

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



**MODELLI NORMALI PERTURBATI
PER MISURARE IL PICCO DI FLUSSO
INSPIRATORIO NASALE**

**PERTURBED NORMAL MODELS FOR
THE MESAUREMENT OF PEAK
NASAL INSIPIRATORY FLOW**

Relatore
Ch.mo Prof. Bruno Scarpa
Dipartimento di Scienze Statistiche

Laureanda
Evis Carapuli
Matricola N 625959

A.A 2013/2014

Dedica
alla mia famiglia

Indice

1	Introduzione	7
2	Descrizione dati	11
2.1	Materiali e Metodi	13
3	Analisi Statistica	15
3.1	Verifica di Normalità	16
3.2	Analisi grafica di osservazioni per il PNIF destro e sinistro seconda misurazione	16
3.3	Descrizione della distribuzione Skew-Normal	19
3.4	Primo metodo: Modello dei massimi per dati longitudinali unilaterali	20
3.4.1	Inferenza Statistica	21
3.4.2	Intervalli di confidenza	22
3.5	Secondo metodo: Il modello Skew-Normal	23
3.5.1	Stima di massima verosimiglianza per modelli Skew- Normal	23
3.5.2	Intervalli di confidenza	24
4	Statistiche d'ordine di una Normale Trivariata in termini di GSN	27
4.1	Analisi grafica di osservazioni per il PNIF complessivo	28
4.2	Primo metodo: Statistiche d'ordine da una distribuzione Nor- male Trivariata	30
4.3	Secondo metodo: Stima dei parametri diretta	32
5	Conclusioni	35
A	Codice sorgente e immagini	39
B	Allegati	51

Capitolo 1

Introduzione

Le deviazioni del setto nasale rappresentano delle malformazioni nasali, congenite o post-natali, frequentissime, specie nella popolazione europea dove pare abbiano un'incidenza elevatissima. Originate da turbe dello sviluppo delle strutture nasali, talora da traumi e microtraumi (frequentissimi nell'età infantile e che agiscono specie sulla porzione anteriore del setto) è stato dimostrato Corey et al. (2000) che le ostruzioni delle vie nasali sono importanti per la qualità di vita in generale (la qualità del sonno, rinosinusite, otite media e l'asma). Per questo motivo, la valutazione delle capacità respiratorie nasali è stata ed è tuttora oggetto di ricerca per molti fisiologi e rinologi sin dal 1959. La maggior parte di questi studi verte sulla misura delle variazioni di pressione durante atti respiratori normali o durante respirazione forzata. Tra i metodi più conosciuti e tuttora in uso per la registrazione delle resistenze nasali vi è la Rinomanometria (Dennis Shusterman e Berkley, 2008) considerata come metodo standard per la misurazione della resistenza respiratoria. Un'altra tecnica recente e molto diffusa è il Picco Di Flusso Inspiratorio Nasale (PNIF, vedi ad esempio Ottaviano et al. (2013) che è stato presentato per la prima volta da Youtlen (1980). Il PNIF è un metodo che permette di misurare la quantità d'aria che riesce a passare attraverso ogni singola narice. A differenza della Rinomanometria, PNIF viene considerato un metodo semplice, economico e facilmente eseguibile anche dai meno esperti. Inoltre, data la facilità d'esecuzione, potrebbe essere utilizzato direttamente dal paziente a casa propria, ovviamente dopo aver ricevuto un'accurata spiegazione sulla maniera di adoperarlo. Ciò permetterebbe un'utile valutazione quotidiana dei risultati di un'eventuale terapia.

Il setto nasale divide il naso nelle due fosse nasali, destra e sinistra ed è utile misurare il funzionamento di ciascuna delle due vie nasali, ad esempio applicando il PNIF solo su una singola narice, tenendo ermeticamente chiusa l'altra. Chiaramente i due flussi risultano correlati tra loro, ma può essere utile per i medici considerare la distribuzione del flusso nasale massimo tra i due.

Lo svolgimento di questo studio, è stato effettuato presso il Dipartimento di Neuroscienze dell'Università di Padova selezionando un campione costituito da 180 soggetti sani (infermieri, medici, pazienti senza problemi respiratori) che non prendevano medicine che potessero influenzare gli esperimenti svolti. Sono state ottenute tre inspirazioni profonde, il valore massimo viene usato come valore di PNIF. È stato misurato il PNIF di destra, di sinistra e quello complessivo. Un primo obiettivo consiste nell'ottenere la distribuzione del massimo tra flusso destro e sinistro, mentre un secondo obiettivo, completamente diverso dal primo, consiste nell'ottenere la distribuzione del massimo tra le tre osservazioni ottenute di seguito. Cioè il secondo obiettivo consiste nel valutare la distribuzione della procedura PNIF, che appunto richiede di effettuare 3 misure e poi di prenderne il massimo.

La prima parte della tesi consiste nel descrivere la raccolta dei dati e spiegare in maniera specifica la procedura eseguita per questo esperimento. Si procede con l'analisi, specificando il modello statistico parametrico che ragionevolmente contenga la distribuzione di probabilità che ha generato i dati. È stata effettuata un'analisi preliminare per verificare se la misura del PNIF unilaterale può essere descritta da un modello normale analizzando visivamente l'istogramma di frequenza dei dati raccolti, il diagramma a scatola con baffi e il Q-Q plot. Si assume che i dati si distribuiscono normalmente e si procede con lo svolgimento dell'analisi seguendo il nostro obiettivo, ossia di ottenere la distribuzione esatta del massimo tra le due variabili unilaterali. Per ottenere tale massimo si procede seguendo due diverse strade: da una parte, ipotizzando la normalità delle osservazioni, sapendo che il massimo di due osservazioni normali si distribuisce come una Normale Assimetrica e a partire dai parametri stimati della distribuzione delle singole osservazioni si ottiene la distribuzione del massimo. Dall'altra parte si ottiene direttamente il massimo delle osservazioni ipotizzando che si distribuiscono come Skew-Normal, stimiamo i parametri direttamente. L'analisi per affrontare il problema del massimo tra destra e sinistra viene effettuata nel terzo capitolo invece nel capitolo quattro si esegua l'analisi tra le tre osservazioni. Il primo metodo analizza i dati partendo dalle osservazioni normali unilaterali destra (Y_D) e sinistra (Y_S). La congiunta delle due variabili è una distribuzione normale bivariata con le marginali standard e coefficiente di correlazione ρ . Sappiamo che la distribuzione del massimo tra le osservazioni controlaterali è una Skew-Normal (si veda Azzalini, 2014 per una trattazione organica dei risultati sulla distribuzione Normale asimmetrica e Loperfido (2002) per una rassegna specifica sui problemi di massimo tra osservazioni controlaterali), con parametro di posizione, di scala, e parametro di forma $\psi = \sqrt{(1 - \rho)/(1 + \rho)}$.

Un secondo metodo per ottenere la distribuzione dei massimi consiste nel stimare direttamente i parametri della distribuzione del massimo delle singole componenti, come visto precedentemente, si distribuisce come una SN. La stima dei parametri si effettua via massima verosimiglianza, usando

l'algoritmo di ottimizzazione descritto da Azzalini (2014). Infine, per ognuno dei metodi, vengono calcolati gli intervalli di confidenza per il parametro di forma della distribuzione Skew-Normal e si effettua il confronto tra i metodi.

Nel quarto capitolo si affronta il problema di determinare la distribuzione della procedura PNIF che viene raccolto come il massimo di tre osservazioni consecutive alle medesime condizioni, vogliamo trovare la distribuzione del pnif, a partire dalle singole distribuzioni delle inspirazioni. Si considerano le tre misure ottenute per il PNIF complessivo (effettuato usando entrambe le narici) e si ottiene il massimo. Analogamente al caso precedente anche qui si deve ottenere la distribuzione di un massimo, ma non più tra due ma tra tre distribuzioni Normali. Jamalizadeh e Balakrishnan (2009) hanno mostrato che il massimo di tre normali correlate tra loro si distribuisce come una GSN (Generalized Skew-Normal, Azzalini, 2014). In entrambi i casi si assume che le singole variabili sono normali e in entrambi i casi quindi usiamo il risultato che il max tra tre normali è una GSN. Nel primo caso stimiamo i parametri dai dati raccolti dalle singole 3 variabili e poi otteniamo i parametri della GSN usando le relazioni tra i parametri delle 3 variabili e quelli della distribuzione del massimo dati da Jamalimazad e Balakrishnan (2009). Nel secondo caso invece stimiamo direttamente i parametri della distribuzione del massimo (che è una GSN) partendo dai massimi osservati. Le distribuzioni ottenute usando i due metodi vengono poi confrontate e vengono proposte alcune riflessioni sulle stime ottenute.

Capitolo 2

Descrizione dati

La valutazione delle capacità respiratorie nasali è stata ed è tuttora oggetto di ricerca per molti fisiologi e rinologi sin dal 1959. La maggior parte di questi studi verteva sulla misura delle variazioni di pressione durante atti respiratori normali o durante respirazione forzata. Tra le metodiche più conosciute e tuttora in uso per la registrazione delle resistenze nasali vi è la rinomanometria (RM), considerata come metodo standard per la misurazione della resistenza respiratoria. Nonostante sia una tecnica accettata e sicura, che possiede un limitato grado d'errore, presenta alcuni limiti: necessita di un'apparecchiatura alquanto costosa, richiede molta esperienza e dispendio di tempo nell'utilizzo da parte del personale medico. Inoltre, non è facilmente trasportabile per il numero degli elementi che la compongono, come mostrato in Figura 2.1.



Fig. 2.1: Rinomanometro (Shusterman, 2008)

Da anni, ricercatori e clinici a livello internazionale conducono studi con lo scopo finale di trovare e perfezionare una metodica efficiente ed efficace, economica e semplice da usare, che consenta di valutare la pervietà nasale. Negli ultimi anni è stato introdotto un nuovo metodo per lo studio della pervietà nasale.

La misurazione del picco di flusso inspiratorio nasale (PNIF, si veda ad es. Ottaviano et al. (2012)) è un metodo semplice, economico e facilmente eseguibile anche dai meno esperti. È adatto per misurazioni seriali e per uso domestico come mostrato in Figura 2.2.



Fig. 2.2: PNIF utilizzato da un paziente Ottaviano et al. (2012))

Inoltre, è stato dimostrato che PNIF è adatto per la misurazione della resistenza delle vie aeree nasali ottenendo risultati simili a quelli della rimanometria Holmstrom et al. (1990). Recentemente, i risultati derivati dalle misurazioni su adulti e bambini sono stati usati in varie pubblicazioni, permettendo di applicare tali risultati anche alla settoplastica. Tutti i dati PNIF raccolti e pubblicati precedentemente a questo studio, sono stati ottenuti analizzando entrambe le narici contemporaneamente, per questo si presenta la necessità di svolgere uno studio unilaterale.

2.1 Materiali e Metodi

Il nostro studio è stato effettuato su un campione costituito da soggetti sani, senza sintomi nasali (ostruzione, rinorrea, ecc.), non fumatori, non asmatici e senza precedenti trattamenti chirurgici al distretto naso-sinusale. Prima dell'inizio del test sono state chieste informazioni circa l'età, il sesso, l'altezza ed eventuali medicinali usati.

Per il misuramento del PNIF è stato usato il Portable Youlten Peak Flow Meter Clement Clark International Ottaviano et al. (2013), il quale è costituito da una parte frontale, la quale è una semplice maschera facciale, e da una parte finale che ha lo scopo di misurare la resistenza respiratoria, come mostrato in Figura 2.3.



Fig. 2.3: **a.** Misuratore Clement Clark International; **b.** Maschera facciale (Fonte: Ottaviano et al. 2012)

La maschera facciale è collegata a un misuratore, che il paziente deve applicare sulla faccia coprendo il naso, ma senza minimamente comprimerlo o toccarlo. Dopo ogni misurazione la maschera va pulita con dell'alcohol. Tutti i soggetti sono stati seduti durante l'esperimento e gli è stato chiesto di inalare in maniera energica attraverso il naso, a bocca chiusa, partendo da

un'espiazione profonda. Il picco di flusso sarà quindi segnalato dal cursore situato all'interno del misuratore. Dalle tre misurazioni eseguite, il valore massimo viene usato come valore di PNIF. La seconda fase delle misurazioni consiste nel bloccare una delle narici, usando un nastro adesivo, e ripetere le tre inalazioni come precedentemente fatto. Anche in questo caso, è stata scelta la misurazione di valore massimo, rispettivamente per la narice destra e quella sinistra.

Per lo svolgimento degli esperimenti sopra esposti è stato selezionato, dal Dipartimento di Neuroscienze dell'Università di Padova (medici, pazienti senza problemi respiratori, infermieri ecc), un gruppo di 180 soggetti sani (che non prendevano medicine che potessero influenzare gli esperimenti svolti) con un'età compresa tra i 13 e gli 80 anni. A tutti i soggetti è stato chiesto di compilare il modulo SNOT 22 presente in Appendice B. Inoltre gli è stato chiesto se presentassero qualche tipo di bloccaggio nasale, se fossero fumatori e/o asmatici, se avessero subito interventi chirurgici o se presentassero qualsiasi altro problema legato all'apparato respiratorio. Partendo dal nostro insieme di dati, ciò che si vuole fare è: addattare un modello probabilistico che ci permetta di riassumere efficacemente i dati e generalizzare il comportamento per il misuramento del PNIF massimo tra narice destra e sinistra. In questo lavoro di tesi stiamo trattando variabili casuali continue dove la struttura delle osservazioni sperimentali vuole essere spiegata tramite un modello statistico per poter scoprire la distribuzione esatta del massimo di PNIF.

Per lo sviluppo di questo studio è stato usato il software R e in particolare il Package `sn`.

Capitolo 3

Analisi Statistica

Per procedere con l'elaborazione dei dati si è inizialmente provveduto all'acquisizione del data-set costituito da 180 volontari sani che sono stati coinvolti in questo studio. La misurazione di PNIF unilaterale, potrebbe essere un metodo utile per lo studio della pervietà della singola narice per aiutare la diagnosi della patologia nasale, soprattutto quando risulta necessario stimare/valutare gli effetti funzionali delle deviazioni unilaterali del setto nasale o in qualsiasi caso dove ci sia un sospetto riguardo all'occlusione nasale unilaterale. Questo studio pilota fornisce dati regolatori iniziali sul PNIF unilaterale.

In questo capitolo presentiamo i modelli probabilistici che pensiamo possano adattarsi meglio al nostro campione dei dati e si discuteranno le situazioni in cui, dato un campione di n osservazioni $y = (y_1, \dots, y_n)$, si vuole specificare un modello statistico parametrico che ragionevolmente contenga la distribuzione di probabilità che ha generato i dati e le distribuzioni di probabilità di interesse. Questi modelli di probabilità appartengono alla famiglia di distribuzioni simmetriche perturbate, la quale può essere abbreviata come classe SS (dall'inglese Skew-Symmetric). Uno dei primi passi da compiere nella specificazione di un modello statistico è l'esame del comportamento dei dati, prestando particolare attenzione a:

- La loro tipologia: quantitativa o qualitativa, quantitativa discreta o continua.
- I loro aspetti notevoli: code corte o pesanti, asimmetria, curtosi, posizione e variabilità.

3.1 Verifica di Normalità

È utile un'analisi preliminare dei dati per verificare se la misura del PNIF può essere descritta da un modello normale. Le distribuzioni normali, sono il più importante modello teorico continuo per modellare variabili continue. Tali distribuzioni sono adatte a modellare variabili la cui distribuzione empirica presenta una forma campanulare simmetrica. Un modo esplicativo per valutare qualitativamente forti scostamenti della normalità è quello di analizzare visivamente l'istogramma di frequenza dei dati raccolti e il *box-plot*. Il diagramma a scatola con baffi è uno strumento grafico utile per valutare le caratteristiche della distribuzione da cui provengono le osservazioni. Questi diagrammi non sono solo utili per verificare la posizione e la variabilità dei dati, ma anche per valutare l'asimmetria e la pesantezza delle code. Infine, si accenna alla possibilità di ricorrere ad una versione lisciata dell'istogramma, vale a dire la densità stimata secondo il metodo nucleo Bowman e Azzalini (1997). Un altro strumento grafico utile è il Q-Q plot.

3.2 Analisi grafica di osservazioni per il PNIF destro e sinistro seconda misurazione

In questo paragrafo portiamo i grafici ottenuti usando come campione dei dati la seconda misurazione del picco di flusso inspiratorio nasale sinistro e destro, le altre analisi effettuate per le altre variabili rimanenti verranno dimostrate in appendice A. Dagli istogrammi presentati in Figura 3.1-a e 3.1-b, si nota una leggera asimmetria, e quindi i nostri dati non sembrano distribuirsi normalmente tuttavia, nel seguito si assumerà come prima approssimazione la normalità delle singole osservazioni. Dall'analisi dei risultati poi si vedrà se e quanto quest'ipotesi ha influenzato i risultati.

Si osserva dal diagramma in Figura 3.2-a e 3.2-b che le lunghezze dei baffi e le altezze dei due rettangoli non sono simili tra loro. Dal grafico in Figura 3.1-c e 3.2-c si nota che i punti non tendono ad allinearsi lungo la bisettrice e questo non comporta un buon adattamento. Quindi dall'analisi grafica effettuata si conferma che i nostri dati sono asimmetrici.

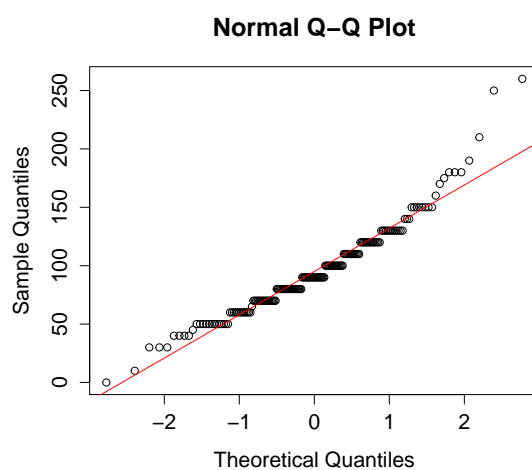
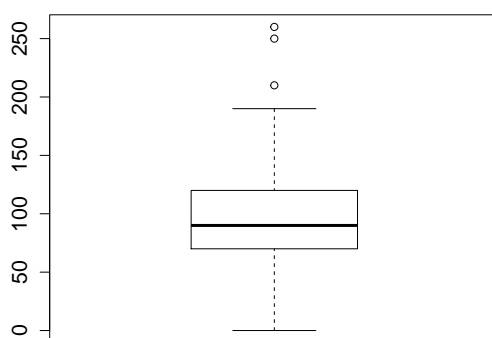
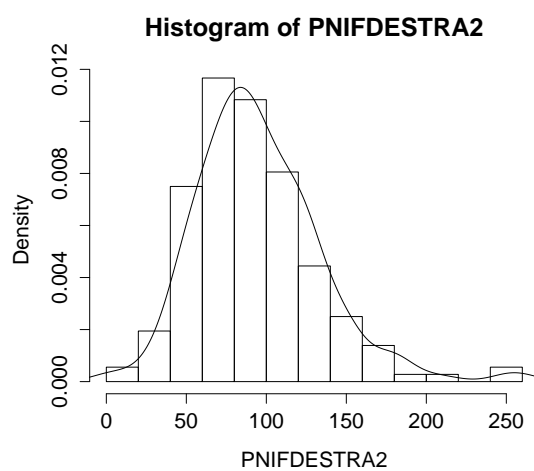


Fig. 3.1: **a.** Istogramma del PNIF.DX.2; **b.** Diagramma a scatola con baffi; **c.** Diagramma Q-Q normal;

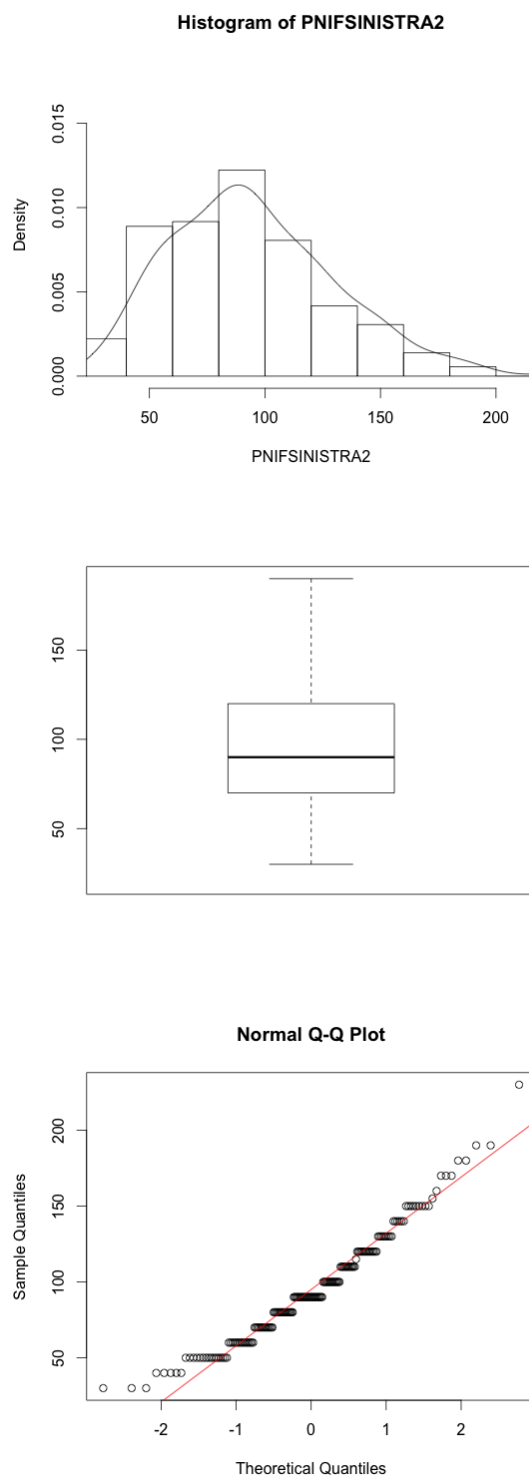


Fig. 3.2: a. Istogramma del PNIF.SX.2; b. Diagramma a scatola con baffi;c. Diagramma Q-Q normal;

Il nostro scopo è quello di ottenere la distribuzione del massimo tra le due osservazioni unilaterali. Possiamo seguire due strade alternative per calcolare quest'ultima: il primo metodo assume che le variabili Y_S e Y_D (Y sinistra e destra rispettivamente), si distribuiscono come una normale con media μ e varianza σ e con il vincolo della correlazione $\rho > 0$. La congiunta di (Y_S, Y_D) si distribuisce come una bivariata normale con il vettore della media μ , matrice di varianze e covarianza Σ e il coefficiente di correlazione $-1 < \rho < 1$. Inoltre sappiamo che il $\max(Y_S, Y_D) \sim SN(\epsilon, \omega, \psi)$ Azzalini (2014) e Loperfido (2002). Nel paragrafo seguente viene spiegato la funzione Skew-Normal (SN).

3.3 Descrizione della distribuzione Skew-Normal

In questi ultimi anni la distribuzione Normale Asimmetrica, indicata con SN (dall'acronimo inglese Skew-Normal) è stata largamente studiata ed utilizzata sia per la sua buona capacità di adattamento ai dati che presentano leggere asimmetrie, sia perché le sue caratteristiche statistiche la rendono molto simile alla distribuzione Normale, e quindi facile da analizzare. Definiamo ora la distribuzione di probabilità Normale Asimmetrica facendo riferimento al Azzalini (1985, Lemma 1.1). Siano $f_0 = \phi$ e $G = \Phi$ la funzione di densità e di ripartizione di una normale standardizzata. Se una v.c. Z ha funzione di densità:

$$\phi(z; \psi) = 2\phi(z)\Phi(\psi z) \quad (-\infty < z < \infty) \quad (3.1)$$

allora si distribuisce come una normale asimmetrica con parametro di forma ψ e scriveremo che $Z \sim SN(\psi)$ dove ψ regola l'intensità dell'asimmetria, mentre $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ e $\Phi(\psi z) = \int_{-\infty}^{\psi z} \phi(t)dt$. L'espressione in (3.1) si generalizza introducendo dei parametri di posizione e scala. Sia $Z \sim SN(\psi)$, $\xi \in R$ e $\omega \in R^+$. Cioè, definita $Y = \xi + \omega Z$, la densità di Y sarà:

$$\phi(z; \xi, \omega, \psi) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left(\psi \frac{y - \xi}{\omega}\right) \quad (-\infty < y < \infty) \quad (3.2)$$

e scriveremo che $Y \sim SN(\xi, \omega, \psi)$. Questo garantisce la costruzione di un'intera famiglia di distribuzioni asimmetriche a partire dalla perturbazione tramite $G(\psi z)$ di una funzione di densità simmetrica f_0 .

Si riportano in seguito alcune importanti proprietà dimostrate in Azzalini (1985) in cui risultano evidenti i legami tra la distribuzione normale asimmetrica e la distribuzione normale. Da un punto di vista teorico, la classe normale asimmetrica ha il vantaggio di essere matematicamente trattabile e di godere di molte proprietà tipiche della distribuzione normale, tra cui:

- se $\psi = 0$ allora $Z \sim SN(0) \sim N(0, 1)$
- se $Z \sim SN(\psi)$, allora $-Z \sim SN(-\psi)$

- se $\psi \rightarrow \infty$ ci si riconduce alla mezza normale, avente densità $2\phi(z)$ per $z \geq 0$
- se $Z \sim SN(\psi)$, allora $Z^2 \sim \chi_1^2$
- per ogni fissato ψ la 3.1 è fortemente unimodale, ossia $\log \phi(z; \psi)$ è una funzione concava di z .

La distribuzione Normale risulta quindi un elemento interno a questa famiglia di distribuzioni.

Il secondo metodo, assume che il massimo delle osservazioni di destra e sinistra si distribuiscono come una Skew-Normal e effettua direttamente la stima dei parametri. Nelle pagine a seguire verranno analizzati e confrontati i due diversi metodi più in specifico.

3.4 Primo metodo: Modello dei massimi per dati longitudinali unilaterali

In questo paragrafo riportiamo la modellazione dei massimi per i dati unilaterali del setto nasale sapendo che il massimo tra le osservazioni unilaterali si distribuisce come una normale asimmetrica (si veda Azzalini, 2014 per una trattazione organica dei risultati sulla distribuzione Normale asimmetrica e Loperfido (2002) per una rassegna specifica sui problemi di massimo tra osservazioni controlaterali). È stata scelta la seconda misurazione dell'inspirazione (PNIF.Destra.2 e PNIF.Sinistra.2) come variabile per dimostrare l'analisi unilaterale di questo capitolo. Tutte le altre analisi effettuate per le variabili rimanenti sono riportate in appendice A. Si assume la normalità delle osservazioni laterali $y_S = (Y_{1S}, \dots, Y_{pS})^T$ e $y_D = (Y_{1D}, \dots, Y_{pD})^T$ dove, y_S e y_D sono i vettori corrispondenti rispettivamente alle osservazioni laterali sinistra e destra in occasioni. La distribuzione corrispondente alla congiunta di questi due vettori è quindi la normale bivariata

$$\begin{pmatrix} Y_S \\ Y_D \end{pmatrix} \sim N_{2p} \left\{ \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \Omega & \Omega - \Lambda \\ \Omega - \Lambda & \Omega \end{pmatrix} \right\}$$

dove $\Omega - \lambda/2$ è la matrice definita positiva e λ è una matrice diagonale dove gli elementi della diagonale $\lambda_1, \dots, \lambda_p$ sono dei scalari non-negativi.

L'obiettivo di questa sezione è quello di trovare la distribuzione del massimo di un vettore casuale bivariato $\max(Y_S, Y_D)$. In Loperfido (2002, teorema 3.1) si dimostra che, una combinazione lineare del massimo di un vettore casuale normale bivariato si distribuisce come una normale asimmetrica con i parametri specificati dalla seguente formula:

$$\max(Y_S, Y_D) \sim SN(\epsilon, \tau^2, \psi) \quad (3.3)$$

dove $\epsilon = \mu$, $\tau^2 = \Omega$ e $\psi = \sqrt{((1 - \rho)/(1 + \rho))}$.

Nel paragrafo seguente si dimostra l'analisi effettuata usando i nostri dati.

3.4.1 Inferenza Statistica

In questa sezione trattiamo i dati relativi alla seconda inspirazione. Si procede con la stima dei parametri assumendo la normalità delle variabili Y_S e Y_D con la corrispondente distribuzione: $Z_1 = \max(Y_S, Y_D) \sim SN(\epsilon, \omega, \psi)$ Loperfido (2002) dove il parametro locale è uguale alla media comune della distribuzione normale bivariata, il parametro di scala è uguale alla varianza comune e il parametro di forma è $\psi = \sqrt{((1 - \rho)/(1 + \rho))}$ dove $\tau^2 = \Omega$ e Ω è la varianza della distribuzione normale bivariata. In Figura (3.3) si riporta il grafico corrispondente alla funzione di densità della nostra distribuzione $SN(\epsilon = 95.40275, \omega = 37.60164, \psi = 0.4095726)$ con i rispettivi parametri e nella Figura (3.4) è riportato l'istogramma delle Z_1 osservate, cioè' del massimo tra osservazione sinistra e destra, ottenuto dai dati così generati con sovrapposta la densità della distribuzione normale asimmetrica.

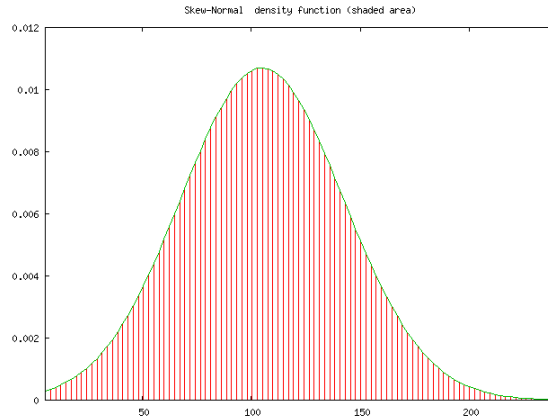


Fig. 3.3: funzione della densità

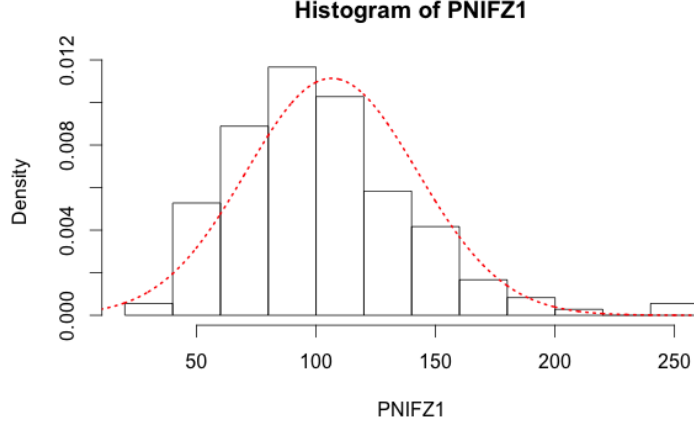


Fig. 3.4: istogramma del massimo tra inspirazione destra e sinistra con sovrapposta la funzione di densità

3.4.2 Intervalli di confidenza

In seguito si riporta il calcolo dell'intervallo di confidenza per il parametro di forma ψ come Hadded and Provost (2011) hanno proposto il calcolo dell'intervallo di confidenza esatto per il parametro di forma. Sia, $Z_1 = (Y_S, Y_D)$, la nostra distribuzione bivariata normale con media comune 0, varianza comune 1 e coefficiente di correlazione $\rho \in (-1, 1)$. La costruzione dell'intervallo di confidenza usa due variabili casuali $D_+ = \sum (Y_{Si} + Y_{Di})^2$ e $D_- = \sum (Y_{Si} - Y_{Di})^2$ dove Y_{Si} e Y_{Di} sono le variabili di sinistra e destra, indipendenti tra loro, per ogni $i = 1, \dots, n$. Siccome $Y_{Si} + Y_{Di}$ e $Y_{Si} - Y_{Di}$ sono indipendenti tra loro, si ha che, $Y_{Si} + Y_{Di} \sim N(0, 2(1 + \rho))$ e $Y_{Si} - Y_{Di} \sim N(0, 2(1 - \rho))$, da questo deriva $D_+(1 - \rho)/D_-(1 + \rho) \sim F_{n-1, n-1}$ dove $F_{n-1, n-1}$ è la distribuzione di Fisher con $(n - 1, n - 1)$ gradi di libertà. Quindi, un intervallo di confidenza esatto per il parametro di forma ψ è:

$$\left\{ \psi \in [0, \infty) : \sqrt{\frac{D_-}{D_+} F_{n-1, n-1}(1 - \alpha/2)} < \psi < \sqrt{\frac{D_-}{D_+} F_{n-1, n-1}(\alpha/2)} \right\} \quad (3.4)$$

dove $F_{n,n}(p)$, con $p \in (0, 1)$, rappresenta il p -esimo quantile della distribuzione di Fisher con $(n - 1, n - 1)$ gradi di libertà.

Calcolando l'intervallo di confidenza per il parametro di forma ψ usando il metodo proposto da Hadded and Provost (2011), il risultato ottenuto è: $\psi \in [0.353; 0.474]$.

3.5 Secondo metodo: Il modello Skew-Normal

Il secondo metodo usato consiste nello stimare direttamente i parametri delle distribuzioni di massimi usando il metodo della massima verosimiglianza (vedi ad es. Azzalini, 2014).

3.5.1 Stima di massima verosimiglianza per modelli Skew-Normal

La distribuzione Normale Asimmetrica, come già accennato, gode di buone proprietà, che le consentono di essere facilmente trattata dal punto di vista matematico. Dal lato pratico, presenta alcuni problemi legate alle stime dei parametri che la caratterizzano, in particolare, quelle relative allo stimatore del parametro di forma. Per campioni di dimensioni moderate, lo stimatore di massima verosimiglianza è infinito con probabilità positiva. In merito alla presenza di problemi legati alle stime dei parametri, sono stati proposti diversi metodi per rendere stimabile in maniera finita il parametro che regola la forma della distribuzione.

Una soluzione, proposta da Azzalini (2014), consiste nell'avviare l'algoritmo di ottimizzazione e farlo terminare quando si osserva che la log-verosimiglianza raggiunge un valore tale da non manifestare successive crescite significative della funzione stessa. Il valore così determinato costituisce la stima di massima verosimiglianza. Si tratta, tuttavia, di un metodo piuttosto arbitrario, in quanto non è definito in maniera univoca un criterio di arresto, e la natura dei dati potrebbe incidere molto sullo stimatore, fornendo stime poco attendibili. Usando l'algoritmo di ottimizzazione otteniamo i seguenti risultati: $SN(\epsilon = 106.3033, \omega = 37.6489, \psi = 0.6809)$. Confrontando con il metodo precedente osserviamo un piccolo aumento del valore di parametro di forma. Portiamo in seguito i grafici corrispondenti alla funzione di densità e l'istogramma con sovrapposta la funzione di densità calcolata.

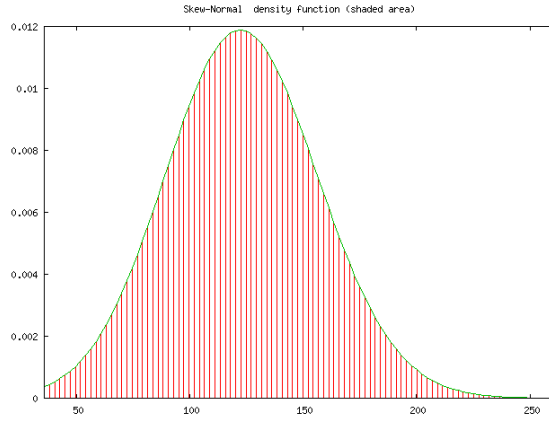


Fig. 3.5: funzione della densità

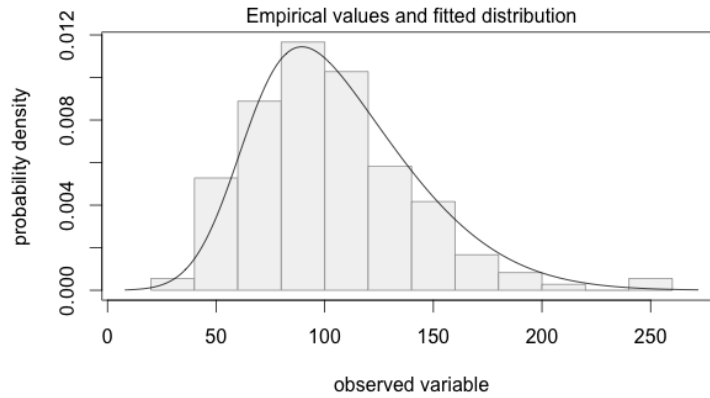


Fig. 3.6: istogramma con sovrapposta la funzione di densità

3.5.2 Intervalli di confidenza

Una volta ottenuta la stima dei parametri è interessante ottenere una stima dell'intervallo di confidenza. Come riportato ad esempio in Azzalini (2014), per ottenere le stime sono state utilizzate le verosimiglianze profilo per ogni parametro d'interesse, per cui si è ottenuto un intervallo di confidenza basato sul *TRV* (test del rapporto di verosimiglianza) a livello 0.95. Una via semplice per reperire la verosimiglianza profilo per un parametro di interesse ψ consiste nel sostituire, nella verosimiglianza complessiva $L(\psi, \omega)$, il parametro di disturbo $\hat{\omega}_\psi$ con la sua stima di massima verosimiglianza per ψ fissato. In un modello statistico parametrico con funzione di verosimiglianza $L(\Theta)$, in cui $\Theta = (\psi, \omega)$ e ψ è il parametro d'interesse, si definisce la

verosimiglianza profilo per ψ la funzione

$$LP(\psi) = L(\psi, \hat{\omega}_\psi) \quad (3.5)$$

$\hat{\omega}_\psi$ è la Stima di Massima Verosimiglianza di ψ nel sotto-modello con ψ fissato e si indica con $lP(\psi) = \log LP(\psi)$ la funzione di log-verosimiglianza profilo.

L'intervallo di confidenza, basato sul test rapporto di verosimiglianza, viene creato considerando i valori di $\hat{\psi}$ che soddisfano la disequazione

$$2(lP(\hat{\psi}) - lP(\psi)) < \chi_{0.95,1}^2 \quad (3.6)$$

Il risultato dell'intervallo di confidenza è $[-0.020, 1.382]$.

Capitolo 4

Statistiche d'ordine di una Normale Trivariata in termini di GSN

Poichè abbiamo visto che il PNIF viene raccolto come il massimo di tre osservazioni consecutive alle medesime condizioni, vogliamo ora trovare la distribuzione del PNIF, a partire dalle singole distribuzioni delle ispirazioni. Ipotizziamo che ciascuna ispirazione sia normale. Questo sappiamo non essere molto vero, ma come prima approssimazione lo diciamo. C'è un risultato di Jamalizadeh e Balakrishnan (2009) che dice che il massimo delle marginali di una normale trivariata con correlazione uguale e pari a ρ è una $GSN(\alpha_1, \alpha_2, \rho)$ dove α_1 e α_2 sono i parametri di forma. La GSN era stata introdotta da Arellano-valle et al.(2004) con la seguente funzione di densità

$$f(x; \alpha_1, \alpha_2) = 2\phi(x)\Phi\left(\frac{\alpha_1 x}{\sqrt{1 + \alpha_2 x^2}}\right) \quad (4.1)$$

dove, $-\infty < x < \infty$ e α_1, α_2 sono i parametri di forma con $\alpha_2 \geq 0$. Inoltre, se $\alpha_2 = 0$ la funzione di densità corrisponde alla distribuzione $SN(0, 1, \alpha_1)$ e se anche $\alpha_1 = 0$ otteniamo la distribuzione normale standard $N(0, 1)$. Una proprietà importante della distribuzione $GSN(x; \alpha_1, \alpha_2)$ è la seguente: la variabile X con funzione di densità (4.1) si può rappresentare in forma mista usando le variabili della distribuzione Skew-Normal (SN) e i parametri di forma della distribuzione $N(a_1, a_2)$. Riferendosi ai nostri dati, l'analisi si effettua basandosi sulla distribuzione $GSN(x; \alpha_1, \alpha_2, \rho)$ Jamalizadeh e Balakrishnan (2009) derivante dalle statistiche d'ordine di una distribuzione normale trivariata $(X_1, X_2, X_3) \sim N(0, \Sigma)$ dove X_1, X_2, X_3 sono le variabili corrispondenti alle misure del PNIF complessivo e ρ è la correlazione. Si organizzano i dati in ordine crescente di grandezza,

$X_{1:3} = \min(X_1, X_2, X_3)_{2:3:3} \leq X_{2:3} \leq X_{3:3} = \max(X_1, X_2, X_3)$, ottenendo così il massimo di $(X_1, X_2, X_3)^T = X_{3:3}$ con funzione di densità (pdf)

$$f(t; \Sigma) = \frac{1}{\sigma_1} a(\theta_1) \phi_{GSN}\left(\frac{t}{\sigma_1}; \theta_1\right) + \frac{1}{\sigma_2} a(\theta_2) \phi_{GSN}\left(\frac{t}{\sigma_2}; \theta_2\right) + \frac{1}{\sigma_3} a(\theta_3) \phi_{GSN}\left(\frac{t}{\sigma_3}; \theta_3\right)$$

dove, $t \in \mathbb{R}$, ϕ_{GSN} denota la funzione di densità della $GSN(\theta)$ e

$$a(\theta) = \frac{1}{2\pi} \cos^{-1}(-(\rho + \alpha_1 \alpha_2)) / \sqrt{1 + \lambda_1^2} \sqrt{1 + \lambda_2^2}$$

con

$$\begin{aligned} \theta_1 &= \left\{ \frac{\frac{\sigma_1}{\sigma_2} - \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \frac{\frac{\sigma_1}{\sigma_3} - \rho_{13}}{\sqrt{1 - \rho_{13}^2}}, \rho_{23} \right\}^T, \\ \theta_2 &= \left\{ \frac{\frac{\sigma_2}{\sigma_1} - \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \frac{\frac{\sigma_2}{\sigma_3} - \rho_{23}}{\sqrt{1 - \rho_{23}^2}}, \rho_{13} \right\}^T, \\ \theta_3 &= \left\{ \frac{\frac{\sigma_3}{\sigma_1} - \rho_{13}}{\sqrt{1 - \rho_{13}^2}}, \frac{\frac{\sigma_3}{\sigma_2} - \rho_{23}}{\sqrt{1 - \rho_{23}^2}}, \rho_{12} \right\}^T. \end{aligned}$$

In questo capitolo, si effettua l'analisi prendendo in considerazione tre variabili corrispondenti a tre misurazioni del PNIF complessivo, con lo scopo di ottenere la distribuzione esatta di un massimo tra tre distribuzioni normali e verrà mostrata l'analisi grafica effettuata solo della seconda misurazione PNIF e in appendice A si presenteranno le altre analisi effettuate per il PNIF destro e sinistro.

4.1 Analisi grafica di osservazioni per il PNIF complessivo

Inizialmente si è effettuato un esame dei dati per verificare se questi sono distribuiti normalmente, usando stimatori di densità non-parametrici e analisi grafica. Dall'istogramma presentato in Figura 4.1 si vede una certa asimmetria e anche una pancia un po' più grossa di quella di una normale. Il boxplot mostrato in Figura 4.2 sembra di più quello di una normale.

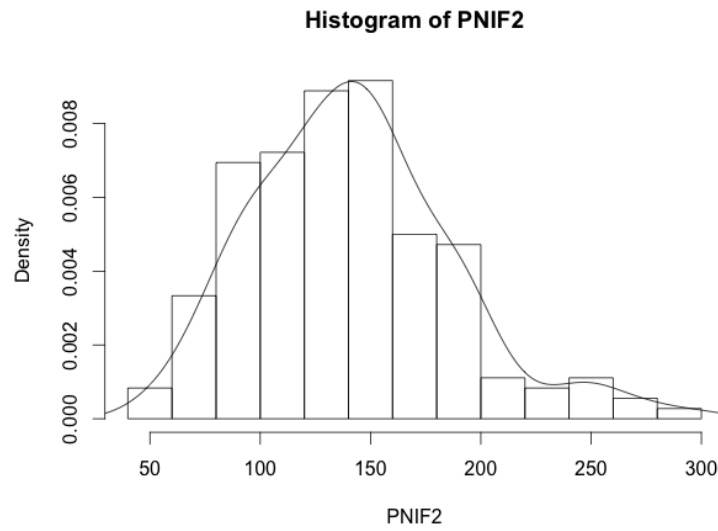


Fig. 4.1: istogramma per le osservazione contenute in PNIF.2

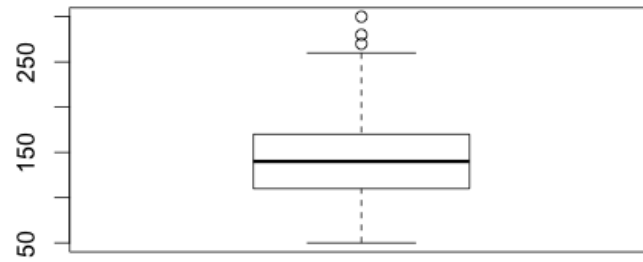


Fig. 4.2: diagramma a scatola con baffi per le osservazione contenute in PNIF.2

Dalla Figura 4.3 si può osservare che i punti tendono ad allinearsi lungo la bisettrice, il che significa che l'adattamento è buono.

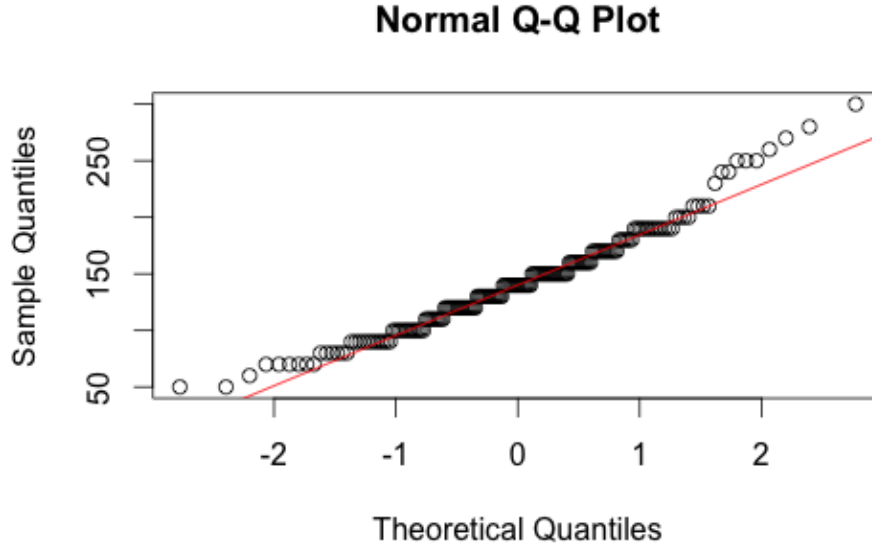


Fig. 4.3: diagramma Q-Q Normali per le osservazione contenute in PNIF.2

Dopo aver verificato la normalità, siamo interessati a fare inferenza statistica sul nostro campione, cioè ad utilizzare le statistiche calcolate sui dati per poi stimare i valori dei corrispondenti parametri. Analogamente al capitolo precedente per effettuare la stima dei nostri parametri si usano due strade: in entrambi i casi, si ipotizza che le singole variabili sono normali e si utilizza il risultato precedentemente mostrato, ossia che il max tra tre normali è una GSN. Nel primo caso vengono stimati i parametri dai dati raccolti dalle singole 3 variabili e poi si ottengono i parametri della GSN usando le relazioni tra i parametri delle 3 variabili e quelli della distribuzione del massimo dati da Jamalizadeh e Balakrishnan (2009). Nel secondo caso invece vengono stimati direttamente i parametri della distribuzione del massimo (che è una GSN) partendo dai massimi osservati. Nelle pagine a seguire verranno analizzati e confrontati i due metodi più in specifico.

4.2 Primo metodo: Statistiche d'ordine da una distribuzione Normale Trivariata

Si prendono in considerazione i dati complessivi per le 3 misurazioni del Picco di Flusso Inspiratorio Nasale (PNIF). La prima misurazione viene chiamata $PNIF_1$, la seconda $PNIF_2$ e la terza $PNIF_3$ si suppone che, la congiunta

di queste tre variabili sia una normale trivariata: $(X_1, X_2, X_3)^T \sim N_3(0, \Sigma)$ con media = 0 e la matrice di varianza-covarianza definita positiva.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \quad (4.2)$$

Dall'articolo di Jamalizadeh e Balakrishnan (2009) sappiamo che il massimo di tre variabili normali, che chiamiamo X , dove, $X = \max(X_1, X_2, X_3)$ si distribuisce come una GSN con funzione di densità

$$\phi_{GSN}(x; \alpha_1, \alpha_2, \rho) = c(\alpha_1, \alpha_2, \rho) \phi(x) \Phi_2(\alpha_1 x, \alpha_2 x, \rho), z \in \mathbb{R}$$

dove, $\alpha_1, \alpha_2 \in \mathbb{R}$, $|\rho| < 1$ e $\Phi_2(\alpha_1 x, \alpha_2 x, \rho)$ denota la funzione di ripartizione (cdf) di $N_2(0, 0, 1, 1, \rho)$ inoltre la $c(\alpha_1, \alpha_2, \rho) = \frac{1}{a(\alpha_1, \alpha_2, \rho)}$.

A partire dai vincoli dove, $\sigma_1 = \sigma_2 = \sigma_3 = 1$ e $\rho_{12} = \rho_{13} = \rho_{23} = \rho^*$ otteniamo le seguenti stime dei parametri $\hat{\rho} = \frac{\rho^*}{1+\rho^*}$, $\hat{\alpha}_1 = \sqrt{\frac{1-\rho^*}{1+\rho^*}}$ e $\hat{\alpha}_2 = \sqrt{\frac{1-\rho^*}{1+\rho^*}}$.

Dopo aver stimato tutti i parametri effettuiamo il grafico della nostra funzione di densità sovrapposta all'istogramma dei dati Normali standardizzati.

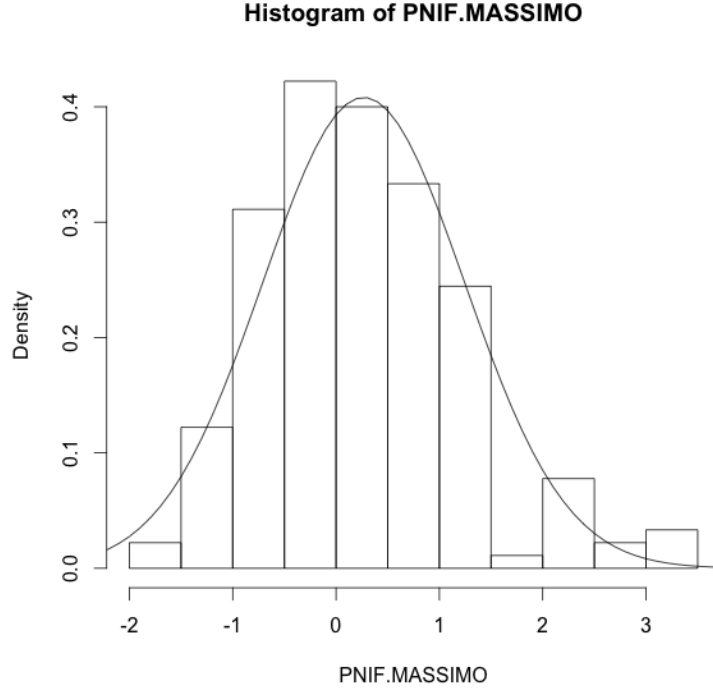


Fig. 4.4: istogramma con sovrapposta la funzione di densità di GSN

4.3 Secondo metodo: Stima dei parametri diretta

Per effettuare la stima dei parametri, questo secondo metodo ci porta ad usare la massima verosimiglianza, cioè, partendo dalla nostra funzione di densità,

$$f(t; \Sigma) = \frac{1}{\sigma_1} a(\theta_1) \phi_{GSN} \left(\frac{t}{\sigma_2}; \theta_2 \right) + \frac{1}{\sigma_2} a(\theta_2) \phi_{GSN} \left(\frac{t}{\sigma_2}; \theta_2 \right) + \frac{1}{\sigma_3} a(\theta_3) \phi_{GSN} \left(\frac{t}{\sigma_3}; \theta_3 \right), t \in \mathbb{R}$$

per indipendenza delle componenti abbiamo $L(t; \alpha_1, \alpha_2, \rho) = c(t)f(t; \Sigma)$.

La corrispondente funzione di log-verosimiglianza è: $l(\alpha_1, \alpha_2, \rho) = \log(L(\alpha_1, \alpha_2, \rho))$, quindi, il massimo di $L(\alpha_1, \alpha_2, \rho)$ si ottiene minimizzando $l(\alpha_1, \alpha_2, \rho)$. Dopo aver effettuato i calcoli, usando i nostri dati, otteniamo le seguenti stime. $\hat{\alpha}_1 = 0.4083673$ $\hat{\alpha}_2 = 0.4083673$ $\hat{\rho} = 0.4166181$. Con i risultati ottenuti proviamo a vedere il grafico della nostra funzione di densità sovrapposta all'istogramma dei dati Normali standardizzati.

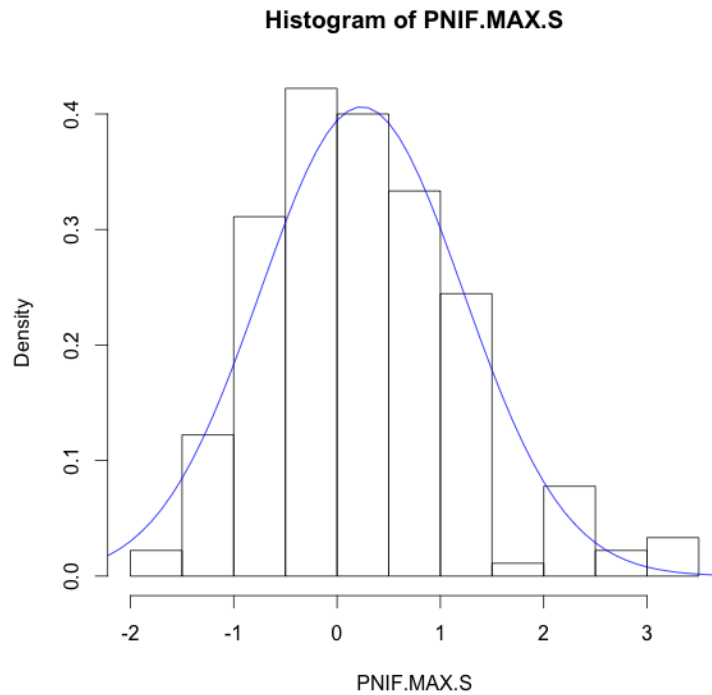


Fig. 4.5: istogramma con sovrapposta la funzione di densità di GSN

Abbiamo quindi sviluppato i due metodi differenti ed in entrambi i casi assumiamo che le singole variabili sono normali e in entrambi i casi quindi otteniamo il risultato che il massimo tra tre normali è una GSN, e infatti le due distribuzioni dei massimi ottenute seguendo i due metodi proposti sono praticamente identiche. Questo viene confermato dal seguente grafico.

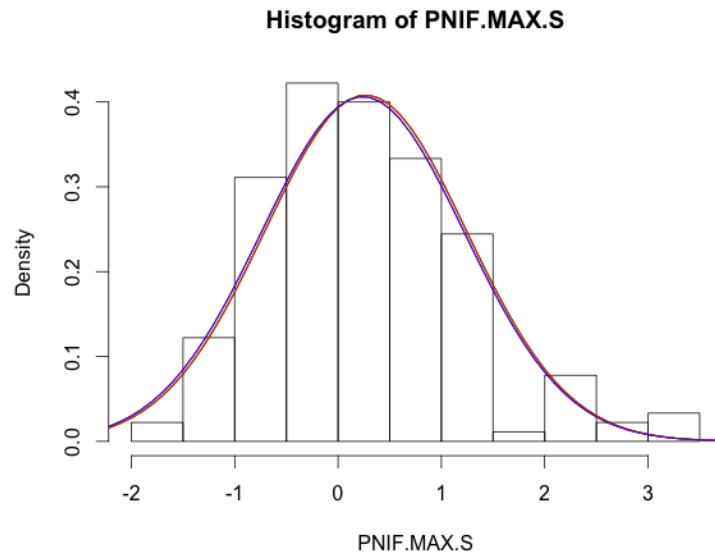


Fig. 4.6: istogramma, stima della densità secondo il primo metodo, linea rossa, e densità secondo il secondo metodo linea blu

Capitolo 5

Conclusioni

In questo lavoro di tesi abbiamo presentato la distribuzione del flusso nasale (PNIF) massimo tra le due narici e il flusso nasale massimo di tre osservazioni consecutive usando entrambe le narici, informazioni utili medicina nella diagnostica e nel trattamento di patologie rinologiche. In particolare, abbiamo studiato la misurazione del funzionamento delle vie nasali applicando il PNIF solo su una singola narice, tenendo ermeticamente chiusa l'altra. Per ottenere il massimo tra le due variabili unilaterali si procede seguendo 2 diversi metodi. Nel primo metodo riportiamo la modellazione dei massimi per i dati unilaterali del setto nasale, sapendo che il massimo tra le 2 variabili unilaterali che assumiamo normali (destra e sinistra), si distribuisce come una normale asimmetrica con parametri che dipendono dal coefficiente di correlazione tra le due normali unilaterali. A partire dai parametri stimati della distribuzione delle singole osservazioni si ottiene così la distribuzione del massimo.

Il secondo metodo si concentra nell'ottenere la distribuzione del massimo tra flusso destro e sinistro ottenendo direttamente il massimo delle osservazioni, ipotizzando che si distribuisce come normale asimmetrica e stimandone i parametri via massima verosimiglianza. Tramite l'analisi preliminare dei dati, abbiamo tuttavia osservato che i nostri dati non sembrano in realtà distribuirsi come una gaussiana perché presentano una leggera asimmetria. Tuttavia l'applicazione dei due diversi metodi per le stime ci ha portato a ottenere una misura di quanto l'ipotesi sulla distribuzione delle singole osservazioni influenza la forma della distribuzione del massimo tra le due variabili unilaterali (destra e sinistra). Ci auguriamo che in future ricerche si possa sviluppare anche la distribuzione del massimo tra due distribuzioni che presentano leggere deviazioni dalla normalità, come ad esempio le distribuzioni normali asimmetriche. Un'analisi dei massimi tra due skew normal potrebbe essere certamente interessante assieme allo studio delle sue proprietà formali e verificandone il comportamento su altri insiemi di dati. Nel capitolo 4 di questa tesi abbiamo anche effettuato uno studio sulla distribuzione del mas-

simo tra tre diverse quantità la cui distribuzione si ipotizza essere normale. Per ottenere il PNIF infatti è necessario ripetere l'operazione di inspirazione per tre volte consecutive nelle medesime condizioni e per ciascuna di queste si registra il risultato. Il PNIF è quindi il massimo tra queste tre misurazioni. Era perciò di interesse individuare la distribuzione del PNIF, a partire dalle singole distribuzioni delle inspirazioni.

Un risultato teorico di Jamalizadeh e Balakrishnan (2009) mostra che il massimo di tre variabili aleatorie normali correlate tra loro si distribuisce come una GSN (Generalized Skew Normal) con parametri legati ai coefficienti di correlazione tra le variabili originali. Anche in questo caso, per stimare la GSN del massimo, sono stati usati due approcci differenti:

Nel primo metodo vengono stimati i parametri dai dati raccolti dalle singole 3 variabili e poi otteniamo i parametri della GSN (Generalized Skew Normal) usando le relazioni tra i parametri delle 3 variabili e quelli della distribuzione del massimo dati da Jamalizadeh e Balakrishnan (2009). Nel secondo caso invece vengono stimati direttamente i parametri della distribuzione del massimo (che è una GSN) partendo dai massimi osservati.

Analogamente a quanto svolto sopra, anche qui è stata effettuata l'analisi preliminare dei dati e come risultato otteniamo che, in entrambi i casi i dati sono prossimi ad una distribuzione normale. Abbiamo quindi sviluppato i due metodi differenti ed in entrambi i casi assumiamo che le singole variabili sono normali e in entrambi i casi quindi otteniamo il risultato che il massimo tra tre normali è una GSN, e infatti le due distribuzioni dei massimi ottenute seguendo i due metodi proposti sono praticamente identiche.

Bibliografia

- [1] Azzalini, A. (2014). The Skew-Normal and Related Families. Institute of Mathematical Statistics Monographs 12-2013.
- [2] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 171-178.
- [3] Azzalini, A. (2000). Una presentazione basata sul concetto di verosimiglianza, 382pp, 2a edizione, Italia, Springer-Vergal.
- [4] Bowman, A. W., Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations*. Oxford University Press.
- [5] Corey, J. P., Houser, S. M., Ng, B. A. (2000). Nasal congestion: a review of its etiology, evaluation, and treatment. *Ear nose and throat journal*, 79(9), 690-703.
- [6] Holmstrom, M., Scadding, G. K., Lund, V. J., Darby, Y. C. (1990). Assessment of nasal obstruction. A comparison between rhinomanometry and nasal inspiratory peak flow. *Rhinology*, 28(3), 191-196.
- [7] Iacus, S. M., Masarotto, G. (2005). *LabstatR: libreria del laboratorio di statistica con R*. McGraw-Hill education.
- [8] Haddad, J. H. and Provost, S. B. (2011). Approximations to the distribution of the sample correlation coefficient. *World Academy of Science, Engineering and Technology* 52 910-915.
- [9] Jamalizadeh, A., Balakrishnan, N. (2009). Order statistics from trivariate normal and tv-distributions in terms of generalized skew-normal and skew-tv distributions. *Journal of Statistical Planning and Inference*, 139(11), 3799-3819.
- [10] Loperfido, N. (2008). A note on skew-elliptical distributions and linear functions of order statistics. *Statistics Probability Letters*, 78(18), 3184-3186.

- [11] Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statistics probability letters*, 56(1), 13-22.
- [12] Ottaviano, G., Scadding, G. K., Scarpa, B., Accordi, D., Staffieri, A., Lund, V. J. (2012). Unilateral peak nasal inspiratory flow, normal values in adult population. *Rhinology*, 50(4), 386-392.
- [13] Ottaviano, G., Lund, V. J., Nardello, E., Scarpa, B., Mylonakis, I., Frasson, G., Staffieri, A. (2014). Peak nasal inspiratory flow: a useful and handy tool for the diagnosis of nasal obstruction in the elderly. *European Archives of Oto-Rhino-Laryngology*, 1-5.
- [14] Youlten, L. J. F. (1980). The peak nasal inspiratory flow meter: a new instrument for the assessment of the response to immunotherapy in seasonal allergic rhinitis. *Allergol immunopathol*, 8, 344.
- [15] <http://coeh.berkeley.edu/ucsfoem/upperair/Rhinomanometry.htm>

Appendice A

Codice sorgente e immagini

PNIF DESTRA SECONDA MISURAZIONE

```
PNIFDESTRA2<-DATLPNIF_ALL$PNIF.DX.2  
  
classi <- seq(min(PNIFDESTRA2), max(PNIFDESTRA2))  
  
hist(PNIFDESTRA2, breaks=classi, prob=TRUE)  
  
lines(density(PNIFDESTRA2))  
  
boxplot(DATLPNIF_ALL$PNIF.DX.2)  
  
qqnorm(DATLPNIF_ALL$PNIF.DX.2)  
  
qqline(DATLPNIF_ALL$PNIF.DX.2, col=2)
```

PNIF SINISTRA SECONDA MISURAZIONE

```
PNIFSINISTRA2<-DATLPNIF_ALL$PNIF.SX.2  
  
classi <- seq(min(PNISINISTRA2), max(PNISINISTRA2))  
hist(PNIFSINISTRA2, breaks=classi, prob=TRUE, ylim=c(0,0.018), xlim=c(30,210))  
  
lines(density(PNIFSINISTRA2))  
  
boxplot(DATLPNIF_ALL$PNIF.SX.2, ylim=c(20,190),)  
  
qqnorm(DATLPNIF_ALL$PNIF.SX.2)  
  
qqline(DATLPNIF_ALL$PNIF.SX.2, col=2)}
```

STIMA USANDO IL PRIMO METODO DI MAX((PNIF.DX.2,PNIF.SX.2)

```
Z1=Z  
  
n<-180  
  
for(i in 1:n)  
{  
  Z1[i]<-max(DATLPNIF_ALL$PNIF.DX.2[i] : DATLPNIF_ALL$PNIF.SX.2[i], na.rm = FALSE)  
}  
  
DATLPNIF_ALL<-cbind(DATLPNIF_ALL,Z1)
```

```

quantile(Z1)

pdf1<-dsn(seq(30,260,by=0.1),xi=95.40275,omega=37.60164,alpha=0.4095726,dp=NULL,log=FALSE,)

PNIFZ1<-DATLPNIF_ALL$Z1

hist(PNIFZ1,breaks=16, prob=TRUE)

s<-seq(30, 260, by=0.1)

lines(s,pdf1,lwd=2,lty=3,col=2)

```

SECONDO METODO: STIMA CON L'ALGORITMO DI OTTIMIZZAZIONE

```

m1 <- selm("Z1 ~ 1", family="SN", data=DATLPNIF_ALL)

print(m1)

plot(m1, param.type="DP")

```

INTERVALLI DI CONFIDENZA

```

max(PNIF.SX.2,PNIF.DX.2)

uso<-DATLPNIF_ALL[,c(12,17)]

x1 <- (uso[,1]-mean(uso[,1]))/sd(uso[,1])
x2 <- (uso[,2]-mean(uso[,2]))/sd(uso[,2])
d1<- sum((x1+x2)^2)
d2<- sum((x1-x2)^2)
sqrt(d2/d1*qf(0.975,180-1,180-1))
sqrt(d2/d1*qf(0.025,180-1,180-1))

```

VERIFICA NORMALITÀ, HISTOGRAMA

```

PNIF2<-DATLPNIF_ALL$PNIF.2

hist(PNIF2, breaks=16, prob=TRUE)

lines(density(PNIF2))

boxplot(DATLPNIF_ALL$PNIF.2)

qqnorm(DATLPNIF_ALL$PNIF.2)

qqline(DATLPNIF_ALL$PNIF.2,col=2)

Z1=PNIF.MAX

attach(DATLPNIF_ALL)

p.d<-(PNIF.DX.2 - mean(PNIF.DX.2))/sd(PNIF.DX.2)

p.s<-(PNIF.SX.2 - mean(PNIF.SX.2))/sd(PNIF.SX.2)

n<-180

for(i in 1:n)
{

```



```

    }
    Z1[i]<-max(PNIF.DX.2[i] : PNIF.SX.2[i], na.rm = FALSE)
  }
  DATLPNIF_ALL<-cbind(DATLPNIF_ALL,Z1)

```

STIMA DEI PARAMETRI CON I DUE METODI PRESENTATI IN CAP. 4

```

DATLPNIF_ALL <- read.csv("DATLPNIF_ALL.csv")
attach(DATLPNIF_ALL)

```

FUNZIONE DI PNIF.MAX STANDARDIZZANDO

```

PNIF.1_s<-(PNIF.1 - mean(PNIF.1))/sd(PNIF.1)
PNIF.2_s<-(PNIF.2 - mean(PNIF.2))/sd(PNIF.2)
PNIF.3_s<-(PNIF.3 - mean(PNIF.3))/sd(PNIF.3)

#Creo la mia colonna del massimo con dati standard
PNIF.MAX.S=NULL

n<-180

for(i in 1:n)
{
  PNIF.MAX.S[i]<-max(PNIF.1_s[i],PNIF.2_s[i],PNIF.3_s[i], na.rm = FALSE)
}

DATLPNIF_ALL<-cbind(DATLPNIF_ALL,PNIF.MAX.S)

summary(PNIF.MAX.S)

library(mnormt)

```

PRIMO METODO

```

rho12=cor(PNIF.1,PNIF.2)
rho13=cor(PNIF.1,PNIF.3)
rho23=cor(PNIF.2,PNIF.3)

rho.s=(rho12+rho13+rho23)/3

# rho è la media di rho12, rho13, rho23
lambda1=lambda2=sqrt((1-rho.s)/(1+rho.s))

rho=rho.s/(1+rho.s)

a_theta3<- 1/(2*pi) * (acos((-rho-lambda1*lambda2)/sqrt((1+lambda1^2)*(1+lambda2^2))))

cost3<-1/a_theta3

PNIF.MASSIMO=PNIF.MAX.S

Sigma<-matrix(c(1,rho,rho,1),ncol=2,nrow=2)

x<-(-30:40)/10

dgsn.1.plot<-NULL

for (i in 1:NROW(x))

```

```

dgsn.1.plot[i]=(1/a_theta3)*dnorm(x[i])*(pmnorm(c(lambda1*x[i],lambda2*x[i]),varcov=Sigma))

scriptsize
hist(PNIF.MASSIMO, freq=F)

lines(x,dgsn.1.plot)

```

SECONDO METODO

```

dpdf.b=function(parametro)
{
  rho<-exp(parametro)/(1+exp(parametro))*2-1

  lambda1=sqrt((1-rho)/(1+rho))

  lambda2=lambda1

  rho.new<- rho/(1+rho)

  Sigma<-matrix(c(1,rho.new,rho.new,1),ncol=2,nrow=2)

  a_theta<-1/(2*pi)*(acos((-rho.new-lambda1*lambda2)/sqrt((1+lambda1^2)*(1+lambda2^2))))

  x = PNIF.MAX.S

  dgsn<-NULL

  for (i in 1:NROW(x))
    dgsn[i]=(1/a_theta)*dnorm(x[i])*(pmnorm(c(lambda1*x[i],lambda2*x[i]),varcov=Sigma))

  return(dgsn)
}

nllik=function(parametro)
{
  -sum(log(dpdf.b(parametro)))
}

out<-nlminb(c(0.47), nllik)

#parametri della distribuzione del massimo
#riscolo il rho della singola X per riportarlo tra -1 e 1
ro<-exp(out$par)/(1+exp(out$par))*2-1

#ottengo il rho per la distribuzione GSN dei massimi
ro.new<-ro/(1+ro)

#ottengo i lambda1=lambda2 per la distribuzione GSN dei massimi
lam<- sqrt((1-ro)/(1+ro))

#creo una funzione per disegnare la GSN con una qualsiasi x
dpdf.bx=function(rho, x)
{
  rho<- exp(rho)/(1+exp(rho))*2-1

  lambda1=sqrt((1-rho)/(1+rho))

  lambda2=lambda1

  rho.new<- rho/(1+rho)

  Sigma<-matrix(c(1,rho.new,rho.new,1),ncol=2,nrow=2)

```

```

a_theta<-1/(2*pi) * (acos((-rho.new-lambda1*lambda2)/sqrt((1+lambda1^2)*(1+lambda2^2)))

dgsn<-NULL

for (i in 1:NROW(x))
{
  dgsn[i]=(1/a_theta)*dnorm(x[i])*(pmnorm(c(lambda1*x[i],lambda2*x[i]),varcov=Sigma))

}

return(dgsn)
}

#calcolo la funzione da disegnare per alcuni valori della
uso<-dpdf.bx(out$par, x=(-30:40)/10)

hist(PNIF.MAX.S, freq=F)

lines((-30:40)/10,uso, col=4)

plot((-30:40)/10,uso, type="l")

hist(PNIF.MAX.S, freq=F, add=T)

```

GRAFICI VERIFICA DELLA NORMALITÀ MISURAZIONE DEL PRIMO, TERZO DEL PNIF DI DESTRA E SINISTRA

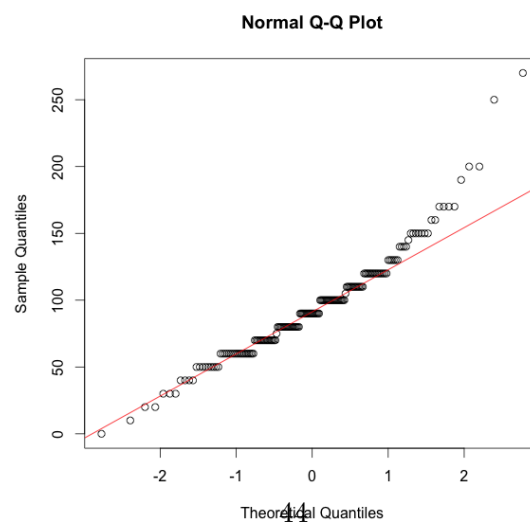
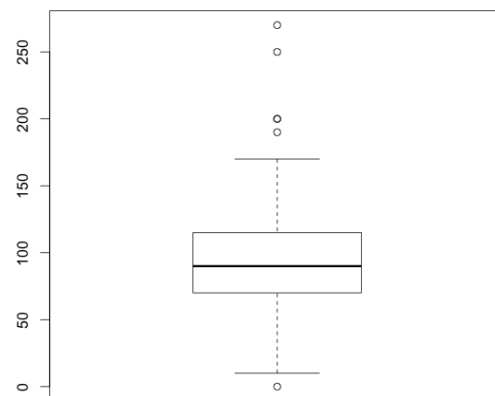
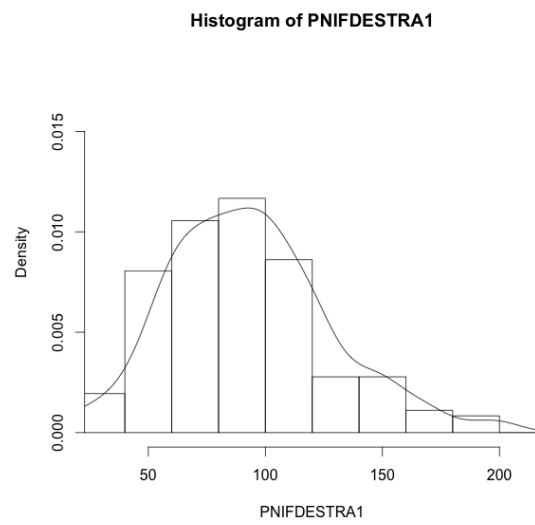


Fig. A.1: a. Istogramma del PNIF.DX.1; b. Diagramma a scatola con baffi; c. Diagramma Q-Q normal.

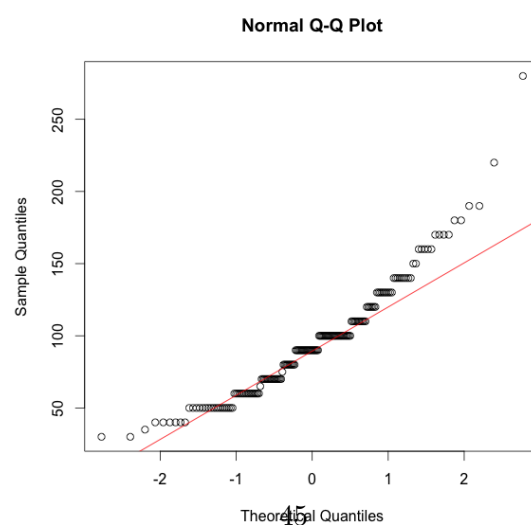
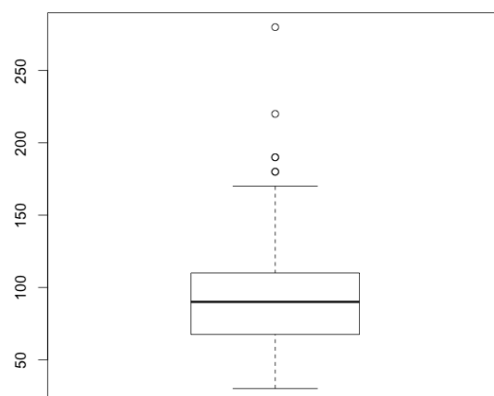
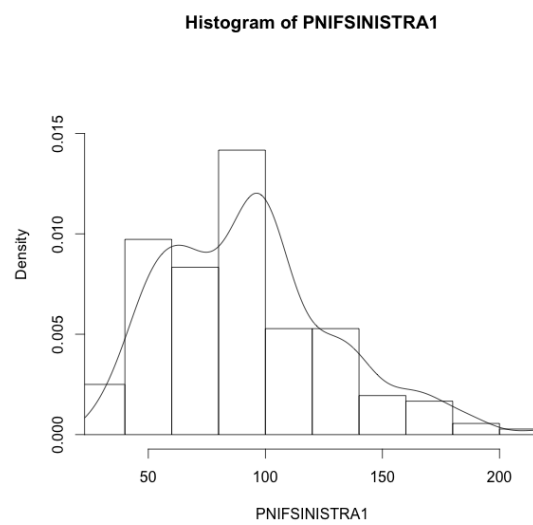


Fig. A.2: **a.** Istogramma del PNIF.SX.1; **b.** Diagramma a scatola con baffi; **c.** Diagramma Q-Q normal.

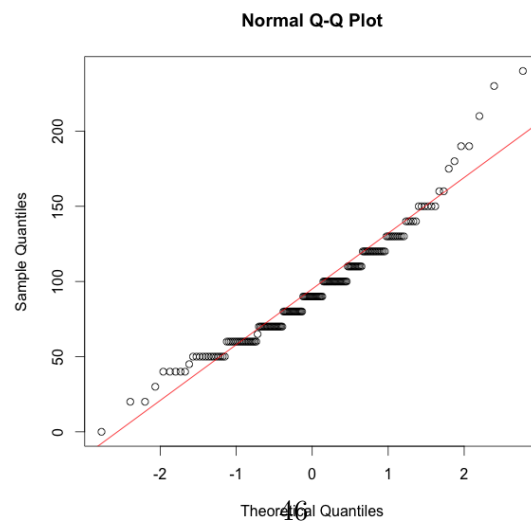
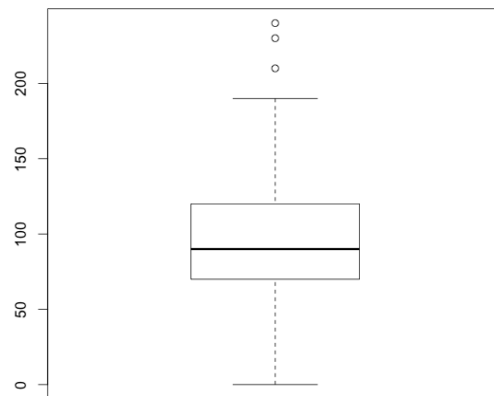
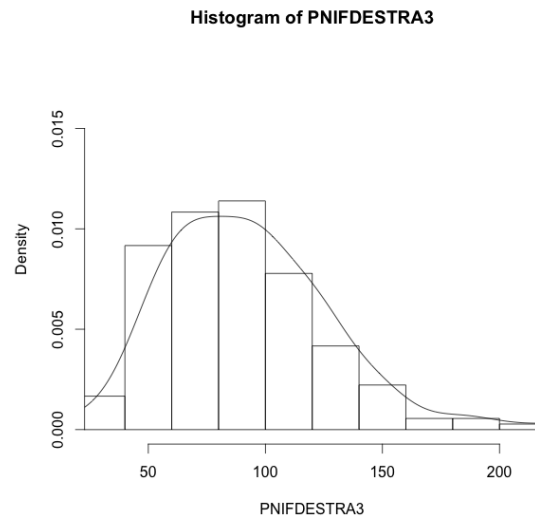


Fig. A.3: a. Istogramma del PNIF.DX.3; b. Diagramma a scatola con baffi; c. Diagramma Q-Q normal.

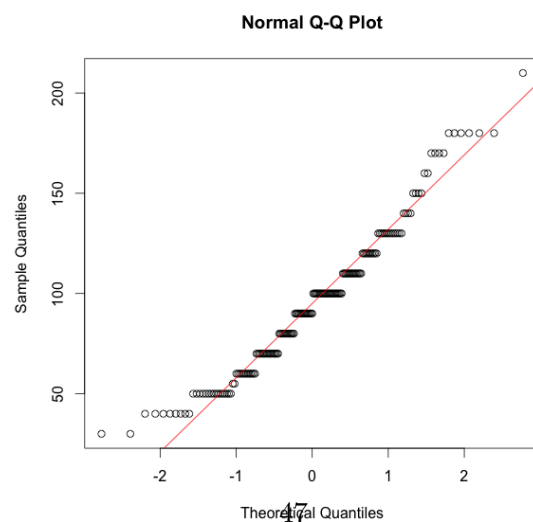
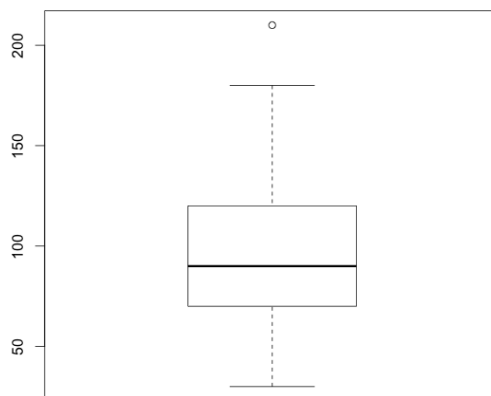
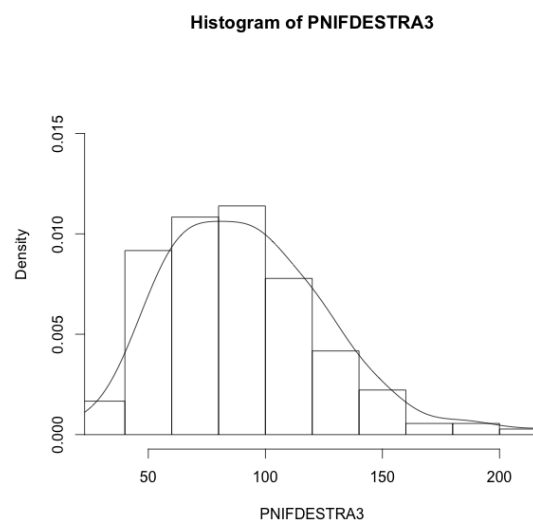


Fig. A.4: **a.** istogramma del PNIF.SX.3; **b.** Diagramma a scatola con baffi; **c.** diagramma Q-Q normal.

CAP3.GRAFICI DELLA FUNZIONE DI DENSITÀ USANDO PRIMO E SECONDO METODO



Fig. A.5: a. istogramma del massimo tra le due osservazioni unilaterali primo metodo; b. istogramma del massimo tra le due osservazioni unilaterali secondo metodo;

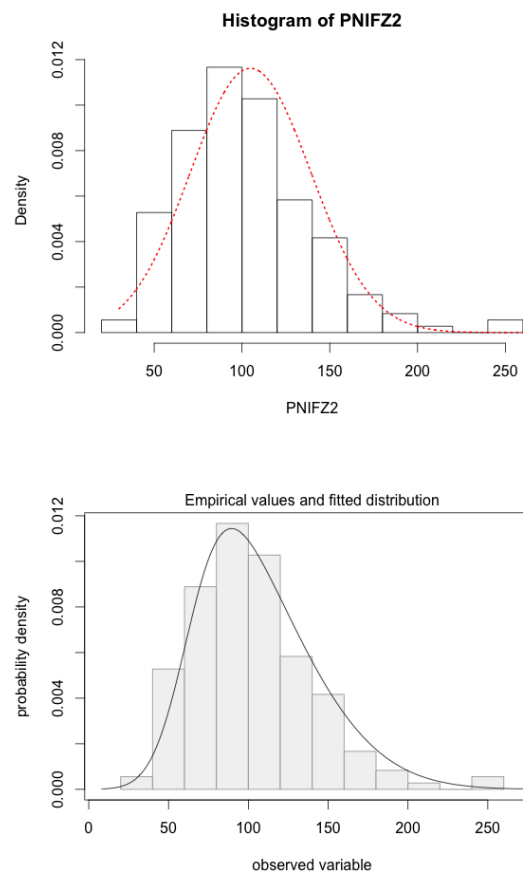


Fig. A.6: **a.** istogramma del massimo tra le due osservazioni unilaterali;
b. diagramma a scatola con baffi.

CAP4.GRAFICI DELLA FUNZIONE DI DENSITÀ USANDO PRIMO E SECONDO METODO

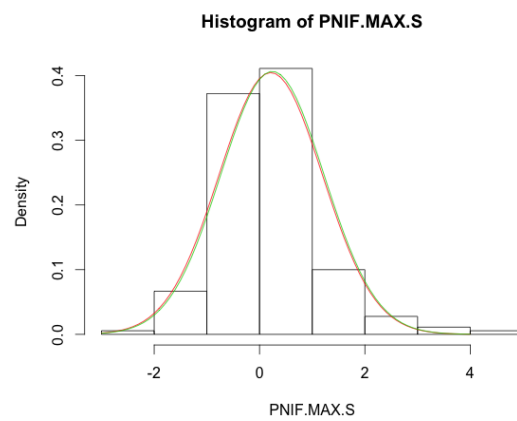
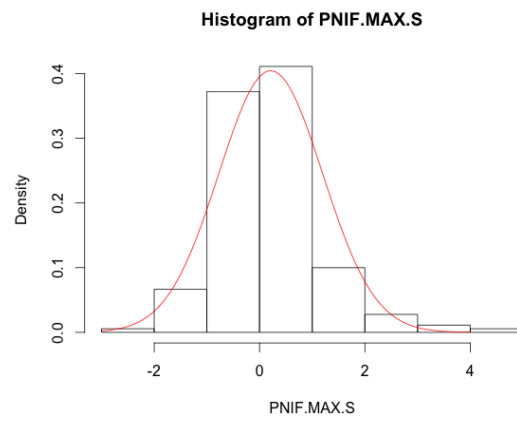
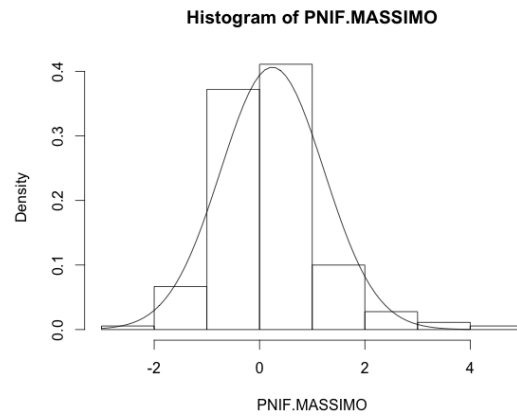


Fig. A.7: **a.** istogramma con la densità del massimo tra le tre osservazioni primo metodo; **b.** istogramma con la densità del massimo tra le tre osservazioni usecondo metodo; **c.** confronto densità.

Appendice B

Allegati



Gavin Setzen, MD

Lawrence S. Kaufman, MD

Diplomates

American Board of Otolaryngology

Fellows

American College of Surgeons

American Academy of Otolaryngic Allergy

Siobhan Kuhar, MD, PhD

Diplomate

American Board of Otolaryngology

American Board of Sleep Medicine

Fellow

American College of Surgeons

American Academy of Otolaryngic Allergy

Nora Perkins, MD, FACS

Diplomate

American Board of Otolaryngology

Fellow American College of Surgeons

John Gavin, MD, FAAP

Fellow American Academy of Pediatrics

Diplomate of American Board of Otolaryngology

Robert Adelson, MD, FACS

Fellow American College of Surgeons

Diplomate American Board of Otolaryngology –

Head and Neck Surgery

Diplomate American Board of Facial Plastic and

Reconstructive Surgery

Michael Dailey, MD

Donna Silvernail, RPA-C

Maggie West-Bump, RPA-C

Robert Nadratowski, RPA-C

Robyn Smith, RPA-C

400 Patroon Creek Boulevard, Suite 205

Albany, New York 12206

Tel : 518 701.2000

Fax: 518 701.2020

www.albanyentandallergy.com

**Pediatric & Adult
Ear, Nose & Throat Disorders**

Sinus, Nasal & Rhinology

Allergy Testing & Treatment

Dizziness & Hearing Loss

Voice, Speech & Swallowing

Snoring & Sleep Apnea

Head & Neck Surgery

Thyroid & Parathyroid Surgery

Facial Plastic Surgery

CAT Scan Imaging

Hearing & Balance Testing

Audiology

Marcia Perretta, AuD

Deanna Ross, AuD

Tricia Brown, AuD

Dana Wilhite, AuD

CAT Scan Imaging

Judith Martin, L.R.T., C.T.M.

Speech Pathology

Susan C. Hare, MS, CCC-SLP

Sino-Nasal Outcome Test (SNOT-22)

Name: _____

Date: _____

Below you will find a list of symptoms and social/emotional consequences of your nasal disorder. We would like to know more about these problems and would appreciate your answering the following questions to the best of your ability. There is no right or wrong answers, and only you can provide us with this information. Please rate your problems as they have been over the past two weeks. Thank you for your participation.

Consider how severe the problem is when you experience it and how frequently it happens, please rate each item below on how "bad" it is by circling the number that corresponds with how you feel using the table. →	No Problem	Very Mild Problem	Mild or Slight Problem	Moderate Problem	Severe Problem	Problem is bad as it can be	Most important 5 items
1. Need to blow nose	0	1	2	3	4	5	<input type="checkbox"/>
2. Sneezing	0	1	2	3	4	5	<input type="checkbox"/>
3. Runny nose	0	1	2	3	4	5	<input type="checkbox"/>
4. Nasal obstruction	0	1	2	3	4	5	<input type="checkbox"/>
5. Loss of smell or taste	0	1	2	3	4	5	<input type="checkbox"/>
6. Cough	0	1	2	3	4	5	<input type="checkbox"/>
7. Post-nasal discharge	0	1	2	3	4	5	<input type="checkbox"/>
8. Thick nasal discharge	0	1	2	3	4	5	<input type="checkbox"/>
9. Ear fullness	0	1	2	3	4	5	<input type="checkbox"/>
10. Dizziness	0	1	2	3	4	5	<input type="checkbox"/>
11. Ear pain	0	1	2	3	4	5	<input type="checkbox"/>
12. Facial pain/pressure	0	1	2	3	4	5	<input type="checkbox"/>
13. Difficulty falling asleep	0	1	2	3	4	5	<input type="checkbox"/>
14. Wake up at night	0	1	2	3	4	5	<input type="checkbox"/>
15. Lack of a good night's sleep	0	1	2	3	4	5	<input type="checkbox"/>
16. Wake up tired	0	1	2	3	4	5	<input type="checkbox"/>
17. Fatigue	0	1	2	3	4	5	<input type="checkbox"/>
18. Reduced productivity	0	1	2	3	4	5	<input type="checkbox"/>
19. Reduced concentration	0	1	2	3	4	5	<input type="checkbox"/>
20. Frustrated/restless/irritable	0	1	2	3	4	5	<input type="checkbox"/>
21. Sad	0	1	2	3	4	5	<input type="checkbox"/>
22. Embarrassed	0	1	2	3	4	5	<input type="checkbox"/>

1. Please check off the 5 most important items affecting your health.....↑

2. List any other items important to you, which you feel are not mentioned above.

Systems reviewed with above questions: General Ear, Nose & Throat · Respiratory · Neurology · Psychology

Ringraziamenti

In quest'ultima pagina voglio ringraziare chi, a suo modo, mi ha aiutato in questi cinque anni di Università. In primo luogo, ringrazio il Prof. Bruno Scarpa per l'aiuto durante questo lavoro di tesi, per le correzioni della lingua italiana e soprattutto per l'esame di data-mining. Un grazie in particolare per le sue buone parole su di me. Ringrazio inoltre il Prof. Guido Masarotto per la disponibilità e per i consigli dati.

Un grazie a tutti i miei amici che mi sono stati vicini. Grazie per il vostro supporto Andrea e Maddalena. Un super grazie a Oriona e Marsel, Ela e Andi che mi sono sempre stati vicini e soprattutto sempre ottimisti. A Tushi (Matilda) per tutti i pranzi e le cene, per il tuo essere in ritardo e le chiacchierate infinite fatte. Un ringraziamento va alla mia personal language trainer (Monika) che mi è sempre stata vicina anche via Skype e a tutti gli amici di Ceccarelli per le serate di studio.

Un particolare ringraziamento va a mia lali (Jenny) per avermi sempre regalato i sorrisi usando la sua terapia telefonica e per le bellissime serate a Jesolo.

Grazie a Henk per tutto (anche se non sei molto bravo ad aiutarmi moralmente) ma nessuno è perfetto.

Ringrazio la mia famiglia per avermi portata fin qui e soprattutto alla mia sorella Fabiola (per la sicurezza che ha su di me e per tutto il resto). Mamma e papà per la preoccupazione dimostrata per gli ultimi sviluppi della tesi e per i consigli durante questi cinque anni. A mia sorella Elona che mi accontentava sempre con i regali dopo ogni sessione d'esame. Un grazie e un saluto va al mio GRUPI (Turi) e a mia sorella Alda per i due fantastici nipotini che basta un loro sorriso su Skype e la zia è sempre felice.

Finito di stampare il 27 giugno 2014 utilizzando L^AT_EX.