20,30,50$  e  $80$ ) con differenti distribuzioni alternative.

I risultati del confronto delle potenze sono presenti nelle tabelle 2-5. Sotto casi simmetrici definiti su  $(-\infty, +\infty)$  il nuovo test  $ELR_Z$  è più performante rispetto agli altri test sotto le distribuzioni alternative considerate, ma leggermente meno rispetto al test JB. Per distribuzioni simmetriche definite in  $(0,1)$  il test  $ELR_Z$  è comparabile col test DB ed ha prestazioni migliori rispetto a tutti gli altri test considerati. Inoltre, quando l'alternativa è Beta  $(0.5, 0.5)$  il test  $ELR_Z$  è comparabile con i test SW e SEELR mentre è più efficiente solo dei test KS, CVM, e JB. Per distribuzioni asimmetriche definite in  $(0, +\infty)$ , i test SW e SEELR hanno le prestazioni migliori, perciò in queste situazioni si dovrebbero preferire questi test. I test AD e DB sono comparabili e hanno prestazioni migliori rispetto agli altri test proposti, come i test KS e CVM. Infine, nella categoria delle distribuzioni alternative asimmetriche definite su  $(-\infty, +\infty)$  il test  $ELR_Z$  è comparabile con il test SEELR per campioni di piccole dimensioni (per esempio:  $n=20,30$ ). In conclusione, solo il test SEELR è comparabile con il test  $ELR_Z$  tra tutti i test studiati. In generale, quando consideriamo tutti i test per la



normalità, è possibile notare che i test JB,  $ELR_Z$  e SW sono generalmente quelli che hanno prestazioni migliori con alternative simmetriche definite su  $(-\infty, +\infty)$ , mentre i test DB e  $ELR_Z$  sono i migliori per le alternative simmetriche definiti in  $(0, 1)$ . E poi, i test SEELR e SW sono i migliori per alternative asimmetriche definite in  $(0, +\infty)$ , mentre i test JB e SW hanno prestazioni migliori per alternative asimmetriche definite su  $(-\infty, +\infty)$ .

**Tabella 2:** Risultati della simulazione Monte Carlo basata su campioni di dimensione  $n$  con distribuzione alternativa simmetrica definita su  $(-\infty; \infty)$ , con  $\alpha=0.05$  potenza dei test.

Distribuzione alternativa simmetrica definita su $(-\infty; \infty)$ , con $\alpha=0.05$									
Distribuzione	$n$	AD	$KS_M$	CVM	JB	SW	DB	SEELR	$ELR_{Z2}$
t(2)	20	0.5068	0.4482	0.5138	0.5632	0.5282	0.2806	0.3774	0.5268
	30	0.6834	0.5832	0.6552	0.7016	0.6908	0.3946	0.4228	0.7004
	50	0.8538	0.7782	0.8370	0.8812	0.8572	0.5640	0.4800	0.8726
	80	0.9602	0.9200	0.9554	0.9646	0.9566	0.8010	0.5420	0.9658
t(4)	20	0.2270	0.1768	0.2114	0.2898	0.2410	0.0922	0.1698	0.2450
	30	0.3002	0.2182	0.2764	0.3788	0.3338	0.1084	0.2164	0.3398
	50	0.4150	0.3176	0.3794	0.5400	0.4520	0.1388	0.2468	0.4784
	80	0.5558	0.3994	0.5210	0.7064	0.6282	0.2094	0.2784	0.6760
t(7)	20	0.1162	0.0952	0.1006	0.1670	0.1398	0.0492	0.1066	0.1346
	30	0.1404	0.1008	0.1306	0.2222	0.1806	0.0552	0.1188	0.1664
	50	0.1806	0.1272	0.1578	0.2954	0.2362	0.0502	0.1422	0.2276
	80	0.2380	0.1618	0.2086	0.4010	0.3122	0.0650	0.1590	0.3324
Cauchy(0,1)	20	0.8780	0.8386	0.8898	0.8622	0.8674	0.7012	0.6368	0.8450
	30	0.9672	0.9410	0.9622	0.9574	0.9610	0.8606	0.6910	0.9542
	50	0.9976	0.9950	0.9964	0.9954	0.9958	0.9712	0.7424	0.9976
	80	1.0000	1.0000	0.9998	0.9998	0.9998	0.9992	0.8882	1.0000
Cauchy(0,5)	20	0.8778	0.8374	0.8796	0.8650	0.8704	0.6902	0.6454	0.8550
	30	0.9628	0.9414	0.9648	0.9512	0.9590	0.8664	0.6950	0.9542
	50	0.9968	0.9948	0.9976	0.9968	0.9966	0.9730	0.7468	0.9962
	80	1.0000	1.0000	1.0000	0.9998	1.0000	0.9996	0.8872	1.0000
Logistica	20	0.1090	0.0872	0.0982	0.1460	0.1138	0.0436	0.0944	0.1158
	30	0.1176	0.0908	0.1220	0.1982	0.1474	0.0452	0.1044	0.1482
	50	0.1562	0.1184	0.1456	0.2620	0.1986	0.0414	0.1216	0.1900
	80	0.2098	0.1406	0.1870	0.3474	0.2662	0.0468	0.1266	0.2908

**Tabella 3:** Risultati della simulazione Monte Carlo basata su campioni di dimensione  $n$  con distribuzione alternativa simmetrica definita su  $(0;1)$ , con  $\alpha=0.05$  potenza dei test.

Distribuzione alternativa simmetrica definita su $(0;1)$ , con $\alpha=0.05$ .									
Distribuzione	$n$	AD	$KS_M$	CVM	JB	SW	DB	SEELR	$ELR_{Z2}$
Beta(2,2)	20	0.0564	0.0544	0.0594	0.0052	0.0516	0.1310	0.0696	0.0970
	30	0.0786	0.0520	0.0812	0.0012	0.0768	0.2004	0.0550	0.1962
	50	0.1222	0.0852	0.1172	0.0010	0.1528	0.3468	0.0628	0.4252
	80	0.2340	0.1256	0.1834	0.0128	0.3170	0.5978	0.1128	0.7204
Beta(3,3)	20	0.0404	0.0474	0.0408	0.0076	0.0372	0.0780	0.0518	0.0620
	30	0.0786	0.0520	0.0812	0.0046	0.0768	0.1112	0.0392	0.1030
	50	0.0736	0.0524	0.0650	0.0014	0.0682	0.1654	0.0326	0.1906
	80	0.1076	0.0762	0.0826	0.0022	0.1128	0.2772	0.0298	0.3458
Beta(0.5,0.5)	20	0.6160	0.3098	0.5058	0.0066	0.7190	0.9094	0.7092	0.7015
	30	0.8576	0.4998	0.7332	0.0052	0.9392	0.9914	0.8830	0.8960
	50	0.9902	0.7976	0.9568	0.3822	0.9992	1.0000	0.9916	0.9956
	80	1.0000	0.9724	0.9990	0.9872	1.0000	1.0000	1.0000	1.0000
Uniform(0,1)	20	0.1640	0.1014	0.1396	0.0040	0.1886	0.4064	0.2598	0.3332
	30	0.3004	0.1422	0.2262	0.0020	0.3894	0.6622	0.3202	0.6002
	50	0.5780	0.2532	0.4282	0.0118	0.7546	0.9358	0.5624	0.9120
	80	0.8636	0.4578	0.7092	0.3706	0.9688	0.9990	0.8730	0.9944
Logit-norm(0,1)	20	0.0648	0.0442	0.0562	0.0056	0.0578	0.1294	0.0700	0.1010
	30	0.0858	0.0574	0.0748	0.0024	0.0796	0.1974	0.0658	0.1990
	50	0.1394	0.0812	0.1220	0.0010	0.1612	0.3420	0.0676	0.4156
	80	0.2630	0.1368	0.2114	0.0126	0.3408	0.5830	0.1094	0.7108
Logit-norm(0,2)	20	0.3758	0.1844	0.2934	0.0046	0.4366	0.7034	0.4806	0.5348
	30	0.6092	0.2884	0.4822	0.0030	0.7342	0.9150	0.6512	0.8258
	50	0.9016	0.5412	0.7814	0.1174	0.9742	0.9976	0.9006	0.9818
	80	0.9942	0.8170	0.9644	0.8594	1.0000	1.0000	0.9958	0.9996

**Tabella 4:** Risultati della simulazione Monte Carlo basata su campioni di dimensione  $n$  con distribuzione alternativa asimmetrica definita su  $(0; \infty)$ , con  $\alpha=0.05$  potenza dei test.

Distribuzione alternativa simmetrica definita su $(0;1)$ , con $\alpha=0.05$ .									
Distribuzione	$n$	AD	$KS_M$	CVM	JB	SW	DB	SEELR	$ELR_{Z2}$
Exp(1)	20	0.7850	0.5722	0.7222	0.6230	0.8334	0.8384	0.8522	0.3642
	30	0.9296	0.7780	0.8922	0.8286	0.9646	0.9754	0.9996	0.4752
	50	0.9972	0.9594	0.9878	0.9756	0.9998	0.9992	1.0000	0.6400
	80	1.0000	0.9990	0.9998	0.9998	1.0000	1.0000	1.0000	0.8114
Gamma(2,1)	20	0.4590	0.3066	0.4136	0.4080	0.5380	0.4420	0.5684	0.2264
	30	0.6662	0.4776	0.6072	0.5852	0.7502	0.6876	0.8094	0.2844
	50	0.8960	0.6926	0.8436	0.8242	0.9500	0.9180	0.9668	0.3822
	80	0.9840	0.8962	0.9682	0.9782	0.9976	0.9914	0.9984	0.5210
Lognorm(0,1)	20	0.9080	0.7760	0.8846	0.8172	0.9350	0.9210	0.9418	0.6036
	30	0.9838	0.9304	0.9730	0.9466	0.9888	0.9906	1.0000	0.7418
	50	1.0000	0.9942	0.9998	0.9976	1.0000	1.0000	1.0000	0.9068
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9838
Lognorm(0,2)	20	0.9986	0.9904	0.9970	0.9840	0.9990	0.9998	0.9999	0.8894
	30	0.9998	0.9998	1.0000	0.9994	1.0000	1.0000	1.0000	0.9684
	50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Weibull(2,1)	20	0.1348	0.0980	0.1142	0.1258	0.1582	0.1264	0.1626	0.0932
	30	0.1828	0.1306	0.1654	0.1704	0.2274	0.1958	0.2718	0.0892
	50	0.3050	0.2000	0.2530	0.2738	0.4086	0.3446	0.5202	0.1120
	80	0.4954	0.3186	0.4200	0.4346	0.6644	0.5634	0.7812	0.1080
Weibull(0.5,1)	20	0.9962	0.9810	0.9954	0.9562	0.9990	0.9996	0.9986	0.8014
	30	1.0000	0.9990	1.0000	0.9972	1.0000	1.0000	1.0000	0.9168
	50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9866
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

**Tabella 5:** Risultati della simulazione Monte Carlo basata su campioni di dimensione  $n$  con distribuzione alternativa asimmetrica definita su  $(-\infty; \infty)$ , con  $\alpha=0.05$  potenza dei test.

Distribuzione alternativa asimmetrica definita su $(-\infty; \infty)$ , con $\alpha=0.05$ .									
Distribuzione	$n$	AD	$KS_M$	CVM	JB	SW	DB	SEELR	$ELR_{Z2}$
$t(\delta = 1, \nu = 2)$	20	0.6446	0.5692	0.6440	0.6556	0.6498	0.4612	0.5688	0.5542
	30	0.8060	0.7178	0.7934	0.8080	0.8072	0.6210	0.6678	0.7242
	50	0.9492	0.8900	0.9394	0.9414	0.9410	0.7820	0.7782	0.8872
	80	0.9928	0.9762	0.9892	0.9924	0.9924	0.9294	0.8410	0.9726
$t(\delta = 1, \nu = 4)$	20	0.3180	0.2368	0.2848	0.3606	0.3142	0.1638	0.2790	0.2744
	30	0.4086	0.3246	0.3884	0.4810	0.4518	0.2262	0.3584	0.3718
	50	0.5912	0.4618	0.5538	0.6592	0.6360	0.3202	0.4830	0.5344
	80	0.7642	0.6370	0.7296	0.8290	0.8108	0.4626	0.5826	0.7084
$t(\delta = 1, \nu = 7)$	20	0.1490	0.1138	0.1404	0.1934	0.1692	0.0766	0.1420	0.1492
	30	0.1958	0.1424	0.1736	0.2722	0.2318	0.0940	0.1876	0.1936
	50	0.2846	0.1968	0.2522	0.3834	0.3372	0.1194	0.2556	0.2738
	80	0.3848	0.2756	0.3542	0.5102	0.4556	0.1620	0.3300	0.3798
SN(0,1,2)	20	0.0896	0.0756	0.0882	0.1054	0.1068	0.0636	0.0978	0.0710
	30	0.1194	0.0912	0.1062	0.1214	0.1422	0.0784	0.1336	0.0792
	50	0.1666	0.1260	0.1488	0.1810	0.1968	0.1164	0.2080	0.0800
	80	0.2434	0.1918	0.2258	0.2712	0.2940	0.1550	0.3074	0.0890
SN(0,1,5)	20	0.2406	0.1744	0.2212	0.2060	0.2660	0.2092	0.2810	0.1144
	30	0.3586	0.2604	0.3152	0.3056	0.4230	0.3346	0.4742	0.1278
	50	0.5796	0.4098	0.5430	0.4768	0.6672	0.5250	0.7110	0.1372
	80	0.8080	0.6084	0.7554	0.7378	0.8888	0.7394	0.9020	0.1722
SC(0,2,5)	20	0.9660	0.9360	0.9658	0.9410	0.9736	0.9436	0.9462	0.8482
	30	0.9978	0.9884	0.9954	0.9910	0.9970	0.9882	0.9774	0.9524
	50	1.0000	0.9998	1.0000	0.9998	0.9998	0.9992	0.9844	0.9940
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9976	0.9998

### 3.4 Aspetti computazionali

E' di primaria importanza determinare il costo computazionale del nuovo algoritmo concentrandosi sul tempo computazionale richiesto dal test proposto rispetto agli altri test esistenti considerati. Per stimarlo, è stato utilizzato il software "R 64 Bit Windows 10 Home" e un processore Intel Core i5-4210U di quarta generazione, con velocità della memoria cache di 1.7 GHz e memoria RAM di 4GB PC3 DDR3L SDRAM. La simulazione è stata eseguita usando 5000 replicazioni per ogni test con un campione di dimensione  $n=80$ .

Nei risultati visibili in tabella 6 è evidenziato solo un chiaro vantaggio del nuovo test proposto rispetto al noto JB test. Inoltre, il test  $ELR_z$  è comparabile con il test SEELR, mentre è inferiore al test DB. I test SW, CVM, KS, e AD risultano comunque i più efficienti dal punto di vista computazionale rispetto al resto dei test considerati.

Tabella 6: Comparazione del tempo computazionale (in secondi) per i test studiati.

Test	Replicazioni	Elapsed	Relative	User.self	Sys.self
AD	5,000	1.14	2.000	1.14	0.00
CVM	5,000	0.82	1.439	0.81	0.00
DB	5,000	17.00	29.825	16.94	0.04
$ELR_z$	5,000	44.83	78.649	44.78	0.01
JB	5,000	252.64	443.228	252.60	0.00
$KM_S$	5,000	0.89	1.561	0.90	0.00
SEELR	5,000	45.42	79.684	45.41	0.00
SW	5,000	0.57	1.000	0.58	0.00

### 3.5 Conclusioni e commenti

In questo capitolo è stato illustrato un test rapporto di verosimiglianza empirica basato sui vincoli dei momenti di una distribuzione mezza normale ( $ELR_Z$ ) e si è posto in confronto con il test da cui gli autori hanno preso spunto (SEELR), e con altri test di normalità presenti in natura. In generale, il test  $ELR_Z$  proposto ha buone proprietà ed ha prestazioni migliori rispetto ai conosciuti test esistenti nel caso di distribuzioni alternative simmetriche. Le proprietà del test  $ELR_Z$  derivano dal metodo EL, in grado di integrare diverse informazioni dall'utilizzo dei primi quattro vincoli dei momenti.

Si può affermare che il test proposto  $ELR_Z$  sia la scelta da preferire quando si deve valutare la divergenza dalla normalità verso distribuzioni alternative simmetriche per campioni con piccole o moderate dimensioni. Ad ogni modo, bisogna sottolineare che questo test ha poca potenza quando si considerano distribuzioni alternative asimmetriche. Per quanto riguarda gli aspetti computazionali il test non comporta vantaggi rispetto ai test considerati tranne rispetto al test JB, ed è comparabile con il test SEELR.

## CAPITOLO 4

### APPLICAZIONE DEI TEST SEELR E ELRz AD UN CASO DI STUDIO REALE

Per verificare le capacità dei due test appena descritti, vediamo ora un'applicazione ad un caso di studio reale.<sup>[16]</sup> Il dataset proviene da uno studio riguardante la malattia coronarica (coronaropatia), condotto su 141 pazienti, del dipartimento di cardiologia del "Teaching Hospital" in Galizia, comunità autonoma nel nord-ovest della Spagna. Attraverso l'uso di un biomarcatore, dopo una angiografia coronaria, 96 di questi pazienti hanno riportato alcune lesioni coronarie, mentre i restanti 45 sono risultati sani. Lavoriamo, dunque, su due dataset distinti, uno dove abbiamo i dati osservati dei pazienti sani e l'altro di quelli malati.

#### 4.1 La curva ROC

Nella medicina moderna, i test diagnostici ideali discriminano perfettamente i malati dai sani, cioè, una volta eseguito il test, gli individui vengono classificati con assoluta certezza come affetti o non affetti dalla malattia di interesse. Però, quando un determinato test diagnostico non riesce a discriminare in maniera netta i malati dai sani, cioè quando le distribuzioni dei risultati del test sono parzialmente sovrapposte negli individui affetti e non affetti da una specifica malattia, è necessario calcolare il grado di incertezza della classificazione. Se il risultato del test è una variabile continua, si può ricorrere all'analisi della curva ROC (Receiver Operating Characteristics). La curva ROC è un metodo per validare un test diagnostico, in quanto è una tecnica statistica che misura l'accuratezza del test lungo tutto il range dei valori possibili. La curva ROC permette anche di identificare il valore soglia ottimale (il cosiddetto best cut-off), cioè il valore del test che massimizza la differenza tra i veri positivi (cioè la proporzione di individui che hanno un valore alterato del test tra tutti quelli realmente affetti dalla malattia) e i falsi positivi (cioè la proporzione di individui che pur avendo un valore alterato del test non sono affetti dalla malattia di interesse).



La curva ROC viene costruita considerando tutti i possibili valori del test e, per ognuno di questi, si calcola la proporzione di veri positivi (la sensibilità) e la proporzione di falsi positivi. La proporzione di falsi positivi si calcola con la formula standard:  $1 - \text{specificità}$  (ovvero, la proporzione di veri negativi). Congiungendo i punti che hanno come coordinate la proporzione di veri positivi e di falsi positivi si ottiene una curva, chiamata appunto curva ROC. L'area sottostante alla curva ROC, denominata AUC ("Area Under the Curve") è una misura di accuratezza diagnostica. Se un ipotetico nuovo test discriminasse perfettamente i malati dai sani, l'AUC avrebbe valore 1, cioè il 100% di accuratezza. Nel caso in cui il nuovo test non discriminasse per niente i malati dai sani, la curva ROC avrebbe un'area di 0.5 (o 50%) che coinciderebbe con l'area sottostante la diagonale principale del quadrato di lato 1 nel primo quadrante del grafico. Nella realtà, si considera adeguato un test diagnostico con un'area sotto la curva  $\geq 80\%$ .

#### **4.2 Introduzione al caso di studio**

L'utilizzo dei biomarcatori per discriminare pazienti sani e pazienti malati è divenuto sempre più frequente negli ultimi tempi. L'assunzione di base di questa metodologia è spesso, ma non sempre, che valori elevati del biomarcatore (indichiamolo con  $Y$ ) vengano associati alla presenza della malattia. Tramite la curva ROC viene valutata l'accuratezza relativa al valore soglia ottimale  $c$  che discrimina tra individui classificati come sani ( $Y < c$ ) o come malati ( $Y \geq c$ ).

Per identificare l'ottimo valore soglia esistono diversi metodi, come quelli discussi nell'articolo analizzato.<sup>[16]</sup>

Spesso le metodologie proposte si basano sull'assunzione di normalità dei dati. Nello studio viene applicato il test di Shapiro-Wilk sui dati del biomarcatore o su una trasformazione Box-Cox di questi per vedere, ad un livello di significatività del 5%, se questi possono essere assunti come normali.

Il test di Shapiro-Wilk effettuato nello studio, lavorando sui dati reali, ha riportato un p-value per il gruppo dei sani (gruppo 0) pari a 0.0746, che ha portato ad accettare l'ipotesi di normalità distributiva dei dati. Per il gruppo 1, ossia quello dei pazienti malati, il p-value risulta minore di 0.0001, per cui viene rifiutata l'ipotesi di normalità distributiva.

Il test è stato riprovato dopo aver eseguito una trasformazione Box-Cox sui dati con  $\hat{\lambda} = 0.1136$ . In questo caso il test ha portato ad accettare l'ipotesi di normalità distributiva solo per i dati dei pazienti malati (p-value pari a 0.0793), mentre quelli sani hanno riportato un p-value pari a 0.0091, che ha portato a rifiutare tale ipotesi.

Ora l'obiettivo è quello di verificare se anche i due test SEELR e  $ELR_z$  forniscono le stesse conclusioni del test di Shapiro-Wilk rispetto all'ipotesi di normalità dei due campioni. Utilizzeremo i due test sia sui campioni dei dati originali, sia su quelli trasformati tramite la trasformazione di Box-Cox:

$$X_{tras} = \frac{X^{\lambda} - 1}{\lambda}$$

dove per il valore di  $\lambda$  consideriamo quello ottenuto tramite la massima verosimiglianza nell'articolo oggetto di studio, ossia pari a 0.1136.

### 4.3 Applicazione dei due test tramite il software R

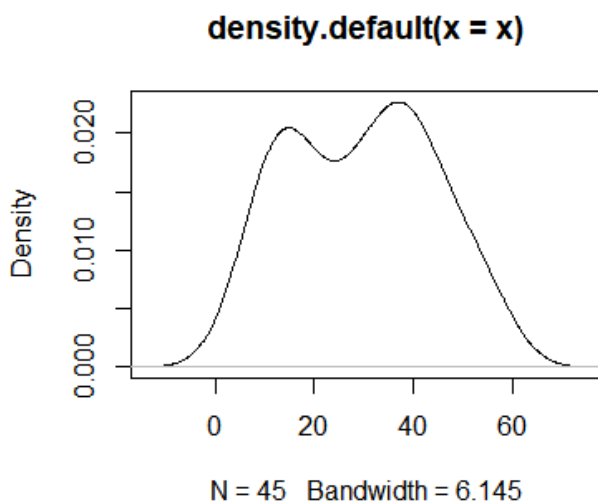
Vengono studiati inizialmente i due campioni dei dati osservati e i due trasformati attraverso una breve analisi esplorativa, e poi, successivamente, verranno applicati i due test.

E' stato utilizzato il software "R 64 Bit Windows 10 Home" e i relativi comandi utilizzati sono riportati nell' "Appendice dei comandi R" (si veda l'ultimo capitolo di questo lavoro). I comandi per formulare i test sono stati rielaborati partendo da quelli presenti nell'articolo.<sup>[18]</sup>

Le funzioni che sono state utilizzate per costruire i due test sono le seguenti:

- $\text{gammaf}(n,k)$  è la funzione che ritorna il valore del k-esimo momento della distribuzione t di Student con  $(n-2)$  gradi di libertà; con  $n$  che indica il numero di unità statistiche.
- $\text{momento}(k)$  è la funzione che ritorna il valore del k-esimo momento della distribuzione mezza-normale.
- $\text{genedata}(n)$  e  $\text{genedataE}(n)$  sono le funzioni che restituiscono i dati originali opportunamente trasformati, rispettivamente per il test SEELR e per il test  $ERL_Z$ .
- $\text{critical}(x)$  e  $\text{criticalE}(x)$  sono le funzioni che restituiscono il valore della statistica test, rispettivamente per il test SEERL e per il test  $ERL_Z$ .
- $\text{SEELR}(n,\alpha,\text{CriticalValue})$  e  $\text{ELRZ}(n,\alpha,\text{CriticalValue})$  sono i due test completi, dove  $\alpha$  è il livello di significatività del test ed è stato fissato pari a 0.05, il valore critico per il test SEELR viene ripreso dalle tabelle presentate nel capitolo 3, mentre per quello per il test  $ERL_Z$  è l'opportuno quantile della distribuzione  $\chi^2$  con 1 grado di libertà.

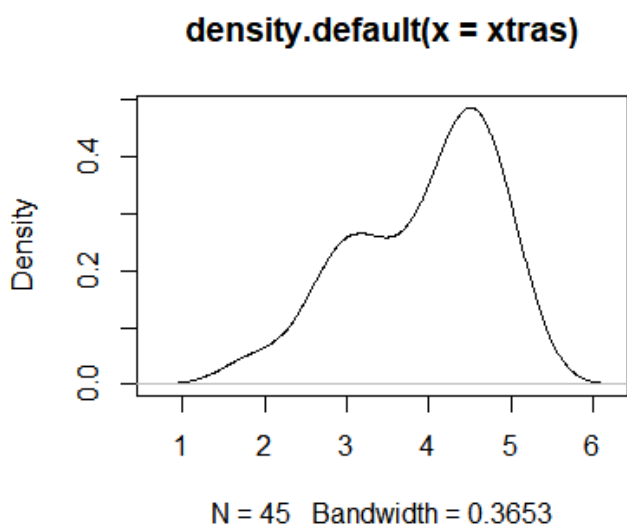
Viene osservato come primo grafico la stima della densità di frequenza dei pazienti considerati sani.



Nonostante la forma della densità sia dissimile da quella normale (si nota, in particolare, la doppia “gobba” nella zona centrale), riprovando ad effettuare il test di Shapiro-Wilk, in effetti, notiamo che l’ipotesi di normalità non viene rifiutata.

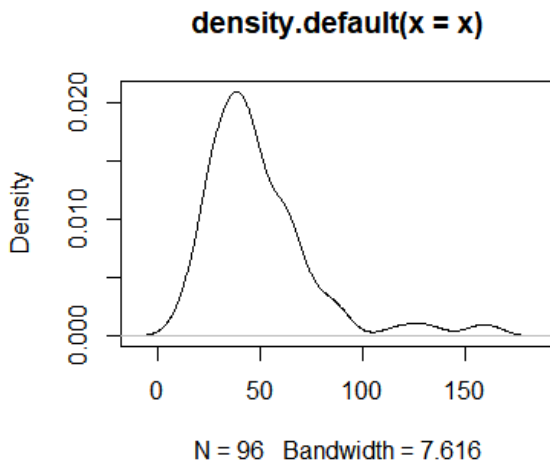
Anche i test SEELR e  $ERL_Z$  confermano l’ipotesi di normalità distributiva dei dati osservati, in quanto riportano rispettivamente un p-value maggiore di 0.09 e uno pari a 0.0728629.

Come secondo grafico, viene osservata la stima della densità di frequenza dei valori dei pazienti sani trasformati tramite la trasformazione di Box-Cox.



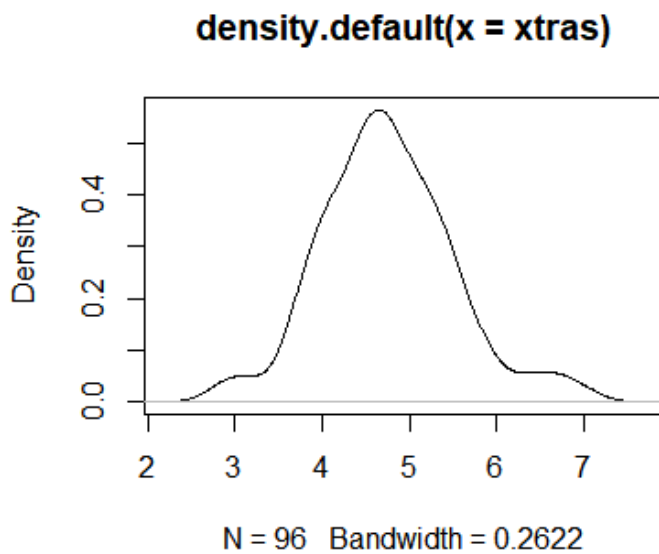
Dalla distribuzione si nota una forma decisamente asimmetrica, con la coda sinistra più accentuata, che non ricorda l’andamento campanulare della distribuzione normale. Infatti riapplicando il test di Shapiro-Wilk, notiamo che, per i dati trasformati dei pazienti sani, l’ipotesi alternativa non viene rifiutata. Anche il test SEELR, riportando un p-value minore di 0.01, porta a rifiutare l’ipotesi nulla. Al contrario, il test  $ERL_Z$  riporta un p-value di 0.4281081, un valore molto elevato, che ci porterebbe a non rifiutare l’ipotesi nulla. Per indagare sulle cause di questo diverso comportamento, sono stati effettuati altri test di normalità su questi dati, quali CVM, LL, JB e AD. I risultati ottenuti, visibili nell’appendice in fondo al capitolo, sono particolari, poiché tutti i test effettuati, tranne il test di Jarque-Bera, che ha riportato un p-value elevato, pari a 0.158, hanno confermato che l’ipotesi nulla andrebbe rifiutata in questo caso.

Viene osservato come terzo grafico la stima della densità di frequenza dei valori osservati dei pazienti malati.



Dalla distribuzione si nota una forma che può ricordare quella di una normale, ma con una coda destra molto lunga e appiattita. Infatti riapplicando il test di Shapiro-Wilk, notiamo che l'ipotesi nulla viene rifiutata. Anche utilizzando i due test considerati in questo lavoro, SEELR e  $ERL_Z$ , si giunge alle medesime conclusioni, infatti otteniamo rispettivamente un p-value molto minore di 0.01, e uno pari a 0.03081359.

Viene osservato come quarto grafico la stima della densità di frequenza dei valori dei pazienti malati trasformati tramite la trasformazione di Box-Cox.



Dalla distribuzione si nota una forma molto simile a quella di una distribuzione normale, nonostante le pendenze siano molto più accentuate. Riapplicando il test di Shapiro-Wilk, si nota che l'ipotesi di normalità non viene rifiutata. Anche i test SEELR e  $ERL_Z$  confermano questa conclusione, in quanto uno riporta un p-value maggiore di 0.09 e l'altro pari a 0.3578957.

#### 4.4 Conclusioni e commenti

I risultati ottenuti sono abbastanza soddisfacenti, in quanto solo una applicazione del test  $ERL_Z$  è risultata essere in disaccordo rispetto al test di Shapiro-Wilk e al test SEELR.

I p-value risultati con i tre test differenti sono stati raggruppati nella tabella 4.1, qui sotto riportata.

Tabella 4.1 Confronto dei p-value nei tre test effettuati.

	Sani	Sani trasformati	Malati	Malati trasformati
<b>Shapiro-Wilk</b>	0.0746	0.0091	<0.0001	0.0793
<b>SEELR</b>	>0.09	<0.01	<<0.01	>0.09
<b><math>ERL_Z</math></b>	0.0728629	<b>0.4281081</b>	0.03081359	0.3578957

I p-value ottenuti tramite il test SEELR sono approssimati, perché i valori soglia del test sono stati tabulati e il valore del livello di significatività osservato del test oscilla tra (0.01 ; 0.09); per cui se il valore osservato della statistica test è minore del valore tabulato corrispondente al livello di significatività di 0.09 è stata usata la notazione  $>0.09$  per indicare il valore del p-value osservato. E se il valore osservato della statistica test è maggiore (molto maggiore) del valore tabulato corrispondente al livello di significatività di 0.01 è stata usata la notazione  $<0.01$  ( $\ll 0.01$ ) per indicare il valore del p-value osservato.

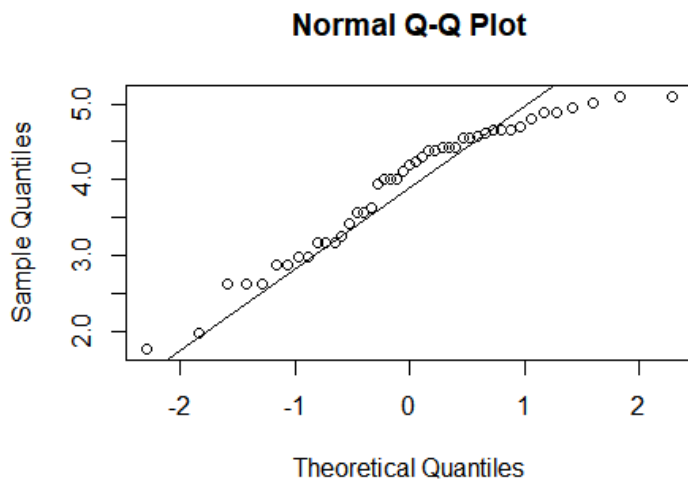
I p-value ottenuti per i dati dei pazienti sani sono tra loro più o meno simili, con quello del test SEELR leggermente più elevato, e, dunque, ci indicano tutti e tre di non rifiutare l'ipotesi di normalità.

I p-value ottenuti per i dati dei pazienti malati ci indicano (se prendiamo come riferimento un livello di significatività del test del 5%) di rifiutare l'ipotesi di normalità. Si nota, però, una differenza netta tra il valore del p-value riportato dal test  $ERL_Z$  rispetto agli altri due test. È un valore più alto, che ci porterebbe a non rifiutare l'ipotesi nulla se facessimo riferimento ad un il livello di significatività del test pari all'1%.

I p-value ottenuti per i dati trasformati tramite Box-Cox dei pazienti malati supportano l'ipotesi di normalità dei dati. Anche in questo caso notiamo che il p-value ottenuto con il test  $ERL_Z$  ha un valore più grande rispetto a quelli trovati con gli altri due test.

Invece, i p-value ottenuti per i dati trasformati tramite Box-Cox dei pazienti sani non forniscono informazioni concordi, in quanto per i test di Shapiro-Wilk e SEELR sono minori di 0.01 e non supportano l'ipotesi di normalità distributiva, mentre il p-value del test  $ERL_Z$  ha un valore molto elevato che sostiene l'ipotesi nulla. Nel paragrafo precedente è stato spiegato che sono stati effettuati altri test per verificare la normalità, e si è notato che anche il test di Jarque-Bera ha riportato un p-value molto elevato.

Per indagare ulteriormente su questa discrepanza di valori abbiamo utilizzato uno strumento grafico, il QQ plot, ottenuto tramite la funzione `qqnorm()` e `qqline()`.



Dal grafico si nota che l'ipotesi di normalità non sembra essere supportata, in quanto nel grafico, i valori si discostano in modo evidente dalla retta ipotizzata.



## Appendice

### A.1 Valori tabulati del test SEELR

**Tabella 1:** Valori critici per il test SEELR.

---

n	Livello di significativa del test ( $\alpha$ )								
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
15	24.3930	19.3261	16.9120	15.4022	14.3492	13.4991	12.8626	12.2387	11.6943
20	17.6161	15.0413	13.6680	12.5892	11.8352	11.1375	10.5549	10.0533	9.6123
25	16.2809	14.0189	12.5428	11.5545	10.7768	10.1616	9.6060	9.1091	8.7200
30	15.5298	13.2217	11.8489	10.8738	10.1421	9.5159	8.9575	8.4924	8.0781
35	15.0866	12.8293	11.3928	10.4383	9.7148	9.1151	8.5892	8.1143	7.7130
40	14.4665	12.3274	10.9934	10.0512	9.3113	8.7334	8.2359	7.7791	7.3046
45	14.4661	12.2105	10.8267	9.8797	9.0569	8.4484	7.9645	7.5208	7.1550
50	14.0975	11.9068	10.5987	9.5877	8.8519	8.2525	7.7661	7.2961	7.3534
55	14.0778	11.7150	10.3508	9.3901	8.6536	8.0371	7.5330	7.0825	7.1281
60	13.5947	11.5098	10.1436	9.2279	8.5075	7.9042	7.4281	6.9912	6.8782
65	13.1798	11.2198	9.8949	8.9337	8.2215	7.5988	7.1381	6.7177	6.6904
70	13.3355	11.0741	9.8095	8.8760	8.1669	7.6308	7.1583	6.7256	6.5998
75	13.3144	11.1228	9.8193	8.9210	8.1576	7.5642	7.0739	6.6729	6.3694
80	13.0226	10.7813	9.4078	8.5643	7.8361	7.2631	6.8206	6.4049	6.3566
85	13.0792	10.8199	9.5703	8.6852	7.9650	7.3351	6.8247	6.4078	6.2947
90	13.0047	10.7469	9.4974	8.5759	7.8112	7.2117	6.7483	6.3164	5.9636
95	12.2724	10.3989	9.1858	8.3127	7.6364	7.0768	6.5738	6.1890	5.8172
100	12.3826	10.2166	9.0359	8.1974	7.5016	6.9129	6.4497	6.0454	5.6997
120	12.2724	10.0474	8.8448	8.0646	7.3898	6.8083	6.3634	5.9694	5.6340
150	11.9573	9.6551	8.4874	7.6337	7.0034	6.4734	6.0403	5.6722	5.3376
170	11.7115	9.4745	8.3023	7.4370	6.8209	6.2712	5.8433	5.4709	5.1503
200	11.2909	9.1640	8.0268	7.2457	6.5874	6.0779	5.6301	5.2719	4.9492

---

---

n	Livello di significatività del test ( $\alpha$ )								
	0.1	0.12	0.13	0.14	0.15	0.17	0.2	0.25	0.3
15	11.1872	10.3578	10.0024	9.6658	9.3559	8.7994	8.0444	7.0402	6.2055
20	9.2128	8.4870	8.1719	7.8987	7.6197	7.1575	6.5516	5.6681	4.9263
25	8.3140	7.6264	7.3018	7.0398	6.7947	6.3177	5.7419	4.9381	4.2787
30	7.7120	7.0646	6.8154	6.5467	6.3014	5.8537	5.3047	4.5101	3.8784
35	7.3469	6.7287	6.4478	6.1878	5.9538	5.4970	4.9419	4.1781	3.5820
40	6.9590	6.3416	6.0788	5.8274	5.5919	5.1940	4.6759	3.9401	3.3770
45	6.8075	6.2033	5.9508	5.7042	5.4653	5.0590	4.5249	3.7973	3.2213
50	7.0189	6.4024	6.1509	5.8895	5.6754	5.2638	4.6993	3.9916	3.3953
55	6.7867	6.1866	5.9378	5.6851	5.4443	5.0261	4.4929	3.7810	3.2209
60	6.5496	5.9565	5.6944	5.4653	5.2354	4.8184	4.3306	3.6575	3.0998
65	6.3830	5.8232	5.5750	5.3330	5.1121	4.7095	4.1938	3.5046	2.9645
70	6.2659	5.6816	5.4237	5.2052	4.9720	4.5625	4.0720	3.4082	2.8769
75	6.0309	5.4974	5.2581	5.0271	4.8127	4.4513	3.9576	3.3083	2.8046
80	6.0107	5.3851	5.1478	4.9460	4.7255	4.3712	3.8744	3.2303	2.7123
85	5.9899	5.4094	5.1615	4.9403	4.7181	4.3290	3.8341	3.1859	2.6693
90	5.6553	5.1019	4.8657	4.6535	4.4620	4.0930	3.6477	3.0070	2.5240
95	5.4968	4.9845	4.7294	4.5290	4.3293	3.9764	3.5405	2.9185	2.4585
100	5.3883	4.8838	4.6526	4.4388	4.2643	3.9082	3.4661	2.8853	2.4180
120	5.3342	4.8117	4.5645	4.3558	4.1620	3.8177	3.3717	2.7622	2.3192
150	5.0627	4.5412	4.3361	4.1235	3.9656	3.6009	3.1835	2.6226	2.1901
170	4.8761	4.3933	4.1760	3.9771	3.8082	3.4946	3.0875	2.5632	2.1302
200	4.6845	4.1926	3.9917	3.7932	3.6241	3.3114	2,9270	2.4185	2.0211

---

## A.2 Comandi R

```
> #Rappresentazione analitica delle distribuzioni
> plot(density(x))

> library(zipfR)
> library(emplik)

> ### TEST SEELR
> gammaf<-function(n,k){
+   a<-(Cgamma((k+1)/2)*Cgamma((n-2-k)/2)*(n-2)**(k/2))/(sqrt(pi)*Cgamma((n-2)/
2))
+ }

> genedata<- function(n){
+   x1<-x
+   for(k in 1:n){x[k]<-(x1[k]-mean(x1))*sqrt(n/(n-1))/(sd(x1[-k]))}
+   return(x)
+ }

> critical<-function(x){
+   #calcoliamo (-2LLR)_k per k=3,4,5,7
+
+   a3=el.test(x**3, m3)$'-2LLR'
+   a4=el.test(x**4, m4)$'-2LLR'
+   a5=el.test(x**5, m5)$'-2LLR'
+   a7=el.test(x**7, m7)$'-2LLR'
+
+   return(max(a3,a4,a5,a7))
+ }

> SEELR<-function(n, alpha, CriticalValue){
+   x<-genedata(n)
+   m3=0 #inizializzo i 4 momenti considerati
+   m5=0
+   m7=0
+   m4=gammaf(n,4)
+
+   if(critical(x)<CriticalValue){
+     test="Non rifiuto Ho"
+   }else{
+     test="Non rifiuto H1"}
+
+   ris=c(test,critical(x))
+   return(ris)
+ }

> ### TEST ELRZ
> n=nrow(sani)

> momento<-function(k){
+   a<-(Cgamma((k+1)/2)*(2**(k/2))/(sqrt(pi)))
+ }

> genedataE<- function(n){
+   x1<-x
+   for(k in 1:n){x[k]<-abs((x1[k]-mean(x1))*sqrt(n/(n-1))/(sd(x1[-k])))}
+   return(x)
+ }

> criticalE<-function(x){
+
+   #calcoliamo (-2LLR)_k per k=1,2,3,4
+
+   a1=el.test(x**1, m1)$'-2LLR'
+   a2=el.test(x**2, m2)$'-2LLR'
```

```

+ a3=e1.test(x**3, m3)$'-2LLR'
+ a4=e1.test(x**4, m4)$'-2LLR'
+
+ return(max(a1,a2,a3,a4))
+ }

> ELRZ<-function(n, alpha, CriticalValue){
+ x<-genedataE(n)
+ m1=momento(1) #inizializzo i primi 4 momenti
+ m2=momento(2)
+ m3=momento(3)
+ m4=momento(4)

+ if(criticalE(x)<CriticalValue){
+ test="Non rifiuto Ho"
+ }else{
+ test="Non rifiuto H1"}
+
+ ris=c(test,criticalE(x))
+ return(ris)
+ }

> #DATI DEI SANI
> sani<-read.table(file.choose(), header=TRUE)
> x<-sani[,1] #dati osservati
> n=nrow(sani)

> shapiro.test(x)

Shapiro-wilk normality test

data: x
W = 0.9544, p-value = 0.07464

> SEELR(45, 0.05, 9.0569)
[1] "Non rifiuto Ho" "6.09223703664708"

> if(6.09223703664708 < 7.1550){ #utilizzo i valori critici tabulati del test
+ alphaoss="è > 0.09"
+ }
> alphaoss
[1] "è > 0.09"

> ELRZ(45, 0.05, 3.841)
[1] "Non rifiuto Ho" "3.21731770072125"

> alphaoss<-1-pchisq(3.21731770072125,1)
> alphaoss
[1] 0.07286293

> # DATI TRASFORMATI DEI SANI con Box-Cox
> landa<-0.1136
> xtras<-((x**landa)-1)/landa #creo dei nuovi dati trasformati

> shapiro.test(xtras)

Shapiro-wilk normality test

data: x
W = 0.92957, p-value = 0.009059
> SEELR(45, 0.05, 9.0569)
[1] "Non rifiuto H1" "15.4962724009997"

> if(15.4962724009997>14.4661){ #utilizzo i valori critici tabulati del test
+ alphaoss="è < 0.01"
+ }

```

```

> alphaoss
[1] "è < 0.01"

> ELRZ(45, 0.05, 3.841)
[1] "Non rifiuto Ho"      "0.627950487596835"

> alphaoss<-1-pchisq(0.627950487596835, 1)
> alphaoss
[1] 0.4281081

> library(tseries)
> jarque.bera.test(x)

      Jarque Bera Test

data:  x
X-squared = 3.6905, df = 2, p-value = 0.158

> library(nortest)
> ad.test(x)

      Anderson-Darling normality test

data:  x
A = 1.1035, p-value = 0.006129

> cvm.test(x)

      Cramer-von Mises normality test

data:  x
W = 0.1966, p-value = 0.005579

> lillie.test(x)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  x
D = 0.14266, p-value = 0.02227

> pearson.test(x)

      Pearson chi-square normality test

data:  x
P = 16.556, p-value = 0.0205

>#DATI OSSERVATI DEI MALATI
> malati<-read.table(file.choose(), header=TRUE) #file y
> x<-malati[,1] #dati osservati
> n=nrow(malati)

> shapiro.test(x)

      Shapiro-wilk normality test

data:  x
W = 0.82236, p-value = 2.193e-09

> SEELR(96, 0.05, 7.6364)
[1] "Non rifiuto H1"      "57.9520082581075"
> if(57.9520082581075>12.2724){
+   alphaoss="è << 0.01"
+ }
> alphaoss
[1] "è << 0.01"

```

```

> ELRZ(96, 0.05, 3.841)
[1] "Non rifiuto H1"      "4.66331831300545"

> alphaoss<-1-pchisq(4.66331831300545,1)
> alphaoss
[1] 0.03081359

> # DATI TRASFORMATI DEI MALATI con Box-Cox
> landa<-0.1136
> xtras<-((x**landa)-1)/landa #creo dei nuovi dati trasformati

> shapiro.test(xtras)

      Shapiro-wilk normality test

data:  xtras
W = 0.97631, p-value = 0.07937

> SEELR(96, 0.05, 7.6364)
[1] "Non rifiuto Ho"      "2.55867774951984"
> if(2.55867774951984<5.8172){
+   alphaoss="è >0.09"
+ }
> alphaoss
[1] "è >0.09"

> ELRZ(96, 0.05, 3.841)
[1] "Non rifiuto Ho"      "0.845263501602507"

> alphaoss<-1-pchisq(0.845263501602507,1)
> alphaoss
[1] 0.3578957

> #Comandi per il qqplot
> qqnorm(x)
> qqline(x)

```

## Bibliografia e sitografia

- [1] Owen, A.B. (1988), Empirical likelihood ratio confidence intervals for a single functional, 1988, *Biometrika*, 75, pp. 237–249;
- [2] Owen, A.B. (1990), Empirical likelihood confidence regions, *Ann. Statist.*, Vol.18, No. 1, pp. 90-120.
- [3] R.B. D’Agostino e A. Belanger, “A suggestion for Using Powerful and Informative Tests of Normality”, *The American Statistician*, Vol. 44, No. 4 (Nov., 1990), pp. 316-321
- [4]”<http://www.real-statistics.com/statistics-tables/shapiro-wilk-table/>”
- [5]”<http://www.real-statistics.com/statistics-tables/kolmogorov-smirnov-table/>”
- [6]”<http://www.real-statistics.com/statistics-tables/lilliefors-test-table/>”
- [7]”<http://www.sixsigmain.it/ebook/Capu13-8.html>”
- [8]”<http://www.real-statistics.com/statistics-tables/anderson-darling-test-table/>”
- [9] L.B. Dong and D.E. Giles “An empirical likelihood ratio test for normality,” *Communications in Statistics-Simulation and Computation*, vol. 36, no. 1-3, pp.197-215, 2007.
- [10] G.Shan, A.Vexler, G.E. Wilding, and A.D. Hutson, “Simple and exact empirical likelihood ratio tests for normality based on moment relations“ , *Communications in Statistics-Simulation and Computation*, vol 40, no.1, pp.129-146, 2010.
- [11] C.S. Marange, Y. Qin, “A simple Empirical Likelihood Ratio Test for Normality Based on the Moment Constraints of a Half-Normal Distribution“, *Journal of Probability and Statistics*, Volume 2018, Article ID 8094146, 10 pages, 12 September 2018.
- [12] C.C. Lin and G.S. Mudholkar, “A simple test for normality against asymmetric alternatives“, *Biometrika*, vol.67, no. 2, pp.451-461, 1980.
- [13] A.P. Prudnikov, Y.A. Brychkov, and O.I. Marichev, *Integrals and Series*, vol.1, Gordon and Breach Science Publishers, 1986.
- [14]”*Statistics in research and development*“ , 1991 (Chapman & Hall), Roland Caucutt
- [15]”[http://www.nephromeet.com/web/eventi/GIN/dl/storico/2011/6/642-647\\_TRIPEPI\\_MASTER.pdf](http://www.nephromeet.com/web/eventi/GIN/dl/storico/2011/6/642-647_TRIPEPI_MASTER.pdf)”
- [16] “Confidence intervals for the symmetry point: an optimal cutpoint in continuous diagnostic tests“, Mónica López-Ratón, Carmen Cadarso-Suárez, Elisa M. Molanes-López, and Emilio Letónc, vol.15, pp. 179-192, 12 January 2016.

[17]“[http://www.nephromeet.com/web/eventi/GIN/dl/storico/2011/6/642-647\\_TRIPEPI\\_MASTER.pdf](http://www.nephromeet.com/web/eventi/GIN/dl/storico/2011/6/642-647_TRIPEPI_MASTER.pdf)”

[18]”<https://www.tandfonline.com/doi/full/10.1080/03610918.2010.532896?scroll=top&%20eedAccess=true&>”