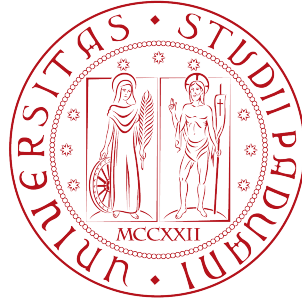


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



Relazione finale

Indici di capacità per processi gaussiani e non-gaussiani

Relatore: Prof. Guido Masarotto
Dipartimento di Scienze Statistiche

Laureando: Matteo Garato
Matricola N. 2003660

Anno Accademico 2022/2023

Indice

| | |
|---|-----------|
| Introduzione | 1 |
| 1 Capacità statistica | 3 |
| 2 Indici di capacità | 6 |
| 2.1 C_p e C_{pk} | 7 |
| 2.1.1 Per distribuzioni normali | 7 |
| 2.1.2 Per distribuzioni generiche | 8 |
| 2.2 C_{pm} e C_{pmk} | 9 |
| 3 Un nuovo indice | 10 |
| 3.1 S_{pmk} | 10 |
| 3.2 S'_{pmk} | 11 |
| 3.2.1 La funzione di perdita | 11 |
| 3.2.2 L'indice | 12 |
| 3.2.3 Bootstrap | 13 |
| 3.2.4 Il codice R | 14 |
| 4 Applicazione degli indici ai dati | 18 |
| 4.1 Dati gaussiani a media variabile | 18 |
| 4.2 Dati Poisson a media variabile | 23 |
| 4.3 Dati reali | 27 |
| 4.3.1 Descrizione dei dati | 27 |
| 4.3.2 Analisi della capacità del processo | 27 |

Introduzione

Nell'ambito della produzione industriale, è fondamentale saper generare prodotti che siano conformi alle specifiche richieste dal cliente. La misura di questa caratteristica prende in nome di *capacità* (in inglese *capability*). Una buona capacità è un requisito molto importante per un'azienda, in quanto un cliente è portato ad acquistare da un fornitore che gli garantisca un alto rapporto di prodotti conformi - non conformi.

Questa può essere misurata tramite un elevato numero di indici, chiamati appunto indici di capacità, ognuno dei quali si differenzia per qualche aspetto: ci sono indici adatti a distribuzioni gaussiane e altri a distribuzioni più generiche, alcuni tengono conto della posizione della media all'interno dei limiti di specifica, altri pesano la differenza dal valore target in modo lineare, quadratico, cubico ecc. L'importante è trovare l'indice più adatto alle proprie esigenze, dal momento che, come vedremo, indici diversi possono dare risultati differenti anche se applicati agli stessi dati.

In questa tesi ne studieremo alcuni e ne verranno analizzate le caratteristiche. In particolare vedremo due gruppi di indici famosi (C_p e C_{pk} , C_{pm} e C_{pmk}) e uno che è stato proposto da Erfanian and Sadeghpour Gildeh (2021), il quale, a differenza degli altri, utilizzando un funzione di perdita asimmetrica, è in grado di pesare in modo differente scarti positivi e negativi dal valore target.

Tutti questi indici verranno calcolati inizialmente su dati simulati derivanti da distribuzioni normali e Poisson, potendone così valutare l'applicazione su dati continui e discreti, simmetrici e asimmetrici. Infine verrà visto un

caso con dei dati reali, in cui il campione deriva da un processo che produce dei tubicini in gomma per una apparecchiatura elettromedicale, dei quali ci soffermeremo nel diametro e ne analizzeremo la capacità.

Capitolo 1

Capacità statistica

In ambito statistico, la capacità misura quanto un processo in controllo è in grado di produrre risultati che siano all'interno dei limiti specifica, riuscendo così a soddisfare le richieste del cliente. Per capire meglio questo concetto possiamo aiutarci con un esempio. Supponiamo che il processo da analizzare sia una linea di produzione di viti la cui lunghezza deve essere compresa tra $50\text{mm} \pm 0.5\text{mm}$. Il processo è tanto più capace quanto è in grado di fornire viti lunghe tra $49,5\text{mm}$ e $50,5\text{mm}$. Estruendo un campione di 500 unità, si potrebbe notare che la lunghezza segue una distribuzione normale di media μ e varianza σ^2 . Alcuni esempi di quello che potrebbe accadere lo possiamo vedere in figura 1.1. Supponiamo che la lunghezza delle viti segua la distribuzione indicata, notiamo che le quattro curve sono per la maggior parte all'interno dei limiti tratteggiate, nonostante ciò la capacità dei processi non è sempre la stessa. La curva a è la più capace: la media è centrata all'interno dei limiti di specifica, infatti coincide con $\frac{USL+LSL}{2}$, e la varianza è bassa. la curva b , nonostante non sia centrata, è ancora capace in quanto la varianza ha un valore basso per cui è completamente all'interno delle linee tratteggiate. Gli ultimi due sono invece processi poco capaci, in particolare la c , anche se centrata, presenta una variabilità troppo elevata, tanto che le due code finiscono per cadere abbondantemente al di fuori dei limiti; la d invece ha la media troppo vicina al limite superiore, motivo per

cui gran parte delle viti finirà per avere lunghezza maggiore di 50.5 mm.

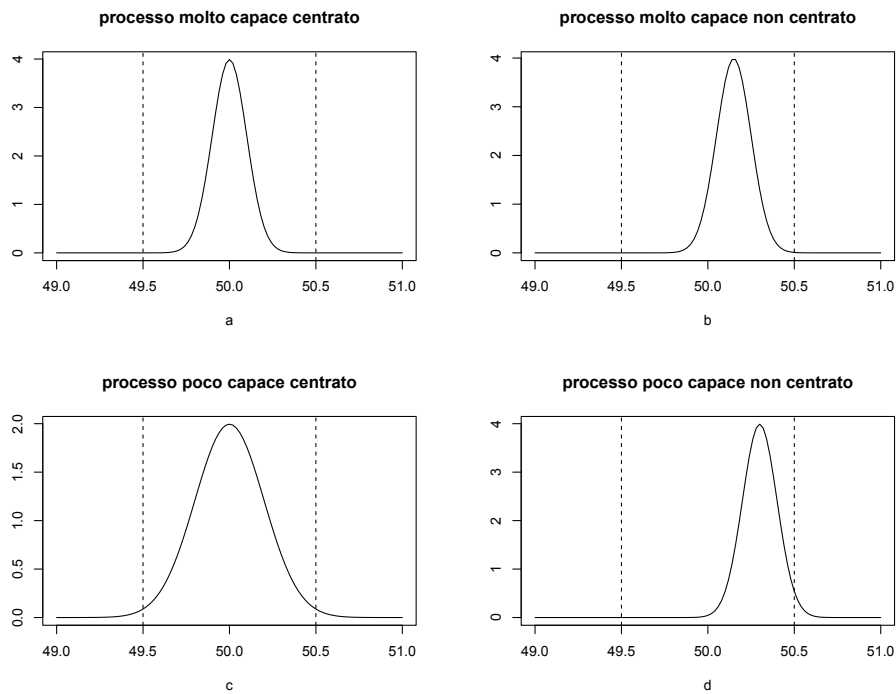


Figura 1.1: Processi molto capaci e poco capaci

Esempi analoghi possono essere fatti anche con casi non gaussiani. Proviamo a vedere come si applica la capacità alla distribuzioni Gamma, che a differenza della normale, è asimmetrica. Supponiamo di avere come limiti di specifica i valori 1 e 9 e un processo con distribuzione Gamma come in figura 1.2. Notiamo che la curva a, la quale ha la media centrata tra i limiti di specifica ha la coda destra un po' al di sopra del limite superiore. Nel caso in cui la media del processo fosse controllabile, conviene spostarla leggermente a sinistra, come nel caso b. In questo modo risulta essere più capace in quanto entrambe le code stanno all'interno dei limiti.

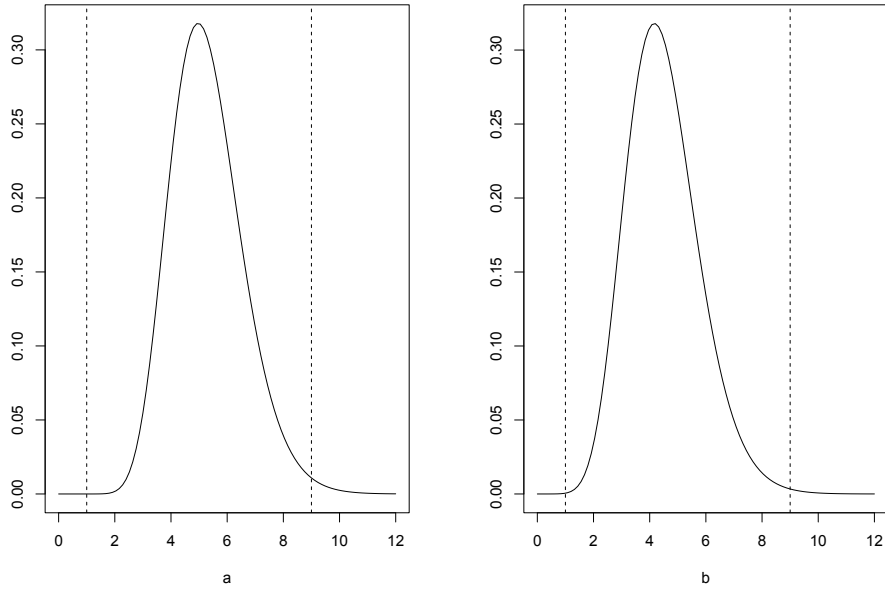


Figura 1.2: Capacità di due processi gamma

Talvolta però la famiglia da cui deriva il processo non è nota e non è quindi possibile tracciare con precisione la curva della distribuzione, che dovrà quindi essere stimata. In questo caso bisogna però tener presente che l'approssimazione sarà buona solo quando la numerosità campionaria è molto elevata, per questo motivo necessitiamo di un campione numeroso.

Capitolo 2

Indici di capacità

Per analizzare la capacità di un processo e non è sufficiente guardare come la distribuzione, o l'istogramma del campione, si posiziona all'interno dei limiti di specifica; ci serve un modo per poterla quantificare in modo da poter anche confrontare processi differenti. A questo scopo ci vengono in aiuto gli indici di capacità (in inglese *capability index*). Generalmente questi sono costruiti in modo che un processo sia definito capace quando l'indice assume un valore maggiore di 1. Questa però è solo una convenzione, in quanto dipende da quali sono i requisiti dell'azienda che, se vuole ottimizzare la propria produzione, vorrà avere valori più elevati. Un esempio sono i processi "Sei Sigma", nei quali si punta ad avere l'indice C_p pari a 2.

L'idea di processi Sei Sigma nasce negli anni '80 da parte dello statistico Mikel J. Harry che lavorava per Motorola. Lo scopo di Harry era quello di migliorare il suo reparto rendendo la produzione più rigorosa possibile. L'obiettivo principale di questi processi è ridurre al minimo la variabilità, così da massimizzare il livello di performance. Un processo gaussiano che rispetta i requisiti Sei Sigma ha infatti un rendimento pari a 99.9997%, che significa avere in media solamente 3,4 non conformità ogni milione di pezzi prodotti.

In questo capitolo vedremo quali sono i principali indici che vengono utilizzati per misurare la capacità dei processi. Per ulteriori approfondimenti

sugli indici di capacità rimando al libro Montgomery (2013, cap. 8).

2.1 C_p e C_{pk}

2.1.1 Per distribuzioni normali

I primi due indici che andremo a vedere sono utilizzati per le distribuzioni normali, questi sono l'indice C_p e la sua versione più flessibile C_{pk} . Il primo è definito come

$$C_p = \frac{USL - LSL}{6 \times \sigma}$$

dove USL e LSL sono i limiti di specifica, mentre σ è la deviazione standard del processo. Dal momento che il vero valore di σ è ignoto, lo andremo a stimare ad esempio mediante la deviazione standard campionaria S . L'indice diventa:

$$\hat{C}_p = \frac{USL - LSL}{6 \times S} \quad (2.1)$$

L'assunzione alla base dell'utilizzo di questo indice è che la media del processo coincida con $\frac{USL+LSL}{2}$, ossia che sia centrata all'interno dei limiti di specifica. Se questo è vero, l'indice coincide con il rapporto tra l'ampiezza dell'intervallo accettabile e la naturale variabilità del processo, dal momento che nella distribuzione normale l'intervallo $\mu \pm 3 \times \sigma$ contiene circa 99,97% di probabilità.

Non sempre è possibile avere la media centrata, in questi casi viene utilizzato l'indice C_{pk} che è definito come

$$C_{pk} = \min \left(\frac{\mu - LSL}{3 \times \sigma}; \frac{USL - \mu}{3 \times \sigma} \right)$$

dove μ rappresenta la vera media del processo, ma dal momento che questa è spesso ignota, viene stimata con la media campionaria \bar{X} . L'indice diventa quindi

$$\hat{C}_{pk} = \min \left(\frac{\bar{X} - LSL}{3 \times S}; \frac{USL - \bar{X}}{3 \times S} \right) \quad (2.2)$$

Come già affermato, si è soliti dire che un processo è capace quando l'indice è maggiore di 1. Può capitare di volere una migliore performance ed un tasso di non conformità minore, in questi casi gli indici assumeranno valori maggiori. Nella tabella 2.1 vediamo come varia C_p rispetto a determinate possibili richieste di tassi di non conformità.

| | | | | | | | | | |
|---------------------------------------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| Numero di non conformità ogni milione | 100000 | 10000 | 1000 | 100 | 10 | 1 | 0,1 | 0,01 | 0,0018 |
| C_p | 0,548 | 0,859 | 1,097 | 1,297 | 1,472 | 1,631 | 1,776 | 1,910 | 2,000 |

Tabella 2.1: Variazione di C_p al diminuire del numero di non conformità accettate

2.1.2 Per distribuzioni generiche

Con un piccolo aggiustamento questi indici si possono adattare anche a distribuzioni non normali. Quello che si deve fare è sostituire il denominatore con una quantità che coincida con quelli usati nelle formule precedenti.

Prendiamo il primo indice dove il denominatore è 6σ . Dal momento che nella normale i valori della funzione di ripartizione in corrispondenza di $\mu + 3\sigma$ e $\mu - 3\sigma$ sono 0,00135 e 0,99865, andremo a sostituire il 6σ con $(Q_{0,99865} - Q_{0,00135})$, dove Q_α rappresenta il quantile di ordine α del campione. La stima dell'indice diventa:

$$\hat{C}_p = \frac{USL - LSL}{\hat{Q}_{0,99865} - \hat{Q}_{0,00135}} \quad (2.3)$$

Sostituendo allo stesso modo i denominatori anche nel secondo indice, otteniamo lo stimatore per per dati non normale di C_{pk} :

$$\hat{C}_{pk} = \min \left(\frac{\hat{Q}_{0,5} - LSL}{\hat{Q}_{0,5} - \hat{Q}_{0,00135}}; \frac{USL - \hat{Q}_{0,5}}{\hat{Q}_{0,99865} - \hat{Q}_{0,5}} \right) \quad (2.4)$$

2.2 C_{pm} e C_{pmk}

I prossimi indici che vediamo sono C_{pm} (chiamato indice di Taguchi) e la sua versione per processi con media non centrata C_{pmk} . Questi a differenza da quelli appena visti che tenevano conto solo della distribuzione del processo all'interno dei limiti di specifica, pongono attenzione anche al valore target; ossia quel valore che si desidera ottenere dal processo, che non necessariamente coincide con il punto centrale dell'intervallo di specifica. L'indice di Taguchi è il seguente:

$$C_{pm} = \frac{USL - LSL}{6 \times \sqrt{\sigma^2 + (\mu - T)^2}}$$

dove T rappresenta il valore target. Nella versione in cui cade l'assunzione che la media sia centrata tra i limiti di specifica, l'indice diventa

$$C_{pmk} = \min \left(\frac{USL - \mu}{3 \times \sqrt{\sigma^2 + (\mu - T)^2}}; \frac{\mu - LSL}{3 \times \sqrt{\sigma^2 + (\mu - T)^2}} \right).$$

Ovviamente μ e σ devono essere stimati, e lo possiamo fare con \bar{X} e S.

La differenza tra questi due indici e quelli visti nel capitolo precedente sta nel denominatore, dove è presente la funzione di perdita $(\mu - T)^2$. Questa permette di tenere conto degli scarti del valore target della media del processo. In particolare la presenza dell'esponente quadrato fa sì che scarti grandi vengano pesati molto di più rispetto a quelli piccoli, che invece vanno a influenzare poco il valore dell'indice. Bisogna notare che, per come è costruita la funzione di perdita, scarti positivi e negativi vengono trattati allo stesso modo.

Capitolo 3

Un nuovo indice

In questo capitolo andiamo ad analizzare un nuovo indice che chiameremo S'_{pmk} , proposto da Erfanian and Sadeghpour Gildeh (2021). Prima di questo vedremo S_{pmk} , ossia l'indice a cui i due statistici si sono ispirati.

3.1 S_{pmk}

L'indice S_{pmk} è definito nel seguente modo:

$$S_{pmk} = \frac{\Phi^{-1}\left(\frac{1+F(USL)-F(LSL)}{2}\right)}{3\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}} = \frac{\Phi^{-1}\left(1-\frac{P}{2}\right)}{3\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}}$$

dove $\Phi^{-1}(\cdot)$ è l'inversa della funzione di probabilità della normale standard, mentre $F(\cdot)$ è la funzione di ripartizione del processo. Nella seconda versione P sta ad indicare la percentuale di non conformi.

La prima differenza che si nota rispetto all'indice C_{pm} è che al denominatore lo scarto tra il valore target T e la media μ viene standardizzata tenendo conto della varianza del processo; la funzione di perdita è infatti $\left(\frac{\mu-T}{\sigma}\right)^2$.

Un'altra grande differenza sta nel numeratore, dove viene usato $\Phi^{-1}(\cdot)$ per calcolare la distanza tra i limiti di specifica. In particolare $F(LSL)$ e $F(USL)$ verranno calcolati con la vera funzione di ripartizione qual'ora fosse nota la distribuzione dei dati, alternativamente si può usare la funzione di

ripartizione empirica. Questo modo di calcolare la distanza tra USL e LSL permette di tenere conto della distribuzione da cui derivano i dati, anche quando questa non è normale.

3.2 S'_{pmk}

Il nuovo indice che andremo a vedere è simile a questo, utilizza però una funzione di perdita diversa.

3.2.1 La funzione di perdita

La funzione di perdita che andiamo ad analizzare è stata introdotta da Varian e Zellner. A differenza di quella vista in precedenza, questa è asimmetrica ed è quindi in grado di pesare in modo differente scarti positivi e negativi.

La formula della funzione è la seguente:

$$L_3(x - T) = 2 \frac{e^{\gamma(x-T)} - \gamma(x - T) - 1}{\gamma^2}$$

dove γ è il parametro che regola il grado di asimmetria. Come si comporta la funzione al variare di γ lo possiamo vedere in figura 3.1.

In particolare vediamo come valori di γ positivi vadano a pesare di più gli scarti positivi rispetto a quelli negativi, mentre se è $\gamma < 0$ si andranno a pesare maggiormente valori di $(x - T)$ negativi. Inoltre all'aumentare di $|\gamma|$ aumenta anche il peso che lo scarto assumerà nel calcolo dell'indice. Viceversa se γ assume valori vicini allo zero, L_3 tende a comportarsi come la funzione di perdita quadratica, infatti:

$$\lim_{\gamma \rightarrow 0} L_3(x - T) = (x - T)^2$$

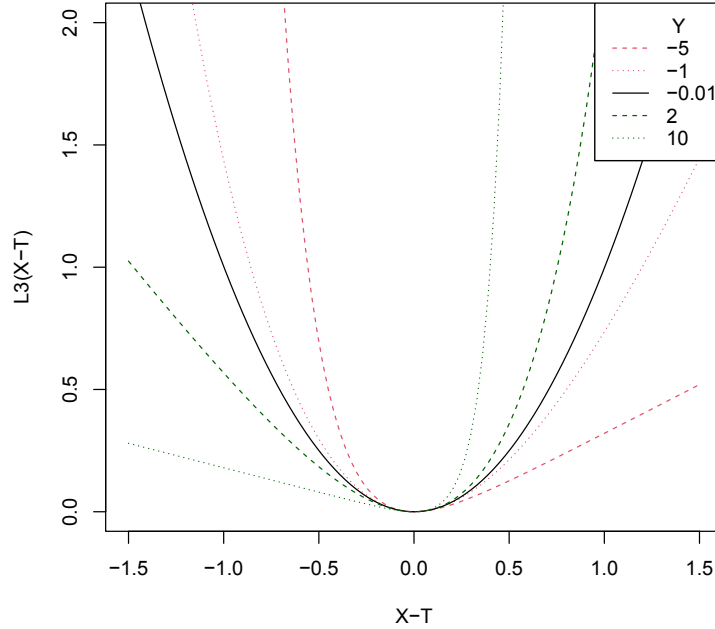


Figura 3.1: andamento di L_3 per diversi valori di γ

3.2.2 L'indice

L'indice proposto da Erfanian and Sadeghpour Gildeh prende spunto da S_{pmk} , al quale si applica però la funzione di perdita asimmetrica L_3 . L'indice diventa:

$$S'_{pmk} = \frac{\Phi^{-1} \left(\frac{1+F(USL)-F(LSL)}{2} \right)}{3\sqrt{1 + \frac{1}{\sigma^2} \times L_3(\mu - T)}}$$

Il vantaggio di utilizzare una funzione asimmetrica è che l'indice sarà in grado, in base al valore di γ , a considerare in maniera differente scarti positivi e negativi da T. Questo può essere utile ad esempio in un processo che produce compresse di un determinato farmaco, le quali devono avere peso μ_0 . In questo caso un peso inferiore di quello nominale potrebbe non essere abbastanza per curare la malattia, mentre un valore superiore potrebbe portare ad un'intossicazione del paziente; per questo motivo è importante tener

conto del segno dello scarto quando si calcola la capacità del processo. Un altro esempio in cui differenze positive o negative dal valore target hanno impatto diverso, si può presentare in una linea di produzione di laminati plastici in cui si vuole misurare la resistenza alla flessione trasversale. In questo caso una resistenza troppo elevata rende l'oggetto difficile da lavorare, viceversa nel caso opposto il laminato sarà troppo poco resistente e il rischio che si rompa sarà maggiore.

Indichiamo con Δ lo scarto $\mu - T$; il comportamento dell'indice al variare di γ è il seguente:

- quando sia Δ che γ sono positivi, S'_{pmk} diminuisce all'aumentare di γ . Questo avviene perchè mi trovo la direzione che ho dichiarato come quella critica, quindi l'indice indicherà una peggior performance del processo;
- quando Δ è negativo e γ è positivo, S'_{pmk} cresce all'aumentare di γ . Questo perchè mi trovo nella direzione opposta dello scarto che è dichiarato critico.

Per valori negativi di γ il comportamento è l'opposto.

Com'è facile notare, il problema sta nell'identificare qual è lo scarto ritenuto più "dannoso" per il processo e di conseguenza trovare il valore γ più opportuno.

3.2.3 Bootstrap

A fianco di un indice è sempre meglio fornire anche un indicatore dell'errore, questo può essere espresso con lo standard error oppure fornendo un intervallo di confidenza con un certo grado di copertura α . Questi possono essere calcolati tramite un approccio bootstrap.

Il bootstrap è un metodo ampiamente utilizzato per ottenere delle informazioni riguardanti l'incertezza associata alle stime, anche quando non vengono fatte assunzioni parametriche sui dati.

Dato un campione di numerosità n usato per stimare una certa quantità Y , l'idea di base dell'approccio bootstrap è quello di generare B campioni di dimensione n che verranno usati per misurare l'incertezza della stima \hat{Y} . I modi di generare i B campioni differenziano il bootstrap in parametrico e non parametrico. Nel bootstrap parametrico vengono fatte assunzioni sulla distribuzione del campione, ipotizzando che questa sia $f(x, \theta)$ e viene usato il campione originale per stimare il parametro θ . A questo punto possiamo generare B campioni di numerosità n dalla distribuzione $f(x, \hat{\theta})$. Nella versione non parametrica, i campioni vengono generati mediante il processo di "campionamento con reinserimento", ossia a partire dal campione originale si ripetono per n volte le operazioni di estrazione casuale di un elemento e il rinserimento di quest'ultimo. Si può notare che lo stesso elemento può essere estratto più volte e quindi può apparire più volte all'interno di un campione.

Arrivati a questo punto abbiamo a disposizione B campioni, che vengono utilizzati per stimare i B valori di \hat{Y} , ottenendo quindi la serie $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_B$ che useremo per fornire un'indicazione sulla varianza dello stimatore e per fornire degli intervalli di confidenza con grado di fiducia $1-\alpha$. Questi si ottengono mettendo in ordine crescente i valori $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_B$ e estraendone i quantili $\frac{\alpha}{2}$ e $1-\frac{\alpha}{2}$.

3.2.4 Il codice R

Qui di seguito vediamo il codice R per calcolare l'indice S'_{pmk} . Per farlo abbiamo innanzitutto bisogno di calcolare la funzione di perdita, che chiameremo L_3 .

INPUT:

- x = media dei dati;
- t = valore Target;
- γ = parametro che regola il grado di asimmetria della funzione.

OUTPUT:

- il valore che la funzione restituisce è un numero reale che rappresenta la distanza tra la media dei dati e il valore target

CODICE

```
1 L3 = function(x,t, gamma){
2   numeratore = exp(gamma * (x-t)) - gamma * (x-t) - 1
3   return(2 * numeratore / (gamma^2) )
4 }
```

A questo punto L_3 viene usata nella funzione che calcola l'indice S'_{pmk} :

INPUT:

- x = dati;
- LSL e USL = limiti di specifica;
- t = valore Target;
- gamma = parametro che regola il grado di asimmetria della funzione;
- cdfLSL e cdfUSL = valori della funzione di ripartizione in LSL e USL.

OUTPUT:

- il valore che la funzione restituisce è il valore dell'indice di capacità del processo.

CODICE:

```
1 Spmk_primo = function(x,LSL,USL,t=0, gamma=0, cdfLSL ,
   cdfUSL){
2   mu = mean(x)
3   s = sd(x)
4   numeratore = qnorm( ( 1 + cdfUSL - cdfLSL) /2)
5   denominatore = 3* sqrt(1 + (1/s^2) * L3(mu,t,gamma) )
6   return( numeratore/denominatore )
7 }
```

Vediamo ora il codice utilizzato per il metodo bootstrap, lo scopo è quello di ottenere un un intervallo di confidenza con copertura $1-\alpha$ e un indicazione sullo standard error dell'indice di capacità considerato.

Vediamo sia il codice del bootstrap parametrico, dove il campione viene generato adattando la distribuzione λ generalizzata di Tukey, sia quello non parametrico dove il campione viene creato tramite estrazione con reinserimento.

INPUT:

- x = campione originale;
- LSL e USL = limiti di specifica;
- indice = nome dell'indice che voglio analizzare
- t = valore Target;
- gamma = parametro che regola il grado di asimmetria della funzione (solo per S'_{pmk})
- alpha = grado di fiducia dell'intervallo di confidenza

OUTPUT:

- stima = stima dell'indice secondo il metodo bootstrap
- quantile0.025 e quantile0.975 = quantili corrispondono al limite inferiore e superiore dell'intervallo di confidenza per l'indice
- sd = deviazione standard della stima dell'indice

```
1 btsParametrico = function(x, LSL,USL, indice,t=0, gamma
  =0,alpha = 0.05){
2   B = 100
3   n = length(x)
4   l.hat = fit.fkml(x)$lambda # stimo i 4 valori di lambda
5   # genero il campione BTS
6   campioniBTS = replicate(B, rgl(n, l.hat[1],l.hat[2],l.
  hat[3],l.hat[4]))
7   # calcolo dell'indice di capacita "indice" sui B
  campioni
```

```

8  indice_sim = apply(campioniBTS, MARGIN = 2,
9                      FUN = function(x) indice(x,LSL,USL,t
10                     =t, gamma = gamma) )
11  risposta = c(stima=mean(indice_sim), quantile0.025=
12              quantile(indice_sim,alpha/2),
13              quantile0.975 = quantile(indice_sim,1-
14              alpha/2), sd = sd(indice_sim))
15  return(risposta)
16 }

```

```

1  btsNonParametrico = function(x, LSL,USL, indice,t=0,
2     gamma=0,alpha = 0.05){
3     B = 100
4     n = length(x)
5     # genero il campione BTS: estrazione
6     campioniBTS = replicate(B, x[trunc(runif(n, 1, n + 1))]
7     )
8     # calcolo dell'indice di capacita "indice" sui B
9     campioni
10    indice_sim = apply(campioniBTS, MARGIN = 2,
11                      FUN = function(x) indice(x,LSL,USL,t
12                     =t, gamma = gamma))
13    risposta = c(stima=mean(indice_sim), quantile0.025=
14                quantile(indice_sim,alpha/2),
15                quantile0.975=quantile(indice_sim,1-alpha/
16                2), sd=sd(indice_sim))
17    return(risposta)
18 }

```

Capitolo 4

Applicazione degli indici ai dati

In questo capitolo andremo ad applicare gli indici che abbiamo visto finora e lo faremo sia a dati simulati, dove quindi conosciamo perfettamente la distribuzione del campione, sia a dati reali, dove quest'ultima è ignota e dobbiamo quindi approssimarla.

Per questione di semplicità, quando trattiamo dati simulati, per ogni caso andiamo a mostrare i risultati per una sola replicazione. Sono comunque stati analizzati altri casi che hanno mostrato risultati stabili, dovuti al fatto che viene usato un campione abbastanza numeroso ($n=150$).

4.1 Dati gaussiani a media variabile

Per prima cosa andiamo ad applicare gli indici ad un campione di $n=150$ osservazioni generato da una normale $N(\mu,1)$, dove μ assume valori $-2, -1, 0, 1, 2$. Poniamo come valore target $T=0$ e come limiti $LSL = -5$ e $USL = 5$. Come si comportano le distribuzioni dentro i limiti di specifica lo vediamo in figura 4.1, nella quale per questioni di leggibilità sono riportate solo le curve con media negativa; poi le conclusioni possono essere riportate anche per quelle positive, dal momento che la normale è una distribuzione simmetrica.

Quando la media è -2 , e di conseguenza anche 2 , il processo sembra non essere molto capace perchè una delle due code finisce al di fuori dei limiti,

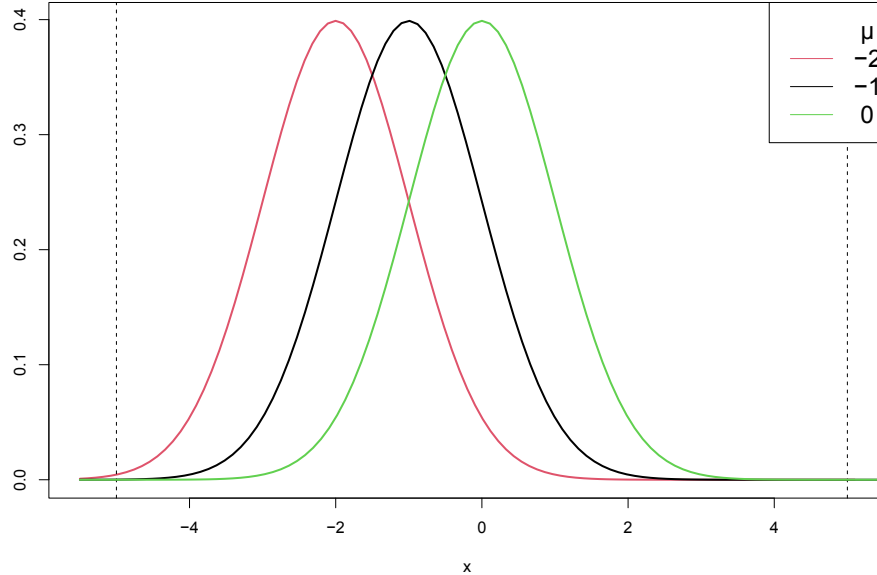


Figura 4.1: distribuzioni normali dentro i limiti -5 e 5

mentre per gli altri valori di μ la capacità sembra essere buona. Proviamo a vedere se gli indici forniscono gli stessi risultati.

Gli indici di capacità che andiamo a calcolare sono inizialmente C_{pk} , C_{pmk} e S_{pmk} ; tralasciamo C_p e C_{pm} perchè nella maggior parte dei casi considerati la media non è centrata, quindi non avrebbe senso usare indici che assumono la centratura. In secondo luogo applicheremo S'_{pmk} per diversi valori di γ . Per i primi ci aspettiamo che campioni di media opposta (ad esempio 2 e -2) forniscano risultati equivalenti, a contrario dell'ultimo che utilizzando una funzione di perdita asimmetrica assumerà un valore dipenderà da γ .

Per ogni indice, oltre al valore ottenuto dal campione, forniremo gli estremi dell'intervallo di confidenza con copertura $1-\alpha$ (che chiameremo $q_{0,025}$ e $q_{0,975}$) e la deviazione standard (sd). Questi ultimi sono ottenuti con il metodo bootstrap nella versione parametrica, dal momento che conosciamo la distribuzione dei dati. Questi dati li vediamo in tabella 4.1.

| media | indice | stima | $q_{0,025}$ | $q_{0,975}$ | sd |
|------------|-----------|-------|-------------|-------------|-------|
| $\mu = -2$ | C_{pk} | 1,010 | 0,926 | 1,138 | 0,052 |
| | C_{pmk} | 0,416 | 0,370 | 0,468 | 0,025 |
| | S_{pmk} | 0,425 | 0,379 | 0,483 | 0,028 |
| $\mu = -1$ | C_{pk} | 1,417 | 1,265 | 1,613 | 0,088 |
| | C_{pmk} | 0,945 | 0,841 | 1,094 | 0,062 |
| | S_{pmk} | 0,933 | 0,804 | 1,067 | 0,070 |
| $\mu = 0$ | C_{pk} | 1,746 | 1,545 | 2,016 | 0,109 |
| | C_{pmk} | 1,746 | 1,569 | 1,898 | 0,086 |
| | S_{pmk} | 1,665 | 1,600 | 1,667 | 0,018 |
| $\mu = 1$ | C_{pk} | 1,412 | 1,252 | 1,635 | 0,094 |
| | C_{pmk} | 0,985 | 0,869 | 1,125 | 0,068 |
| | S_{pmk} | 0,973 | 0,859 | 1,102 | 0,065 |
| $\mu = 2$ | C_{pk} | 1,061 | 0,941 | 1,229 | 0,074 |
| | C_{pmk} | 0,460 | 0,413 | 0,518 | 0,028 |
| | S_{pmk} | 0,466 | 0,409 | 0,534 | 0,032 |

Tabella 4.1: indici C_{pk} , C_{pmk} e S_{pmk} al variare della media μ

Come ci aspettavamo per valori generati da distribuzioni con medie opposte i risultati sono equivalenti, a meno di un piccolo errore. La prima cosa che si nota è che \hat{C}_{pk} è sempre più grande rispetto agli altri indici, tanto che anche quando la media è 2 o -2 questo è pari a 1, mentre \hat{C}_{pmk} e \hat{S}_{pmk} assumono valori molto bassi, entrambi circa 0,4. Quando lo scarto tra μ e t diminuisce, gli indici assumono sempre valori più elevati. Nel caso in cui la media del processo coincide con il valore target, tutti gli indici ci dicono che il processo è molto capace, in quanto assumono valori che si aggirano intorno a 1,7. È interessante notare come, in quest'ultimo caso, l'indice \hat{C}_{pk} è pari a

1,74 con un intervallo di confidenza compreso tra 1,57 e 1,90, questo significa che il processo non è ancora abbastanza capace per entrare a far parte dei cosiddetti "Sei Sigma", bisognerebbe quindi, oltre a centrare la media tra i limiti, riuscire a ridurre la varianza.

Con gli stessi dati calcoliamo l'indice \hat{S}'_{pmk} , con γ uguale a 0.01, 0.5, 1, 5, 10. Consideriamo solo valori maggiori di zero, quindi andremo a pesare maggiormente scarti positivi dal valore target. Le stesse conclusioni si possono trarre considerando valori negativi. Come si comporta l'indice lo vediamo in tabella 4.2.

Guardiamo i primi due blocchi della tabella, dove la media, e di conseguenza anche lo scarto da t , è negativa. La prima cosa che si nota è che all'aumentare di γ aumenta anche il valore dell'indice. Questo accade perchè, come si vede dalla figura 3.1, quando γ aumenta gli scarti negativi vengono pesati sempre meno, facendo quindi assumere all'indice valori maggiori. Il contrario accade se μ diventa positivo, infatti all'aumentare di γ gli scarti maggiori di zero vengono pesati sempre di più. Il ruolo di γ è fondamentale: se prendiamo $\mu = 1$, per valori piccoli di γ il processo potrebbe essere capace, in quanto l'intervallo di confidenza comprende valori maggiori di 1, viceversa per valori grandi del parametro la capacità diventa molto piccola. Per questo motivo bisogna scegliere un valore di gamma adeguato a quello che si vuole misurare. L'ultimo caso che ci resta da vedere è quello in cui la media del processo coincide esattamente con il valore target, in questo caso, qualsiasi sia il valore di γ , l'indice ci fornisce valori stabili che indicano una buona capacità del processo.

Inoltre vale la pena osservare è che quando γ è molto piccolo, ad esempio 0,01, \hat{S}'_{pmk} assume valori simili a \hat{S}_{pmk} . Questo accade perchè, come abbiamo visto nel capitolo 3, per γ che tende a zero, le funzioni di perdita utilizzate sono equivalenti e quindi i due indici tendono ad essere uguali.

| media | γ | stima | $q_{0,025}$ | $q_{0,975}$ | sd |
|------------|----------|-------|-------------|-------------|--------|
| $\mu = -2$ | 0,01 | 0,431 | 0,384 | 0,494 | 0,028 |
| | 0,5 | 0,489 | 0,436 | 0,553 | 0,030 |
| | 1 | 0,541 | 0,492 | 0,591 | 0,027 |
| | 5 | 0,766 | 0,697 | 0,820 | 0,034 |
| | 10 | 0,863 | 0,815 | 0,916 | 0,027 |
| $\mu = -1$ | 0,01 | 0,451 | 0,393 | 0,508 | 0,034 |
| | 0,5 | 0,509 | 0,453 | 0,571 | 0,033 |
| | 1 | 0,562 | 0,485 | 0,628 | 0,037 |
| | 5 | 0,789 | 0,703 | 0,845 | 0,035 |
| | 10 | 0,888 | 0,828 | 0,942 | 0,030 |
| $\mu = 0$ | 0,01 | 1,666 | 1,616 | 1,667 | 0,014 |
| | 0,5 | 1,666 | 1,610 | 1,667 | 0,018 |
| | 1 | 1,666 | 1,600 | 1,667 | 0,018 |
| | 5 | 1,666 | 1,611 | 1,667 | 0,015 |
| | 10 | 1,666 | 1,627 | 1,667 | 0,012 |
| $\mu = 1$ | 0,01 | 0,972 | 0,849 | 1,085 | 0,059 |
| | 0,5 | 0,930 | 0,793 | 1,060 | 0,067 |
| | 1 | 0,882 | 0,752 | 1,032 | 0,075 |
| | 5 | 0,400 | 0,289 | 0,579 | 0,082 |
| | 10 | 0,071 | 0,0369 | 0,152 | 0,032 |
| $\mu = 2$ | 0,01 | 0,465 | 0,405 | 0,516 | 0,030 |
| | 0,5 | 0,401 | 0,342 | 0,462 | 0,030 |
| | 1 | 0,334 | 0,285 | 0,397 | 0,030 |
| | 5 | 0,026 | 0,018 | 0,039 | 0,006 |
| | 10 | 0,001 | 0,001 | 0,001 | 0,0001 |

Tabella 4.2: indice S'_{pmk} al variare della media μ

4.2 Dati Poisson a media variabile

Andiamo ora fare le stesse analisi per un campione di $n = 150$ osservazioni generato da una distribuzione di Poisson con media λ pari a 6, 7, 8, 9 o 10. I limiti di specifica sono 0 e 16 mentre il valore target è 8. Prima di iniziare l'analisi vale la pena far notare che stiamo analizzando una distribuzione asimmetrica e a valori discreti. Inoltre, a differenza dell'esempio di prima dove cambiava solamente la media, ora a cambiare è anche la varianza, dal momento che in una distribuzione Poisson media e varianza coincidono.

In figura 4.2 si può vedere come si comporta la Poisson al variare di λ , le linee tratteggiate indicano i limiti di specifica. Dal grafico si può dedurre che i processi più capaci siano quelli in cui la media vale 6 e 7. La curva nera, nonostante abbia $\mu=T$, e quindi produrrà osservazioni che maggiormente saranno vicine al valore desiderate, ha la coda destra che cade leggermente al di sopra del limite USL, per questo motivo sembra essere leggermente meno capace. Meno performanti sono invece i processi in cui la media vale 9 e 10, in questi casi le code sinistre sono abbondantemente al di sopra del limite di specifica superiore, per questo motivo ci aspettiamo che anche gli indici di capacità assumano valori inferiori di 1. Osserviamo inoltre che, dal momento che la Poisson è una distribuzione che ha come supporto i numeri naturali, la probabilità di osservare valori più piccoli di LSL è zero.

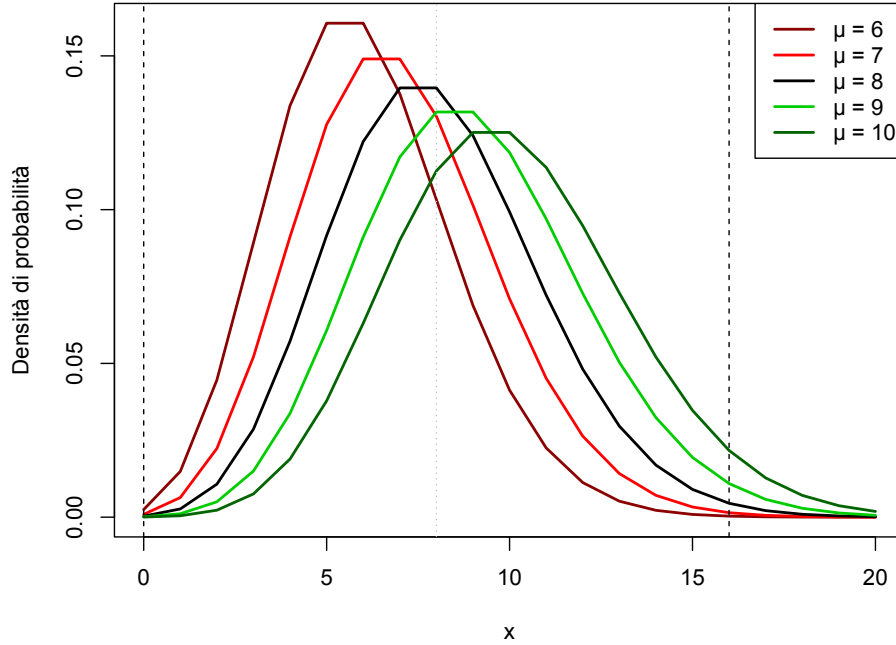


Figura 4.2: distribuzioni Poisson dentro i limiti 0 e 16

Andiamo ora a calcolare \hat{C}_{pk} , \hat{C}_{pmk} , \hat{S}_{pmk} e \hat{S}'_{pmk} : come per il caso gaussiano tralasciamo gli indici che assumono centratura della media. I risultati dei primi tre li vediamo nella tabella 4.3 L'indice C_{pk} ci conferma quanto detto: valori più alti dell'indice li abbiamo per le medie più basse. Questo perchè tiene conto solo del comportamento della distribuzione all'interno dei limiti di specifica e non della differenza delle osservazioni dal valore target. Il discorso è differente per gli altri due indici, dove utilizzando una funzione di perdita, assumono valori più elevati quando μ uguale a 8, e valori sempre più bassi mano a mano che ci allontaniamo da T. Vale la pena sottolineare che se prendiamo due processi in cui lo scarto della media da T è uguale, ad esempio i processi con μ pari a 7 e a 9, l'indice non assume lo stesso valore, ma ci dice che quello con media più bassa è più capace. Questo accade perchè, oltre ad avere varianza inferiore, la probabilità di eccedere dal limite superiore è più bassa mentre resta nulla quella di eccedere da quello inferiore.

| media | indice | stima | $q_{0,025}$ | $q_{0,975}$ | sd |
|----------------|-----------|-------|-------------|-------------|--------|
| $\lambda = 6$ | C_{pk} | 1,018 | 0,843 | 1,460 | 0,161 |
| | C_{pmk} | 0,643 | 0,574 | 0,741 | 0,046 |
| | S_{pmk} | 0,772 | 0,674 | 0,851 | 0,0474 |
| $\lambda = 7$ | C_{pk} | 1,036 | 0,7848 | 1,373 | 0,138 |
| | C_{pmk} | 0,855 | 0,775 | 0,975 | 0,054 |
| | S_{pmk} | 0,970 | 0,907 | 0,991 | 0,022 |
| $\lambda = 8$ | C_{pk} | 0,841 | 0,650 | 1,121 | 0,129 |
| | C_{pmk} | 0,949 | 0,852 | 1,063 | 0,055 |
| | S_{pmk} | 0,951 | 0,898 | 0,989 | 0,025 |
| $\lambda = 9$ | C_{pk} | 0,724 | 0,564 | 0,990 | 0,111 |
| | C_{pmk} | 0,740 | 0,645 | 0,872 | 0,063 |
| | S_{pmk} | 0,790 | 0,709 | 0,880 | 0,044 |
| $\lambda = 10$ | C_{pk} | 0,571 | 0,405 | 0,781 | 0,098 |
| | C_{pmk} | 0,591 | 0,497 | 0,705 | 0,056 |
| | S_{pmk} | 0,653 | 0,569 | 0,738 | 0,046 |

Tabella 4.3: indici C_{pk} , C_{pmk} e S_{pmk} al variare di λ

Calcoliamo ora \hat{S}'_{pmk} dove attribuiamo a γ gli stessi valori considerati nell'esempio di dati gaussiani. Nella tabella 4.4 vediamo il comportamento dell'indice. Quest'ultima ci suggerisce che il processo più capace è quello con $\lambda=7$: qualsiasi sia il valore di γ , \hat{S}'_{pmk} assume valori prossimi a 1. Quando $\lambda=6$, all'aumentare di gamma aumenta anche la capacità, questo succede perchè vengono pesati sempre meno gli scarti negativi, che sono molto più frequenti di quelli positivi. Quando la media coincide con T, l'indice assume valori abbastanza costanti al variare di γ , mentre gli ultimi due processi sono sempre ritenuti poco capaci, e a maggior ragione lo sono all'aumentare del parametro, in quanto vengono pesati molto i numerosi scarti positivi.

| media | γ | stima | $q_{0,025}$ | $q_{0,975}$ | sd |
|----------------|----------|--------|-------------|-------------|--------|
| $\lambda = 6$ | 0,01 | 0,773 | 0,684 | 0,880 | 0,051 |
| | 0,5 | 0,817 | 0,737 | 0,888 | 0,043 |
| | 1 | 0,850 | 0,772 | 0,905 | 0,037 |
| | 5 | 0,946 | 0,899 | 0,983 | 0,022 |
| | 10 | 0,972 | 0,937 | 1,004 | 0,0195 |
| $\lambda = 7$ | 0,01 | 0,970 | 0,922 | 0,991 | 0,020 |
| | 0,5 | 0,978 | 0,929 | 0,993 | 0,017 |
| | 1 | 0,986 | 0,946 | 0,996 | 0,013 |
| | 5 | 1,012 | 0,995 | 1,015 | 0,006 |
| | 10 | 1,022 | 1,004 | 1,025 | 0,005 |
| $\lambda = 8$ | 0,01 | 0,951 | 0,902 | 0,983 | 0,025 |
| | 0,5 | 0,951 | 0,892 | 0,989 | 0,027 |
| | 1 | 0,951 | 0,893 | 0,986 | 0,025 |
| | 5 | 0,951 | 0,899 | 0,991 | 0,027 |
| | 10 | 0,951 | 0,897 | 0,987 | 0,091 |
| $\lambda = 9$ | 0,01 | 0,790 | 0,710 | 0,883 | 0,046 |
| | 0,5 | 0,781 | 0,688 | 0,859 | 0,049 |
| | 1 | 0,770 | 0,632 | 0,881 | 0,061 |
| | 5 | 0,508 | 0,180 | 0,886 | 0,197 |
| | 10 | 0,090 | 0,017 | 0,487 | 0,134 |
| $\lambda = 10$ | 0,01 | 0,652 | 0,569 | 0,736 | 0,046 |
| | 0,5 | 0,623 | 0,527 | 0,717 | 0,050 |
| | 1 | 0,582 | 0,444 | 0,705 | 0,066 |
| | 5 | 0,092 | 0,021 | 0,246 | 0,066 |
| | 10 | 0,0027 | 0,0002 | 0,0191 | 0,0054 |

Tabella 4.4: indice S'_{pmk} al variare di λ

4.3 Dati reali

Andiamo ora ad applicare i quattro indici non più a dati simulati, ma a dati reali, in cui non conosciamo la vera distribuzione e dobbiamo in qualche modo ricavarla.

4.3.1 Descrizione dei dati

I dati che andiamo ad analizzare fanno riferimento al diametro di $n=411$ tubicini in gomma che devono essere montati in un apparecchiatura medicale. Questi per essere conformi alle specifiche, devono avere un diametro di $2\text{mm} \pm 0,1\text{mm}$. Assumiamo che il valore target sia la media dei due limiti di specifica, ossia 2mm .

4.3.2 Analisi della capacità del processo

Assumiammo che i dati siano stati raccolti con il processo in controllo. Per analizzarne la capacità procediamo come abbiamo fatto per gli altri due esempi precedenti: prima vediamo come si comportano le osservazioni dentro i limiti di specifica e poi calcoliamo gli indici di capacità.

Per il primo passo andiamo a costruire un istogramma di frequenza con i dati del campione. Come possiamo notare in figura 4.3, la distribuzione dei dati non può essere normale, perchè la curva destra è troppo pesante. Questo ci viene confermato anche dal test di Shapiro Wilk che, con un p-value di $2.863\text{e-}15$, ci porta a rifiutare l'ipotesi nulla di gaussianità dei dati. Per questo motivo è stato deciso di adattare i dati con la distribuzione γ generalizzata di Tukey, la quale, come si può vedere dalla curva nel grafico, si adatta molto bene ai dati. Questa distribuzione la useremo per calcolare i quantili da inserire nelle formule degli indici di capacità. Guardando l'istogramma e la curva, il processo sembra essere molto capace in quanto, oltre ad avere la media abbastanza centrata dentro i limiti di specifica ($\hat{\mu} = 1,98735$), entrambe le code non si protraggono oltre LSL e USL. Vediamo se gli indici di capacità confermano questa intuizione.

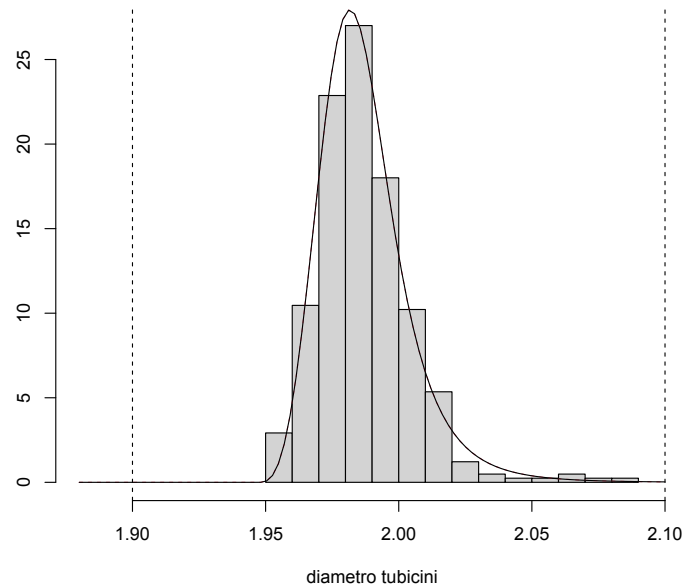


Figura 4.3: istogramma dei tubicini con andamento della distribuzione γ

Iniziamo l'analisi degli indici calcolando C_{pk} , C_{pmk} , S_{pmk} e S'_{pmk} . Dal momento che la distribuzione dei dati non è nota, il primo indice lo andiamo a calcolare in tre modi differenti:

- $C_{pk-norm}$: assumendo normalità, la formula è quella classica 2.2;
- $C_{pk-\lambda}$: *estraendo* i quantili 0,00135 e 0,99865 dei dati, formula ??;
- C_{pk-q} : adattando ai dati la distribuzione lambda generalizzata di Tukey e calcolandone i quantili 0,00135 e 0,99865.

I risultati li vediamo in tabella 4.5. Per prima cosa notiamo che $C_{pk-norm}$ ci indica che il processo è molto capace, però ricordiamo che l'assunzione di normalità dei dati viene violata, quindi il risultato non è molto affidabile. Molto più affidabili sono gli indici $C_{pk-\lambda}$ e C_{pk-q} i quali assumono valori molto simili tra loro e ci dicono che il processo è abbastanza capace, anche se andando a vedere gli intervalli di confidenza, questi potrebbe assumere valori inferiori a 1.

| indice | stima | $q_{0,025}$ | $q_{0,975}$ | sd |
|------------------|-------|-------------|-------------|-------|
| $C_{pk-norm}$ | 1,658 | 1,471 | 1,880 | 0,103 |
| $C_{pk-\lambda}$ | 1,181 | 0,838 | 1,660 | 0,213 |
| C_{pk-q} | 1,169 | 0,827 | 1,930 | 0,292 |
| C_{pmk} | 1,345 | 1,247 | 1,459 | 0,053 |
| S_{pmk} | 0,925 | 0,846 | 1,136 | 0,079 |

Tabella 4.5: indici C_{pk} , C_{pmk} e S_{pmk} per i tubicini in gomma

Andiamo ora ad analizzare l'indice S'_{pmk} il quale essendo più flessibile per quanto riguarda la funzione di perdita, potrebbe darci qualche informazione in più sulla capacità del processo. In questo studio andiamo a calcolare l'indice anche per valori di gamma negativi. Facciamo questo perchè a differenza di quanto fatto in precedenza, abbiamo un campione che è fisso e non possiamo quindi "muovere" la media intorno al valore target. I risultati li possiamo vedere in tabella 4.6.

Notiamo che qualsiasi sia il valore di γ , il processo è sempre ritenuto poco capace, in quanto l'indice si assesta sempre in un valore poco superiore a 0,9.

| γ | stima | $q_{0,025}$ | $q_{0,975}$ | sd |
|----------|-------|-------------|-------------|-------|
| -10 | 0,918 | 0,837 | 0,988 | 0,039 |
| -5 | 0,922 | 0,843 | 0,993 | 0,039 |
| -1 | 0,924 | 0,841 | 0,996 | 0,038 |
| -0,5 | 0,925 | 0,848 | 0,993 | 0,037 |
| -0,1 | 0,925 | 0,839 | 0,997 | 0,041 |
| -0,01 | 0,925 | 0,844 | 0,989 | 0,038 |
| 0,01 | 0,925 | 0,846 | 0,989 | 0,037 |
| 0,5 | 0,925 | 0,846 | 0,987 | 0,038 |
| 1 | 0,926 | 0,843 | 0,993 | 0,038 |
| 5 | 0,928 | 0,856 | 0,992 | 0,035 |
| 10 | 0,932 | 0,848 | 0,996 | 0,037 |

Tabella 4.6: indice S'_{pmk} per i tubicini in gomma per diversi valori di γ

C'è un modo per poter rendere il processo più capace? Dal momento che la bassa capacità è probabilmente dovuta alla pesante curva destra che si avvicina troppo a USL, qual'ora fosse possibile, possiamo pensare di abbassare la media del processo di 0,035mm. In questo caso tutta la distribuzione verrebbe traslata verso sinistra. In figura 4.4 vediamo come si comporta la nuova distribuzione all'interno dei limiti di specifica, dalla quale possiamo notare che la coda sinistra resta abbondantemente al di sopra di LSL mentre quella destra si è allontanata da USL. I risultati degli indici di capacità quando andiamo a sottrarre ai dati il valore 0,035 li vediamo in tabella 4.7.

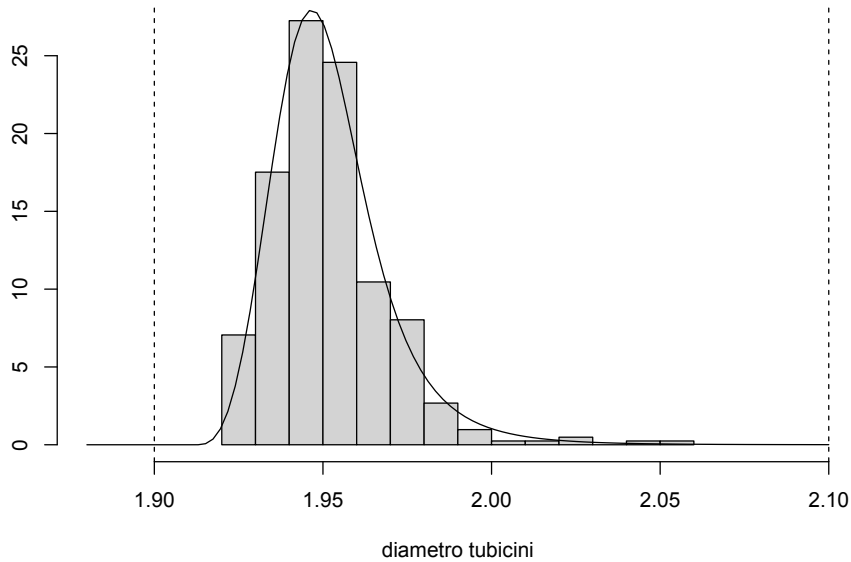


Figura 4.4: Nuova distribuzione dentro i limiti di specifica

Si può notare che non tutti gli indici hanno avuto lo stesso comportamento. In particolare sono migliorati $\hat{C}_{pk-\lambda}$ e \hat{C}_{pk-q} , in quanto sono gli indici che tengono conto dei quantili all'interno dei limiti di specifica, mentre \hat{C}_{pmk} , \hat{S}_{pmk} e \hat{S}'_{pmk} sono peggiorati di molto, perchè questi tengono conto della distanza tra la media dei dati dal valore target, la quale è cambiata da 1,987mm a 1,952mm.

Abbiamo quindi avuto un miglioramento per quanto riguarda il fatto di stare all'interno dei limiti di specifica, però così facendo abbiamo dovuto allontanare la media del diametro dei tubicini prodotti dal valore target di 2mm: bisognerà quindi valutare quale delle due situazioni è considerata migliore.

| indice | stima |
|------------------|-------|
| $C_{pk-norm}$ | 0,993 |
| $C_{pk-\lambda}$ | 1,540 |
| C_{pk-q} | 1,525 |
| C_{pmk} | 0,344 |
| S_{pmk} | 0,434 |

| S'_{pmk} | | | |
|------------|-------|----------|-------|
| γ | stima | γ | stima |
| -10 | 0,404 | 0,01 | 0,434 |
| -5 | 0,419 | 0,1 | 0,435 |
| -1 | 0,431 | 0,5 | 0,436 |
| -0,5 | 0,433 | 1 | 0,437 |
| -0,1 | 0,434 | 5 | 0,443 |
| -0,01 | 0,434 | 10 | 0,464 |

Tabella 4.7: Indici ricalcolati per i nuovi dati traslati

Conclusione

In questa relazione sono stati analizzati indici diversi con lo scopo di misurare la capacità di un processo, con particolare attenzione a quello che abbiamo chiamato S'_{pmk} . Ogni indice, dopo essere stato descritto, è stato applicato sia a dati simulati che a dati reali. Fare questa differenza è importante, dal momento che se i dati sono simulati, il calcolo di quantili estremi e valori della funzione di ripartizione fornisce risultati accurati, mentre quando sono reali questi vanno stimati empiricamente, con la possibilità di ottenere stime non buone. Ad esempio nel calcolo del numeratore degli indici S_{pmk} e S'_{pmk} viene utilizzata $F(\text{USL})$ e $F(\text{LSL})$, ovvero il valore della funzione di ripartizione dei limiti di specifica. Quando però abbiamo un campione in cui nessuna osservazione esce dai limiti, questi assumeranno rispettivamente valori 1 e 0, con la conseguenza di dover calcolare l'inversa della funzione di probabilità della normale standard in 1, avendo quindi come risultato infinito. Per superare questo problema fingeremo di avere due osservazione al di fuori dai limiti così da non ottenere $F(\text{USL})=1$ e $F(\text{LSL})=0$.

Come è stato più volte detto durante la relazione, γ ricopre un ruolo fondamentale per il calcolo di S'_{pmk} , in quanto, per valori diversi del parametro, l'indice può passare da indicare il processo come molto capace o poco capace. Trovare il giusto valore del γ è però complicato, in quanto è semplice stabilirne il segno, ma non il valore preciso.

Infine come abbiamo visto nell'ultima tabella può accadere che indici diversi vadano a dare informazioni contrastanti tra loro. Per questo motivo è importante conoscere le caratteristiche dell'indice che si va ad utilizzare e

capire quali sono le cause che portano a determinati risultati; inoltre affiancare alla stima un grafico in cui si vedono i limiti di specifica può aiutare a comprendere bene la capacità del processo.

Bibliografia

- Erfanian, M. and Sadeghpour Gildeh, B. (2021) A new capability index for non-normal distributions based on linex loss function. *Quality Engineering*, **33**, 76–84.
- Montgomery, D. C. (2013) *Introduction to Statistical Quality Control*. New York: Wiley, 7th edn.
- Varian, H. R. (1975) A bayesian approach to real estate assessment. In *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage* (eds. S. Fienberg and A. Zellner), 195–208. Amsterdam: North-Holland.
- Zellner, A. (1986) Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, **81**, 446–451.