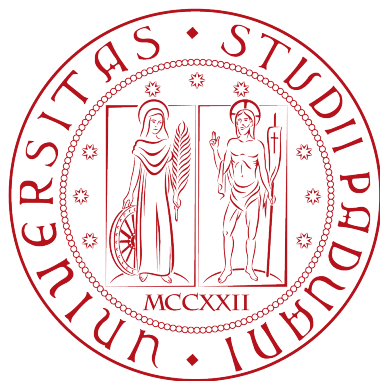


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



STATISTICA ITERAZIONE  
**GRUPPO CONDENSATORI**

Lorenzo Cifelli  
Francesco Martella  
Gianmarco Spadetto  
Daniele Zampieri

Anno Accademico 2021/2022

# Indice

<b>1</b>	<b>Introduzione</b>	<b>2</b>
1.1	Descrizione del problema . . . . .	2
1.2	Obiettivi . . . . .	2
<b>2</b>	<b>Parte 1 - Analisi descrittiva</b>	<b>4</b>
2.1	Indici riassuntivi . . . . .	4
2.2	Studio degli andamenti . . . . .	4
2.3	Distribuzione Marginale . . . . .	6
2.4	Dipendenza . . . . .	7
2.5	Concetti fondamentali individuati . . . . .	7
<b>3</b>	<b>Parte 2 - Modellazione</b>	<b>8</b>
3.1	Modello anova . . . . .	8
3.2	Modellazione della relazioni tra fasi . . . . .	8
<b>4</b>	<b>Parte 3 - Difettosità finale</b>	<b>13</b>
4.1	Variabile latente . . . . .	13
4.2	Risultati . . . . .	14
4.3	Limiti . . . . .	16
<b>5</b>	<b>Conclusioni</b>	<b>17</b>
<b>6</b>	<b>Appendice A: Materiale Supplementare</b>	<b>18</b>
6.1	1 - Individuazione di andamenti . . . . .	18
6.2	2 - Summary del modello ANOVA . . . . .	22
6.3	3 - Summary & Grafici dei Modelli Beta-binomiali . . . . .	23
<b>7</b>	<b>Appendice B: Metodo dei momenti</b>	<b>29</b>

# 1 Introduzione

## 1.1 Descrizione del problema

I dati sono stati raccolti in una azienda che produce condensatori elettrolitici in alluminio (Aluminum Electrolytic Capacitor, AEC). Si tratta di una produzione di massa. In particolare, in ogni ciclo di produzione vengono prodotti migliaia di pezzi. Il processo di produzione richiede 4 fasi che devono essere svolte in sequenza e che possiamo indicare con il loro nome inglese, ovvero:

- Rolling
- Soaking
- Assembly
- Cleaning

Ognuna delle quattro fasi può introdurre dei difetti sia funzionali che estetici.

Per ogni ciclo di produzione, 200 condensatori sono scelti casualmente alla fine di ogni fase e classificati come difettosi o meno. Il file *AEC.dat*, di cui di seguito si presenta un estratto, contiene per 1180 cicli di produzione il numero totale di pezzi difettosi trovato in ogni fase.

Rolling	Soaking	Assembly	Cleaning
0	3	6	12
0	0	5	6
5	1	4	7
2	2	11	13
2	1	6	9
1	2	4	7

## 1.2 Obiettivi

Il primo obiettivo è costruire un modello che racconti come la probabilità di essere difettoso si evolve da fase a fase.

Un obiettivo secondario, ma importante, è anche capire come sia possibile stimare la difettosità finale di un ciclo di produzione (ovvero quella rilevata direttamente dopo la fase di *Cleaning*) utilizzando tutte e 4 le misure (e se ne vale la pena).

Osservazioni aggiuntive:

1. I pezzi trovati difettosi durante il controllo sono eliminati. Nonostante questo gli ingegneri responsabili della produzione sono fortemente convinti che la probabilità di essere difettoso cresce da fase a fase per ogni ciclo di produzione. Ovvero, ogni fase (incluso il controllo–eliminazione difettosi trovati) aggiunge difetti, non li cura. Dato che i 200 pezzi ispezionati sono differenti da fase a fase questa monotonia può non essere rispettata dai dati.

2. Sempre secondo gli ingegneri responsabili della produzione ci sono dei fattori stocastici non da loro controllabili che “affliggono” i differenti cicli di produzione. La probabilità di essere difettoso dovrebbe quindi essere non solo differente da fase a fase ma anche da ciclo a ciclo. Dopo qualche analisi preliminare potrebbe essere utile verificare anche in maniera più o meno formale questo aspetto.

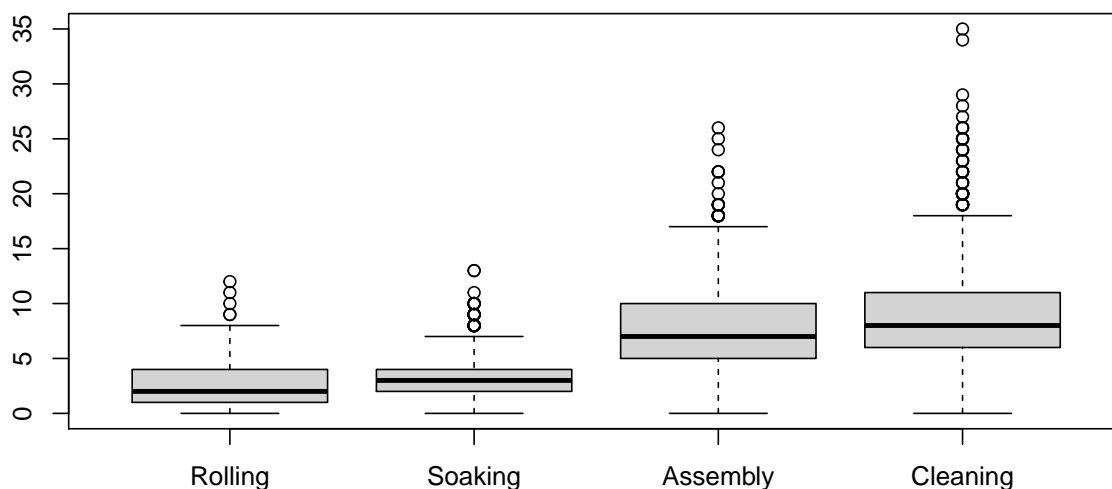
## 2 Parte 1 - Analisi descrittiva

### 2.1 Indici riassuntivi

In prima istanza, si riportano media e varianza del numero di condensatori difettosi per ogni fase.

	Media	Varianza
Rolling	2.572881	3.045574
Soaking	3.183051	4.156455
Assembly	7.696610	15.965552
Cleaning	8.777119	22.156389

Emerge come entrambi gli indici crescano da fase da fase, specialmente evidenziando un salto tra la seconda (*Soaking*) e la terza (*Assembly*). Le stesse caratteristiche si possono notare dalla figura seguente, dove vengono riportati i boxplot per le 4 fasi. Inoltre si evidenzia che per l'ultima fase ci siano molti valori estremi (una probabile indicazione che la quarta fase sia quella più critica della produzione).

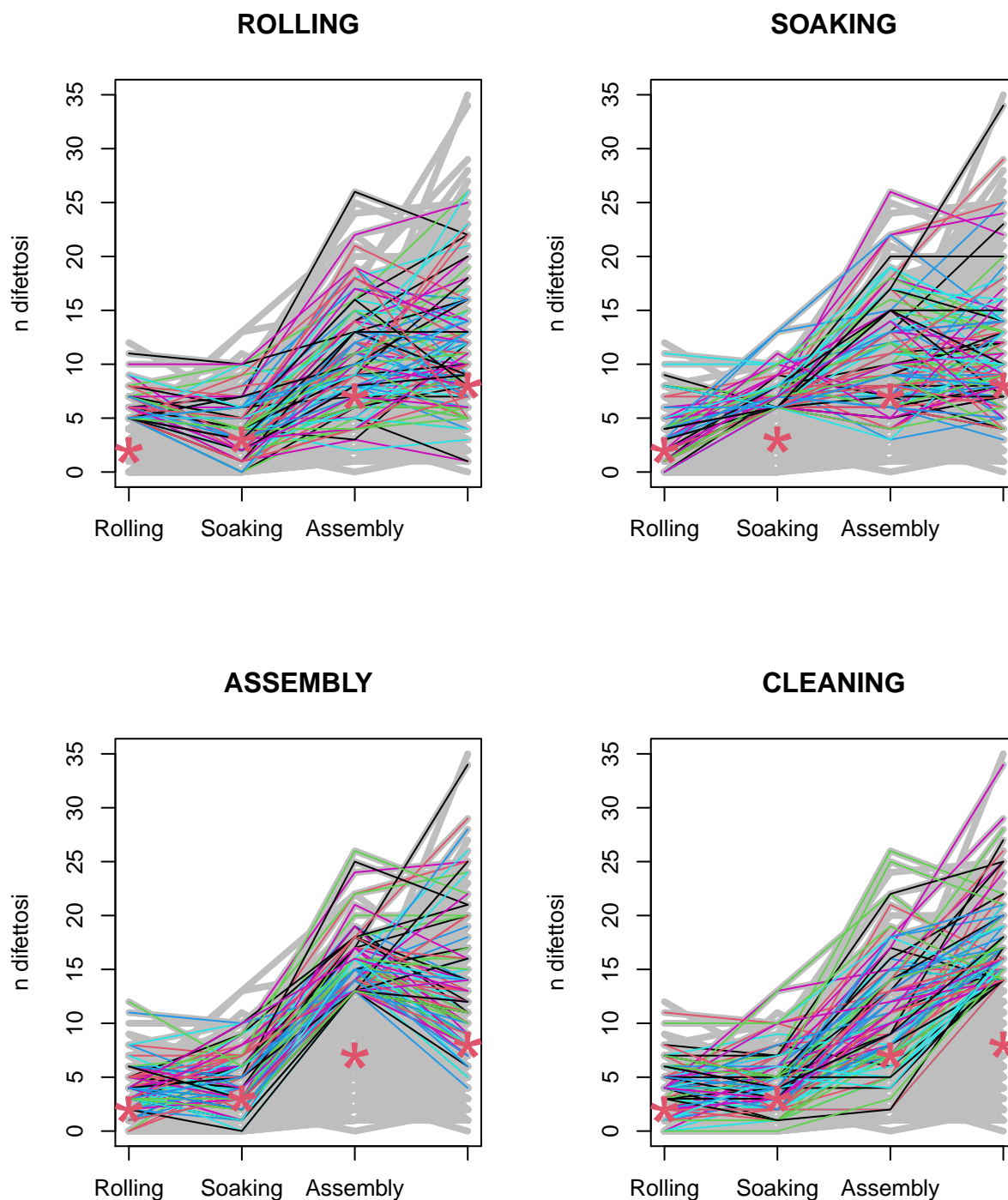


Queste informazioni ci portano a ritenere sensata l'ipotesi degli ingegneri su un possibile aumento della probabilità di difettosità da fase a fase.

### 2.2 Studio degli andamenti

Successivamente si è indagato la presenza di comportamenti ricorrenti tra le fasi. Per esplorare questo aspetto si sono rappresentati graficamente 100 cicli di produzione selezionati casualmente tra quelli con valori al di sopra del 85° quantile per ogni fase. Sono stati

considerati tali valori perché ritenuti problematici dal punto di vista del processo produttivo. Inoltre, viene anche segnato la mediana di ogni fase con un asterisco rosso.

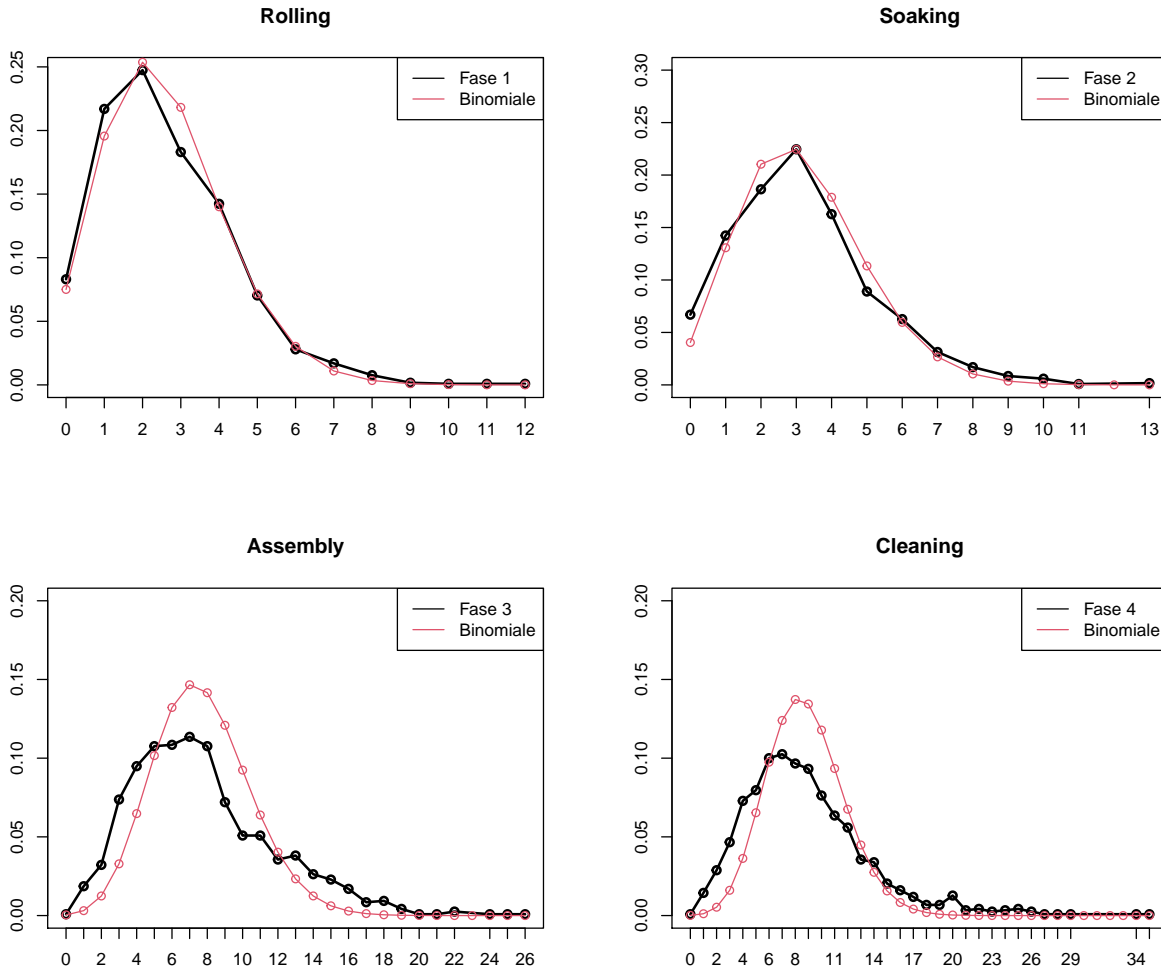


Dalla figura si evince che un numero elevato di pezzi difettosi, nelle prime due fasi, non ha

andamenti evidenti nelle fasi successive se non per un valore leggermente al di sopra delle rispettive mediane. Similmente, i valori più elevati in fase di Assembly tendono ad essere al di sopra della mediana anche in Cleaning. Infine, i cicli con un maggior numero di difettosi in Cleaning presentano un salto elevato rispetto alla fase precedente.

## 2.3 Distribuzione Marginale

Nel determinare quale modello probabilistico sia più adatto a descrivere il fenomeno, si sono studiate le distribuzioni marginali delle 4 fasi. Nella figura successiva vengono confrontate le distribuzioni osservate con quella di una distribuzione binomiale con media pari alla media campionaria. Mentre per le fasi 1 e 2 la distribuzione binomiale sembra adeguata, per le fasi 3 e 4 si nota uno forte scostamento. Questo indica come dalla fase 2 alla fase 3 il cambiamento non coinvolge solamente il parametro di posizione, ma l'intera distribuzione.

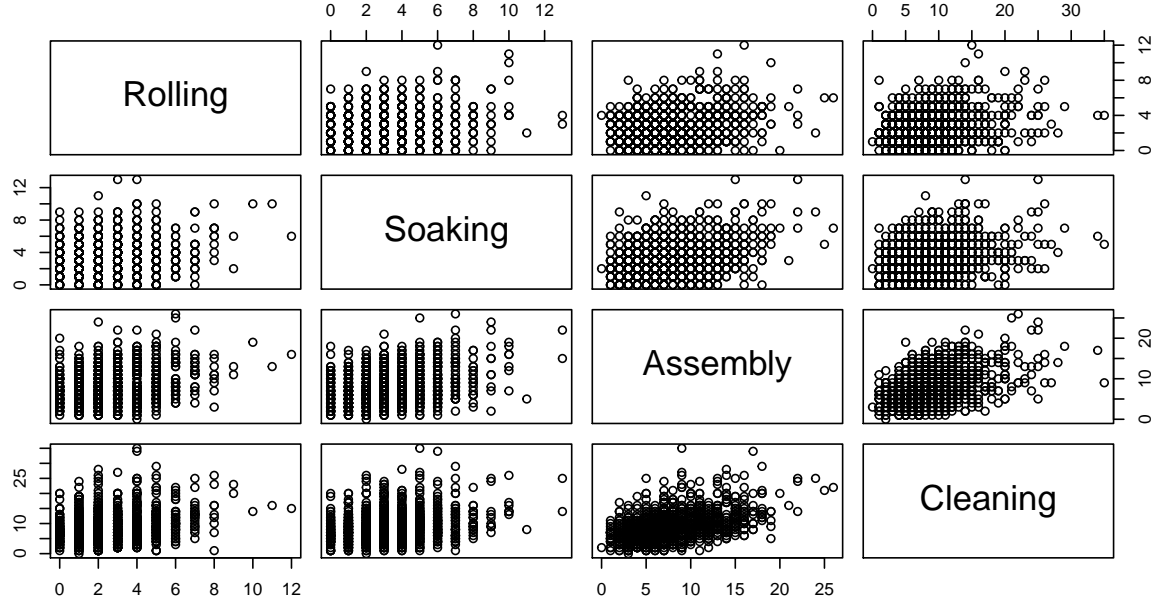


Alla luce di questo risultato, si è stimato il parametro di dispersione attraverso il metodo dei momenti (vedasi Appendice C). Si ottiene che la distribuzione delle fasi 3 e 4 hanno rispettivamente una varianza 2.16 e 2.64 volte quella di una distribuzione binomiale, che ci

conferma il forte scostamento annunciato in precedenza.

## 2.4 Dipendenza

Infine si è analizzata la struttura di dipendenza tra le 4 fasi.



Dalla figura precedente non si nota un'evidente struttura di correlazione. Questo risulta contro-intuitivo rispetto alla natura del problema: ci si aspetta infatti che il numero di difettosi cresca o rimanga uguale da fase a fase. Tale dinamica deriva dal campionamento, infatti ad ogni fase i 200 pezzi estratti vengono distrutti e i pezzi raccolti nella fase successiva sono nuovi. Si può quindi ottenere un numero di difettosi nella fase successiva anche minore.

## 2.5 Concetti fondamentali individuati

In conclusione, si evidenziano i seguenti aspetti riscontrati:

- presenza di un aumento di difettosi rilevati dal passaggio da una fase all'altra, specialmente tra la seconda e la terza si evidenzia un salto;
- presenza di sovradisersione crescente;
- la distribuzione beta-binomiale risulta la distribuzione più adatta
- correlazione lineare più evidente tra la terza e quarta fase



## 3 Parte 2 - Modellazione

### 3.1 Modello anova

In prima istanza si è verificato che le medie e le varianze del numero di difettosi nella 4 fasi siano significativamente diverse tra di loro. Si è quindi stimato un modello anova assumendo, alla luce dei risultati ottenute dell'analisi esplorativa sulle distribuzioni marginali, una distribuzione Beta-Binomiale per il numero di difettosi. Posto  $Y_{ij}$  il numero di difettosi nel ciclo  $i$ -esimo nella fase  $j$ -esima, il modello che ne consegue è il seguente:

$$\begin{aligned} Y_{ij} &\sim \text{BetaBin}(200, p_j, \psi_j) \\ p_j &= \text{logit}^{-1}(\beta_0 + \beta_1 I(j=1) + \beta_2 I(j=2) + \beta_3 I(j=4)) \\ \psi_j &= \text{logit}^{-1}(\gamma_0 + \gamma_1 I(j=1) + \gamma_2 I(j=2) + \gamma_3 I(j=4)) \end{aligned}$$

per  $i=1, \dots, 1180$ ,  $j=1, \dots, 4$ .

Dal modello risulta un effetto significativo delle varie fasi di produzione sia nel probabilità del numero di difetti che nella varianza. Si riportano di seguito le stime delle probabilità e della varianza stimate per ciascuna fase.

	Prob.	Dev. st.
Rolling	0.013	1.745
Soaking	0.016	2.038
Assembly	0.038	3.915
Cleaning	0.044	4.558

Si nota che la probabilità di difettosità stimata aumenta di fase in fase raggiungendo il massimo nella fase di Cleaning. Tale probabilità rimane comunque sempre al di sotto del 5%. Anche la variabilità segue lo stesso andamento come già dedotto dalle analisi esplorative.

### 3.2 Modellazione della relazioni tra fasi

Per studiare le relazioni tra i difettosi nelle 4 fasi, si è scelto di propendere per un procedimento sequenziale, ossia si è modellata la probabilità di guasto di una fase sfruttando solo le fasi ad essa antecedenti, tenendo anche conto della sovradisersione. Nello specifico, sono stati definiti i seguenti modelli di regressione assumendo un modello Beta-Binomiale:

- 1) variabile risposta gli  $n$  difettosi in Soaking e variabile esplicativa  $n$  difettosi in Rolling
- 2) variabile risposta gli  $n$  difettosi in Assembly e variabili esplicative il numero  $n$  di difettosi in Soaking e Rolling
- 3) variabile risposta gli  $n$  difettosi in Cleaning e variabili esplicative il numero  $n$  di difettosi in Assembly, Soaking, Rolling.

Formalmente il modello assume la seguente struttura:

$$Y_{ij} \sim BetaBin(\pi_j, e_j^\phi)$$

$$\pi_j = g^{-1}(\beta_0 + \sum_{p=1}^{j-1} \beta_p Y_{ip})$$

con  $j=2,3,4$ ;  $g(\cdot)$  funzione di legame *logit* e  $\phi_j$  parametri di dispersione.

Iniziando dal primo modello, denominato *m1* la stima mostra come il numero di difettosi della fase di Rolling risulti significativo nel prevedere la medesima quantità nella seconda fase. In particolare ad un aumento unitario in termini di difettosi per la prima fase corrisponde un aumento della probabilità di difettosità per la seconda fase pari al 5.9%.

Tabella 4: Stime dei coefficienti del predittore lineare per il modello M1

	Stima	Std. Error	t value	p-value
Intercetta	-4.278	0.034	-126.877	0
Rolling	0.058	0.010	5.657	0

Si è effettuata un'analisi dei residui allo scopo di verificare la bontà del modello.

Il pacchetto utilizzato per implementare il modello beta-binomiale sfrutta una particolare tipologia di residui chiamati residui-quantile (Dunn e Smyth, 1996) utile in condizione di sovradisersione. Questi residui sono, per risposte con distribuzione continua, definiti come

$$r_i = \Phi^{-1}(F(y_i; \hat{\theta})) \quad (1)$$

dove  $F(y_i; \theta)$  è la funzione di ripartizione della risposta e  $\Phi^{-1}(\cdot)$  è l'inversa della funzione di ripartizione della normale standard. Se il modello è correttamente specificato i residui avranno distribuzione approssimata normale standard. Per variabili risposta discrete, a tale definizione si aggiunge una perturbazione casuale.

Dall'analisi dei residui presente in Figura 1 relativo a *m1* si può notare come questo modello appaia correttamente specificato. Anche il test di Shapiro-Wilk conduce a non rifiutare l'ipotesi di normalità.

Proseguendo con il secondo modello si evince come il numero di difetti della fase Rolling risulti significativo nel prevedere il numero di difetti della fase di Assembly.

In particolare ad un aumento unitario in termini di difettosi per la prima fase corrisponde un aumento della probabilità di essere difettoso per la terza fase pari al 6.4%. Anche l'effetto della fase di Soaking risulta significativo, e in particolare presenta un effetto che comporta un aumento della probabilità di difettosità per la terza fase pari al 6.9%.

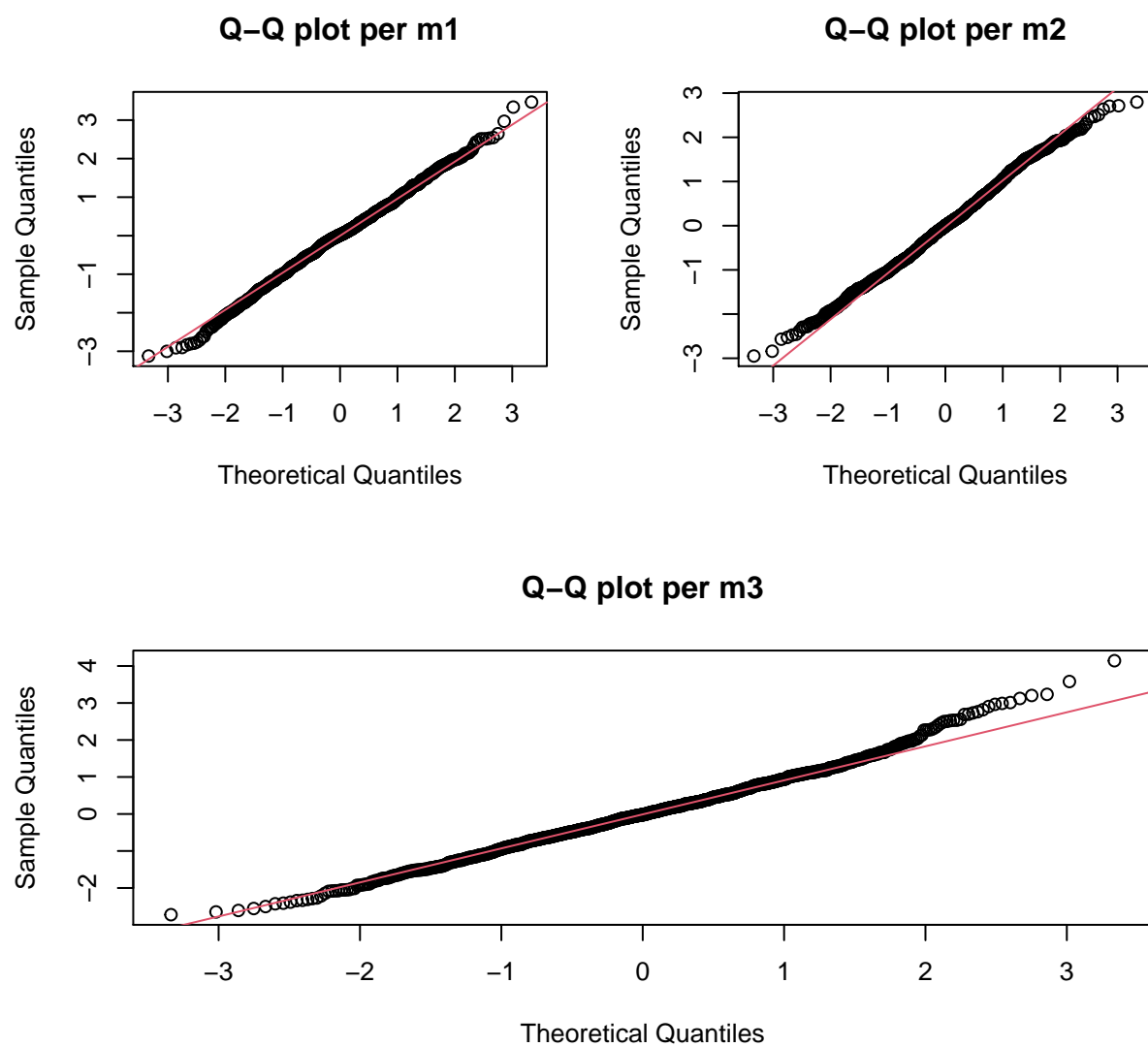


Figura 1: Q-Q plot dei 3 modelli beta-binomiali.

Tabella 5: Stime dei coefficienti del predittore lineare per il modello M2

	Stima	Std. Error	t value	p-value
Intercetta	-3.609	0.032	-112.069	0
Rolling	0.062	0.008	7.526	0
Soaking	0.067	0.007	9.918	0

Per quanto riguarda i residui del modello  $m2$ , sempre in riferimento alla Figura 1, il qqplot suggerisce un buon adattamento, confermato anche dal test di Shapiro-Wilk.

Infine, anche per il terzo modello (chiamato  $m3$ ) si ottiene che il numero di difettosi delle tre fasi precedenti sia significativo nel prevedere il numero di difettosi della fase di Assembly. Nello specifico, con la medesima interpretazione dei modelli precedenti, le fasi antecedenti hanno un effetto sulla probabilità di difettosità dell'ultima fase pari al 3.5%, 3.2% e 5.2% , in ordine dalla prima alla terza.

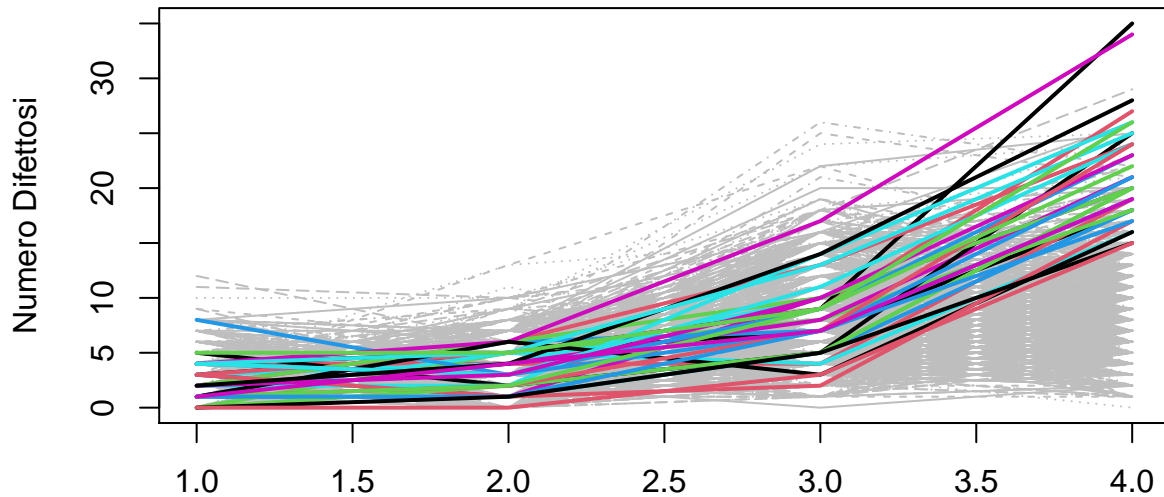
Tabella 6: Stime dei coefficienti del predittore lineare per il modello M3

	Stima	Std. Error	t value	p-value
Intercetta	-3.686	0.035	-106.590	0
Rolling	0.034	0.008	4.227	0
Soaking	0.031	0.007	4.626	0
Assembly	0.050	0.003	14.517	0

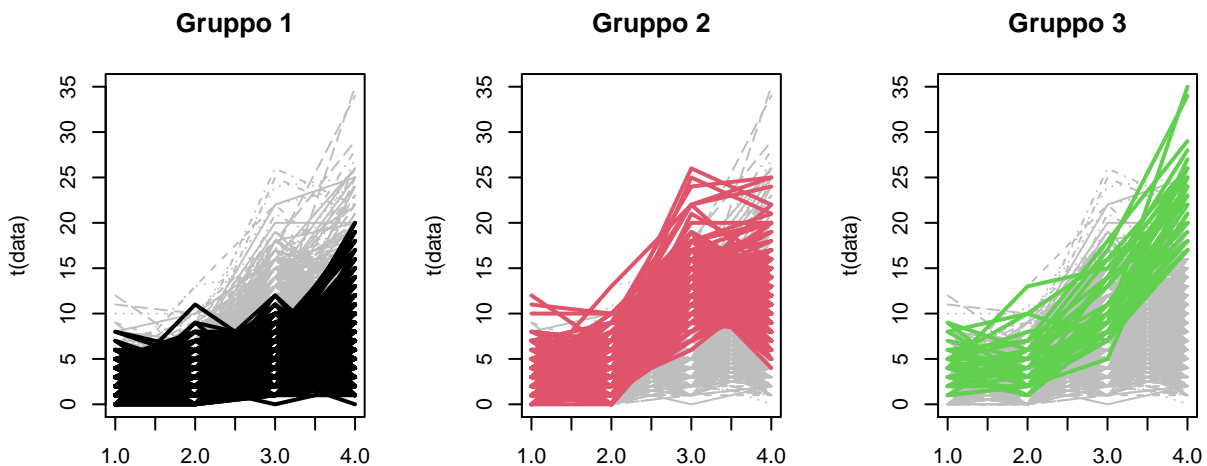
Tuttavia per questo modello, l'analisi dei residui riportati in Figura 1, mostra un allontanamento dalle ipotesi distributive, nello specifico si nota che la coda destra si discosta dai valori di riferimento della normale standard . In aggiunta il test di Shapiro-Wilk porta al rifiuto dell'ipotesi nulla.

Sia per il modello 2 che per il 3 si è valutato l'inserimento di interazioni significative tramite una procedura *stepwise* con criterio di penalizzazione AIC, tuttavia in entrambi i casi non sono stati apportati cambiamenti ai modelli. Per gli stessi modelli il termine relativo alla varianza è risultato significativo ad evidenza della necessità di passare dal modello binomiale al beta-binomiale.

Succesivamente si sono analizzate i cicli che il modello non riesce a prevedere in modo adeguato. Si riporta il loro andamento tra le 4 fasi nella seguente figura:



E' evidente un comportamento comune che si manifesta nel numero molto elevato di difetti nell'ultima fase rispetto alla terza. Questo suggerisce la presenza di sottopopolazioni di cicli caratterizzati da andamenti differenti probabilmente dovuti a fattori stocastici non misurati. Di conseguenza è sembrato adeguato effettuare un raggruppamento delle osservazioni tramite clustering agglomerativo da cui risultano tre gruppi distinti evidenti dalla seguente figura.



Il primo gruppo identificato si distingue per un andamento sostanzialmente stabile nelle prime tre fasi e da una dispersione nell'ultima fase. Nel secondo gruppo si nota un forte incremento della difettosità tra le fasi di Soaking e Assembly e da un assestamento in Cleaning. Infine il

terzo gruppo è caratterizzato da una crescita costante del numero di pezzi difettosi attraverso tutto il processo. Questo gruppo comprende gran parte dei cicli che non mostravano un buon adattamento. Uno sviluppo futuro può essere quello di implementare un modello che nella stima dei coefficienti di regressione tenga conto della diversa natura delle osservazioni.

## 4 Parte 3 - Difettosità finale

Un obiettivo secondario posto dagli ingegneri, ma importante, è quello di capire come sia possibile stimare la difettosità finale di un ciclo di produzione utilizzando la quarta ed ultima fase. Gli ingegneri, inoltre, si chiedevano se fosse più o meno opportuno usare solo il numero di difettosi dell'ultima fase o tutte 4 le misure. La strategia da noi adottata è stata quella di utilizzare tutte e 4 le fasi: dal momento che si è in presenza di un campionamento casuale, costruire la metrica su tutte le 4 misure ci permette di utilizzare tutta l'informazione a disposizione, e di conseguenza ottenere una stima della metrica di interesse più precisa.

### 4.1 Variabile latente

Per rispondere al quesito si è definita una modellazione basata su una variabile latente. Nello specifico, si assume che la distribuzione del numero di difettosi in ogni fase dipenda da una variabile latente (quindi non osservabile) comune per tutto il ciclo. Definiamo con  $Z_i$  la variabile latente e con  $Y_i = (Y_{ij})_{j=1,\dots,4}$  il numero di difettosi per il ciclo  $i$ -esimo e la fase  $j$ -esima. Il modello che ne consegue è il seguente:

$$\begin{aligned} Z_i &\sim N(0, 1) \\ Y_{ij}|Z_i = z &\sim \begin{cases} \text{Bin}(200, q_{jz}) & j = 1, 2 \\ \text{BetaBin}(200, q_{jz}, \phi_j) & j = 3, 4 \end{cases} \\ q_{jz} &= \text{logit}(\alpha_j + \beta z) \end{aligned}$$

- si assume l'indipendenza da ciclo a ciclo e l'indipendenza condizionata da fase a fase data la variabile latente;
- la relazione, definita dal coefficiente *beta*, della probabilità di difettosità con la variabile latente è la stessa per ogni fase, varia invece l'intercetta del predittore lineare. Secondo questa formulazione, ogni fase, avrà una probabilità di difettosità di base diversa (come dopotutto è auspicabile);
- per la fase di Rolling e quella di Soaking si assume una distribuzione binomiale, mentre, per quanto visto nelle analisi precedenti, sembra più adatta una distribuzione beta-binomiale per le 2 fasi finali.
- i parametri di dispersione  $\phi_j$  vengono fissato a priori per questioni computazionali e posti pari alle stime ottenute tramite il metodo dei momenti.

Per la stima del vettore  $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta)$  si è adottato un approccio di stima di massima verosimiglianza.

La distribuzione congiunta è:

$$P(Z_i, Y_i) = P(Y_{i1}|Z_i = z) \dots P(Y_{i4}|Z_i = z)P(Z_i = z)$$

Integrando  $Z_i$  si ottiene la distribuzione marginale:

$$P(Y_i) = \int P(Z_i, Y_i) dZ_i$$

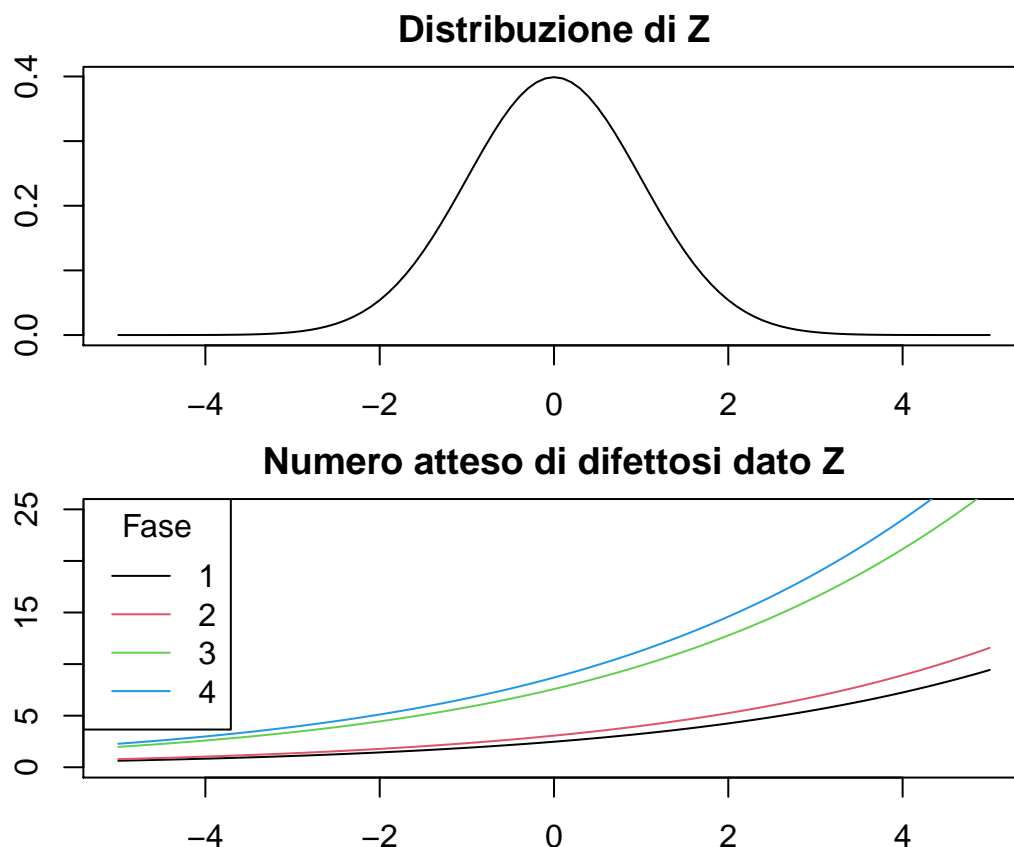
La funzione di log-verosimiglianza allora è:

$$l(\theta; Y) = \sum_{i=1}^{1180} \log(P(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}; \theta))$$

dalla cui massimizzazione si ottengono le stime di  $\theta$ . La massimizzazione può avvenire secondo una qualche procedura numerica e nello specifico è stata utilizzata la funzione `nlminb` di *R*.

## 4.2 Risultati

Nella figura seguente vengono riportati la distribuzione marginale di  $Z$  e il valore atteso del numero di difettosi. La figura ci conferma quanto rilevato dalle analisi precedenti: fissata la variabile latente, il numero atteso di difettosi aumenta da fase a fase con un salto netto tra la fase di Soaking e quella di Assembly.



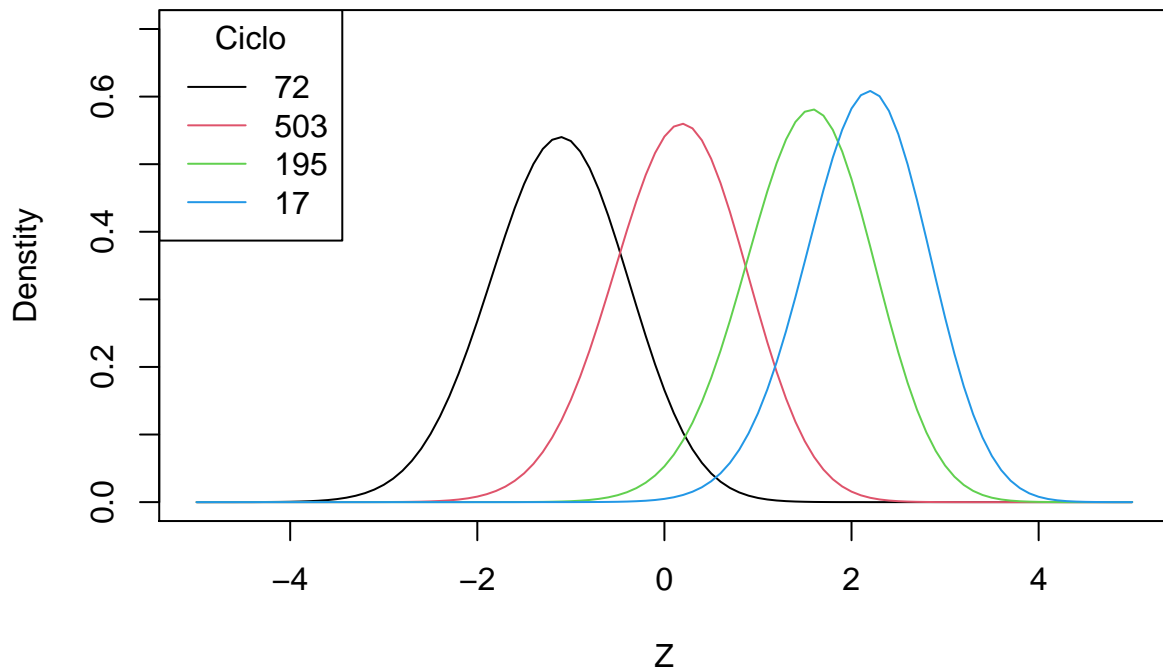
La componente di più interesse fornita da questa modellazione è quella riguardante la distribuzione a posteriori di  $Z_i$ . Si ottiene facilmente dall'equazione:

$$P(Z_i|Y_i) = \frac{P(Y_{i1}|Z_i = z) \dots P(Y_{i4}|Z_i = z)P(Z = z)}{P(Y_i)}$$

Osservando tale distribuzione, per ogni ciclo si ha un'indicazione della difettosità generale.

Nella figura successiva viene rappresentata la distribuzione a posteriori della variabile latente per alcuni cicli. Nello specifico, sono stati selezionati 4 cicli con caratteristiche diverse:

- il ciclo 72 presenta un numero di difettosi piccolo per tutte le fasi;
- il ciclo 503 presenta un andamento crescente;
- il ciclo 195 è quello con il numero di difettosi più elevato nell'ultima fase;
- il ciclo 17 presenta un numero molto elevato di difettosi nell'ultime due fasi.



Si può osservare come all'aumentare del numero di difettosi complessivo osservato la distribuzione si sposta verso destra, cioè verso valori di difettosità più elevati. Inoltre, la varianza tende a diminuire anche se in modo molto lieve. Da notare anche il fatto che il ciclo 17 abbia una distribuzione più a destra rispetto al ciclo 195 nonostante abbia un numero di difettosi per l'ultima fase molto minore. Questo deriva dal fatto che la distribuzione a posteriori, per costruzione, tiene conto di tutte e 4 le fasi e non solo dell'ultima come desiderato all'inizio di questa sezione.



Nella tabella seguente vengono riportati i 4 cicli selezionati con le rispettive mode e quantili della distribuzione di interesse. Come esempio, si riporta anche il ciclo 73 a cui viene attribuita una moda molto alta. Questo perché presenta sia nella fase di Rolling che in quella di Soaking un numero di difettosi molto elevato e poco consono al processo.

Ciclo	Rolling	Soaking	Assembly	Cleaning	Moda	25%	75%
72	3	1	2	2	-1.11	-1.64	-0.64
503	4	2	8	9	0.19	-0.33	0.64
195	4	5	9	35	1.58	1.08	2.00
17	6	5	25	21	2.19	1.71	2.60
73	10	10	19	14	2.65	2.21	3.02

### 4.3 Limiti

Il limite maggiore di questa modellazione è che assume la variabile latente costante per tutte le fasi. Di conseguenza, non coglie la dinamica mostrata dalle analisi precedenti per cui in alcuni cicli sembra esserci un cambio netto di difettosità dopo una o più fasi. Alcuni tentativi sono stati fatti in questa direzione ma non si è riuscito a formulare un algoritmo efficiente per la stima delle quantità di interesse.

## 5 Conclusioni

L'analisi effettuata ha consentito di confermare l'andamento monotono del rischio di difettosità, il quale infatti cresce ad ogni fase successiva nonostante l'eliminazione dei pezzi difettosi durante il processo, così come ipotizzato dagli ingegneri responsabili della produzione. Si è potuto inoltre modellare la struttura di dipendenza tra le varie fasi, la quale si manifesta in una correlazione positiva del numero di pezzi difettosi di ogni ciclo, non solo con la fase immediatamente precedente, ma con tutte le altre. Il processo si caratterizza inoltre per una crescente dispersione, in particolare nel passaggio tra le fasi di Assembly e Cleaning, che la modellistica implementata non è in grado di cogliere appieno sfruttando l'informazione sulla difettosità delle fasi precedenti. Un'analisi ad hoc tramite clustering, mette in risalto la presenza di un gruppo ben distinguibile che coincide in gran parte con le osservazioni che si discostano dal modello. Questo fornisce ulteriore evidenza all'ipotesi secondo cui esistono fattori stocastici ignoti che affliggono i cicli di produzione. Il modello a variabile latente ha consentito di fornire una misura di difettosità finale dei cicli di produzione che considera tutte e 4 le fasi. Anche questa procedura conferma una marcata diversificazione dei vari cicli.

## 6 Appendice A: Materiale Supplementare

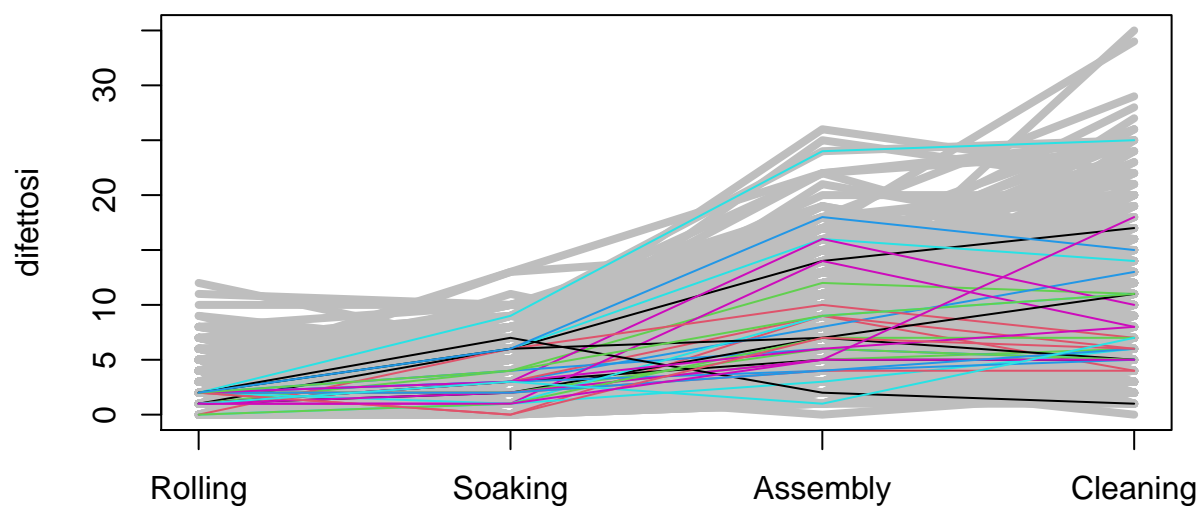
### 6.1 1 - Individuazione di andamenti

Per esplorare la presenza di comportamenti ricorrenti si sono rappresentati graficamente in totale 60 cicli di produzione, selezionati casualmente in base al numero di difettosi di una fase, e suddivisi nelle due seguenti situazioni:

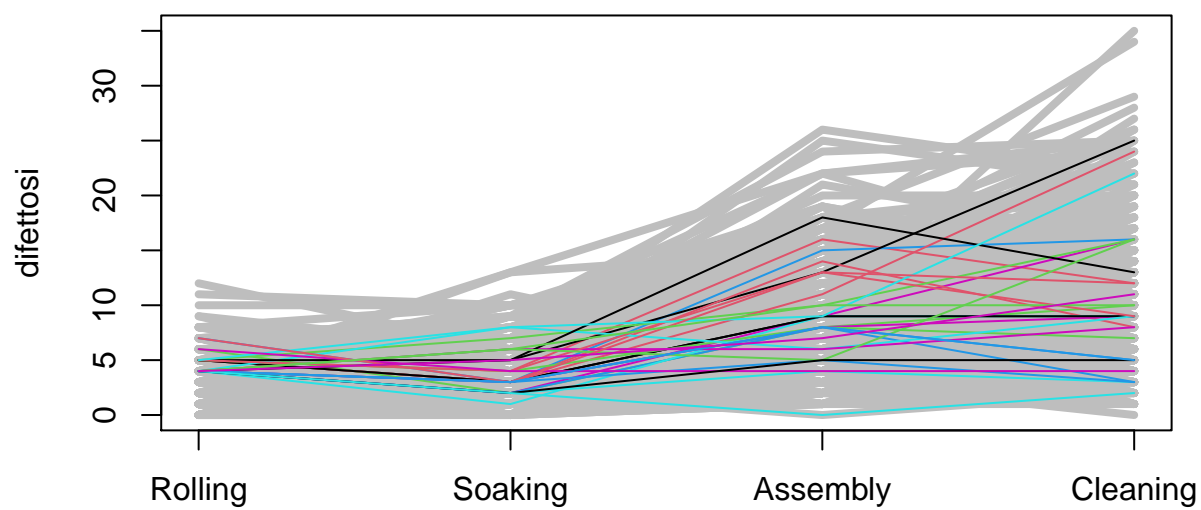
1. il numero di difettosi è compreso tra 0 e la mediana della fase
2. il numero di difettosi è compreso tra la mediana e il massimo della fase

Quindi per esempio prendendo la fase di *Soaking*, verranno estratti casualmente 60 cicli di produzione dove, in quella specifica fase, 30 avranno il numero di difettosi compreso tra 0 e la mediana e 30 con il numero di difettosi tra la mediana e il massimo.

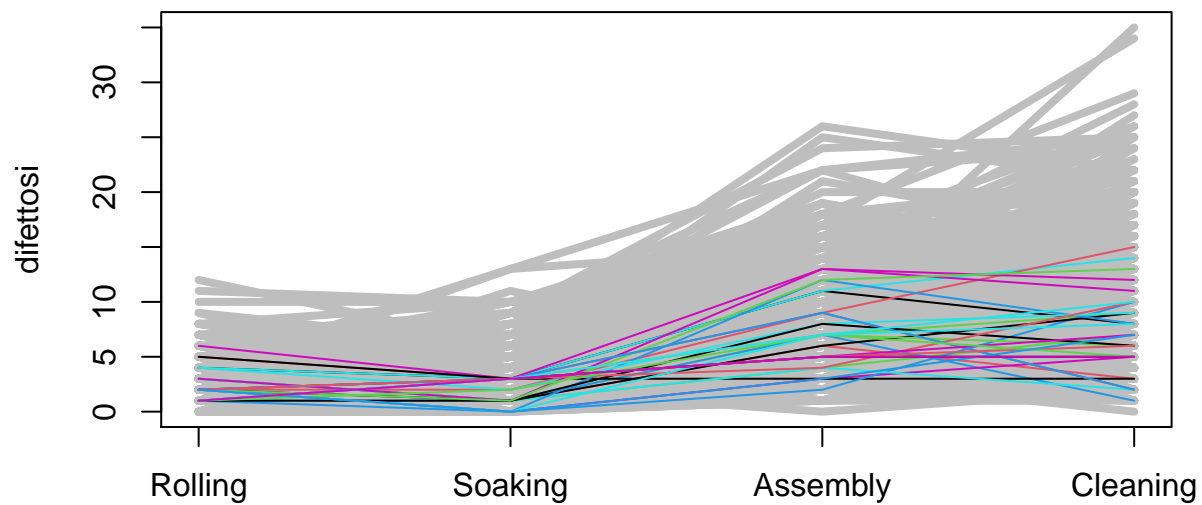
### Andamento di cicli casuali con difettosi bassi in ROLLING



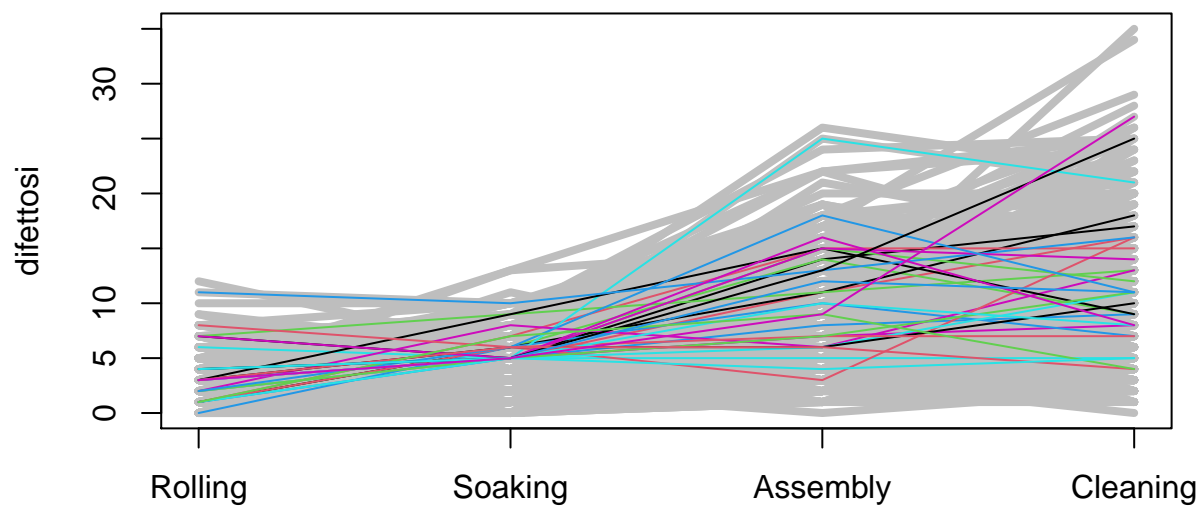
### Andamento di cicli casuali con difettosi alti in ROLLING



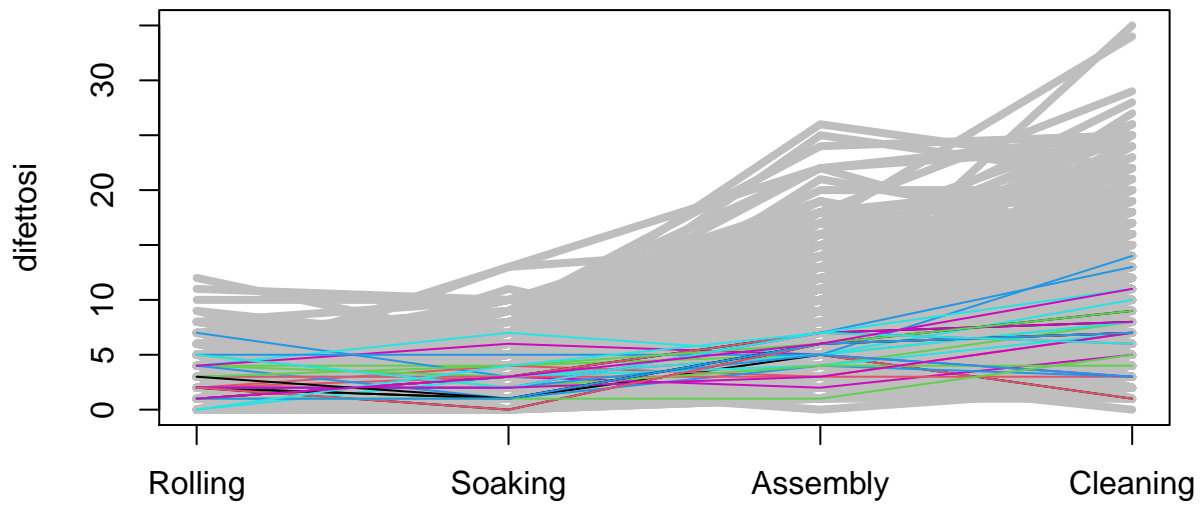
### Andamento di cicli casuali con difettosi alti in SOAKING



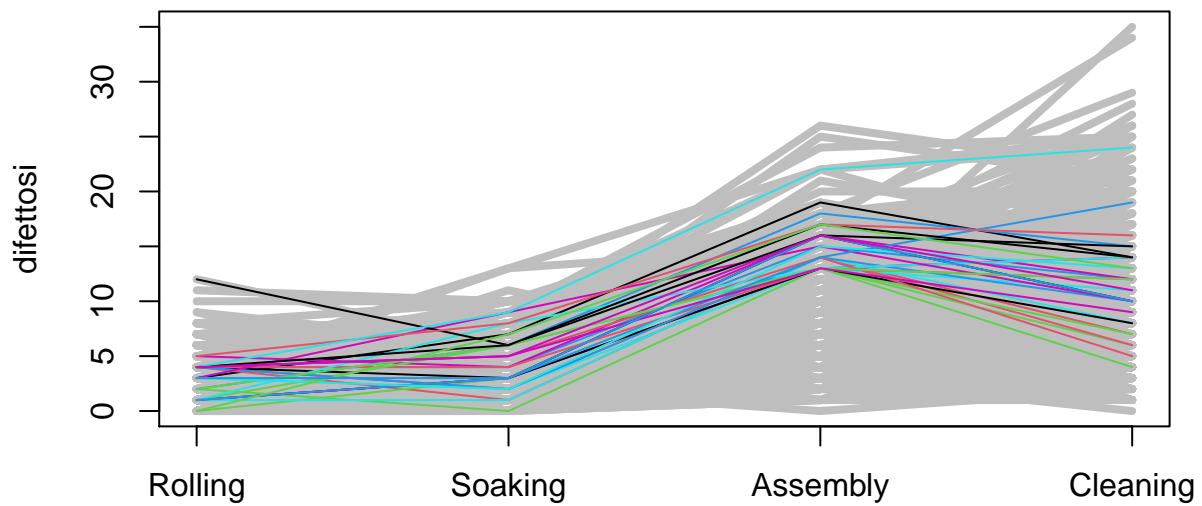
### Andamento di cicli casuali con difettosi alti in SOKING



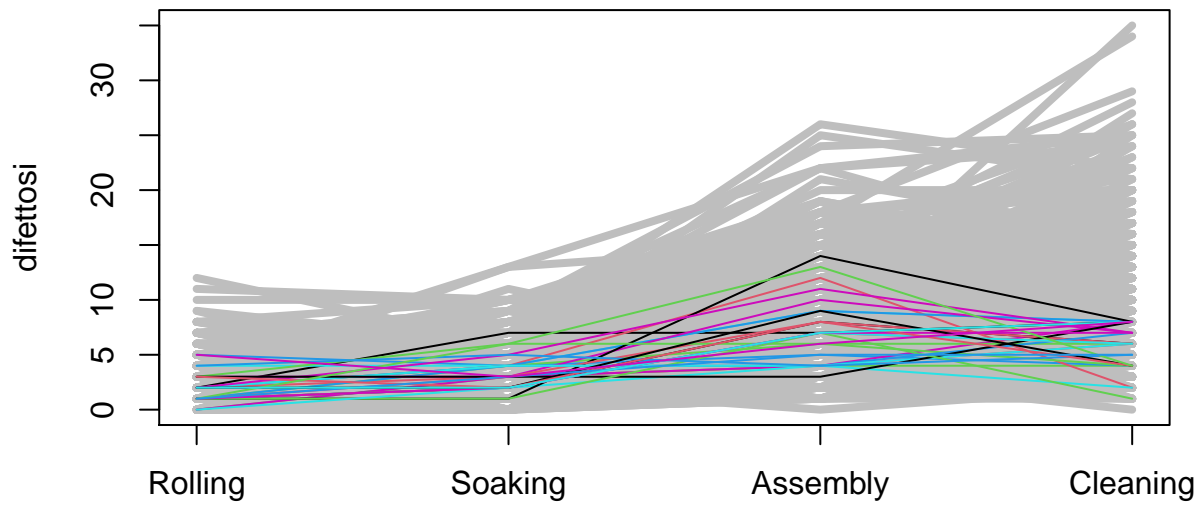
### Andamento di cicli casuali con difettosi bassi in ASSEMBLY



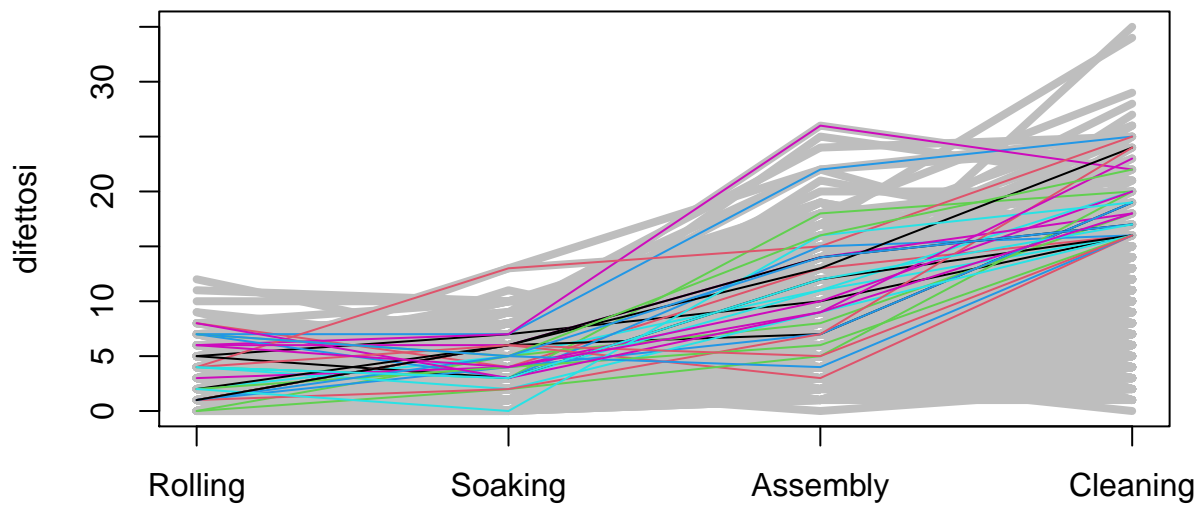
### Andamento di cicli casuali con difettosi alti in ASSEMBLY



### Andamento di cicli casuali con difettosi alti in CLEANING



### Andamento di cicli casuali con difettosi alti in CLEANING



## 6.2 2 - Summary del modello ANOVA

##

## Call:

```
## vglm(formula = cbind(y, 200 - y) ~ Fase, family = betabinomial(zero = 0),
##      data = long.data)
##
## Coefficients:
##              Estimate Std. Error  z value Pr(>|z|)
## (Intercept):1 -3.21821    0.01549 -207.729 < 2e-16 ***
## (Intercept):2 -5.19604    0.08143  -63.810 < 2e-16 ***
## FaseCleaning:1  0.13720    0.02212   6.203 5.55e-10 ***
## FaseCleaning:2  0.29498    0.10848   2.719 0.00654 **
## FaseRolling:1  -1.12213    0.02522 -44.495 < 2e-16 ***
## FaseRolling:2  -1.77365    0.28088  -6.315 2.71e-10 ***
## FaseSoaking:1  -0.90622    0.02443 -37.092 < 2e-16 ***
## FaseSoaking:2  -1.23777    0.19417  -6.375 1.83e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(mu), logitlink(rho)
##
## Log-likelihood: -11286.65 on 9432 degrees of freedom
##
## Number of Fisher scoring iterations: 7
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1', '(Intercept):2', 'FaseRolling:2'
```

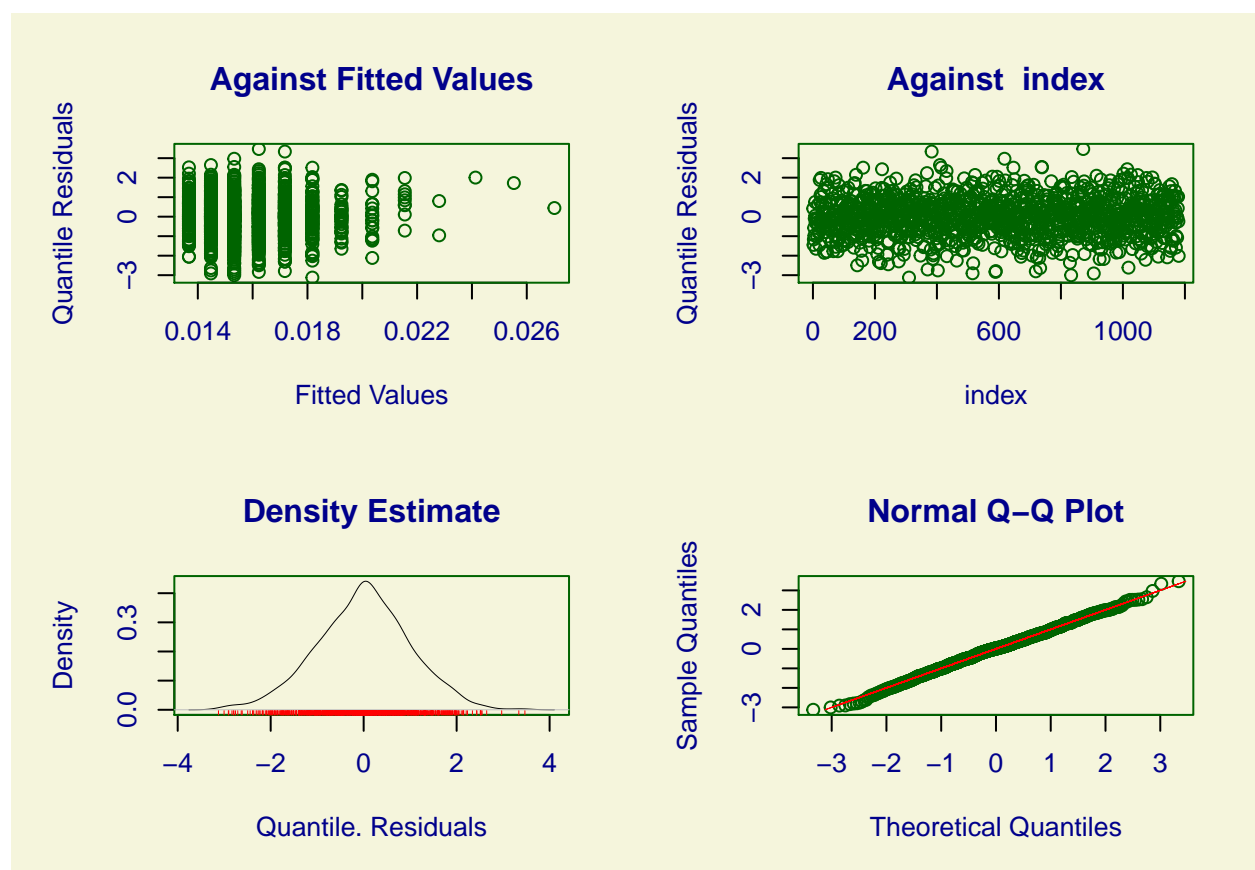
## 6.3 3 - Summary & Grafici dei Modelli Beta-binomiali

### 6.3.1 Modello *m1*

```
## *****
## Family:  c("BB", "Beta Binomial")
##
## Call:  gamlss(formula = cbind(Soaking, 200 - Soaking) ~ Rolling,
##      sigma.formula = ~1, family = BB, data = data, trace = FALSE)
##
## Fitting method: RS()
##
## -----
## Mu link function:  logit
## Mu Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -4.27832    0.03372 -126.877 < 2e-16 ***
## Rolling      0.05784    0.01022   5.657 1.93e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
##
## -----
## Sigma link function:  log
## Sigma Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.5717      0.1953  -33.64  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## No. of observations in the fit:  1180
## Degrees of Freedom for the fit:   3
##      Residual Deg. of Freedom: 1177
##                               at cycle: 10
##
## Global Deviance:      4842.548
##           AIC:        4848.548
##           SBC:        4863.768
## *****
```



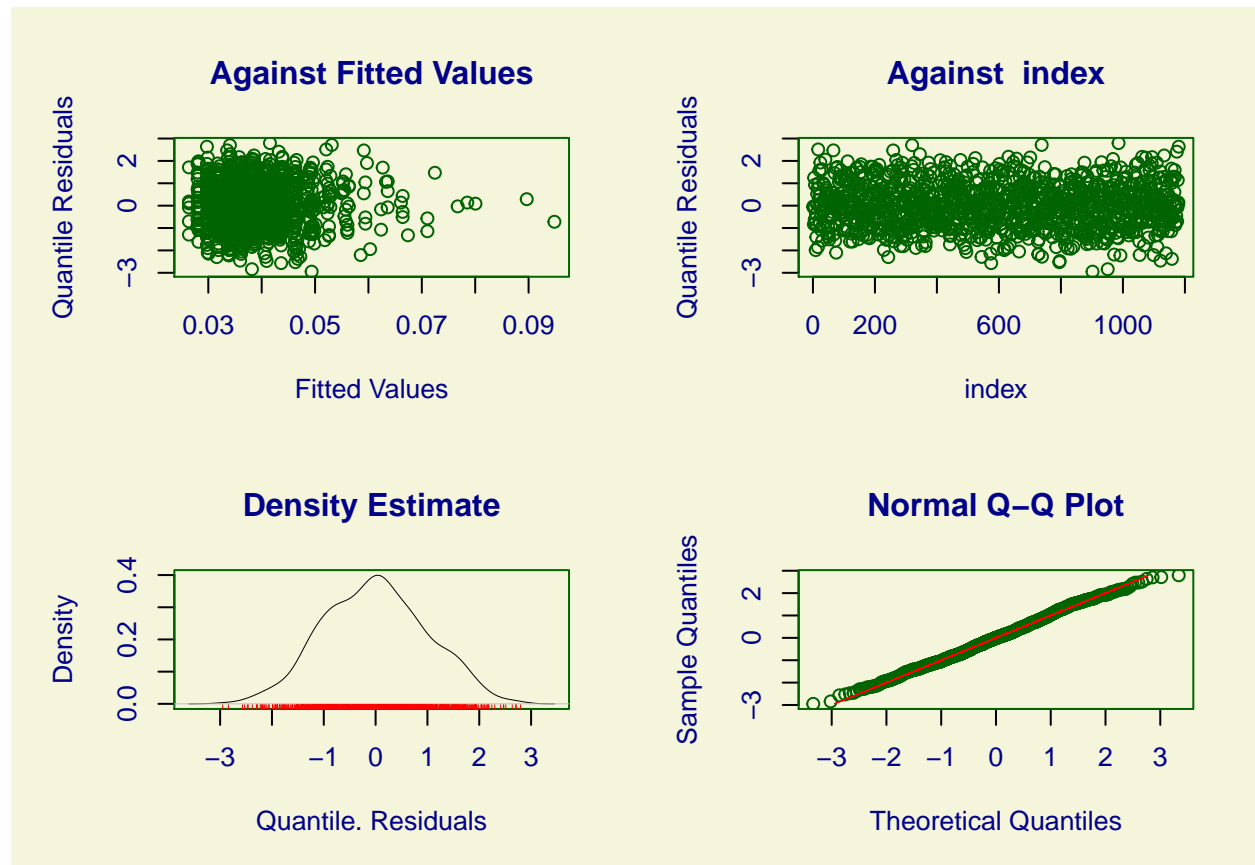
```
## *****
## Summary of the Randomised Quantile Residuals
```

```
##                      mean    = -0.003139984
##                      variance  = 1.006396
##                      coef. of skewness = -0.06831473
##                      coef. of kurtosis  = 3.163151
## Filliben correlation coefficient = 0.9991112
## *****
##
##  Shapiro-Wilk normality test
##
## data:  resid(m1.2)
## W = 0.99814, p-value = 0.2182
```

### 6.3.2 Modello *m2*

```
## *****
## Family:  c("BB", "Beta Binomial")
##
## Call:  gamlss(formula = cbind(Assembly, 200 - Assembly) ~
##      Rolling + Soaking, sigma.formula = ~Rolling, family = BB,      data = data)
##
## Fitting method: RS()
##
## -----
## Mu link function:  logit
## Mu Coefficients:
##      Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -3.608617   0.032200 -112.069 < 2e-16 ***
## Rolling      0.061906   0.008226   7.526 1.04e-13 ***
## Soaking      0.067170   0.006772   9.918 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Sigma link function:  log
## Sigma Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.84317   0.17595 -33.209 <2e-16 ***
## Rolling      0.11675   0.05052  2.311  0.021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## No. of observations in the fit:  1180
```

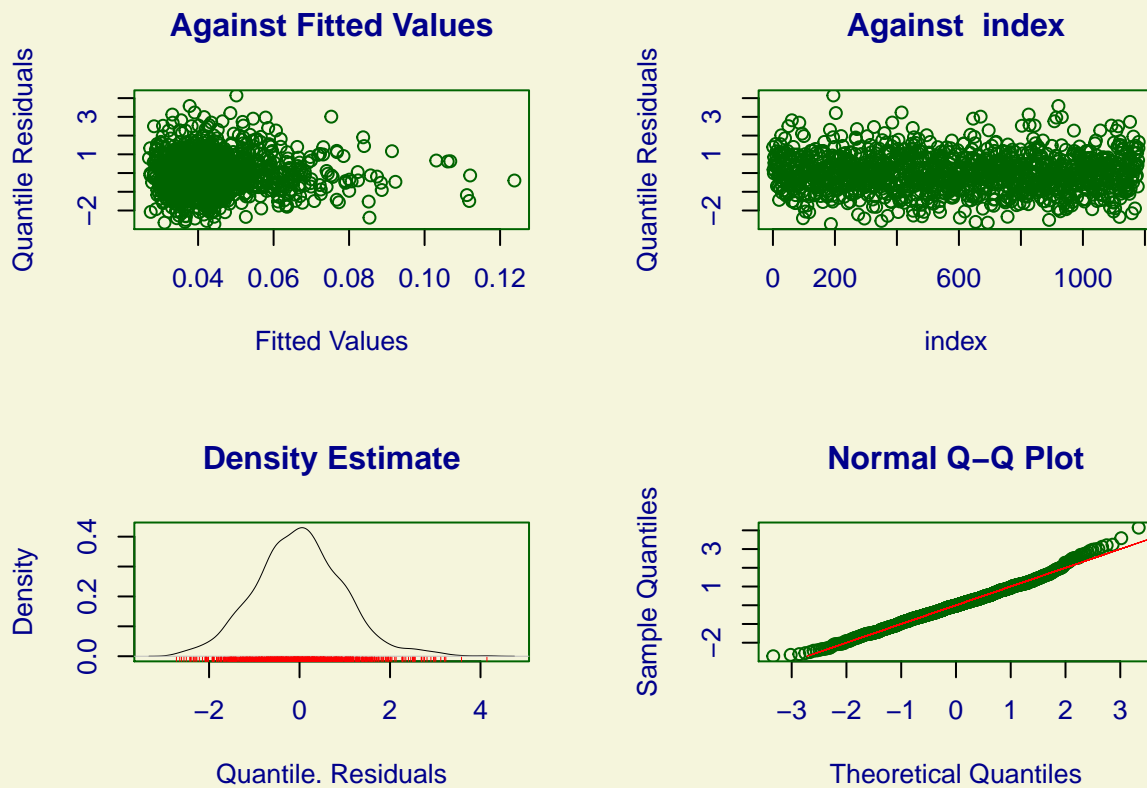
```
## Degrees of Freedom for the fit: 5
##      Residual Deg. of Freedom: 1175
##                               at cycle: 9
##
## Global Deviance:      6271.226
##           AIC:        6281.226
##           SBC:        6306.593
## *****
```



```
## *****
##      Summary of the Randomised Quantile Residuals
##                               mean    = -6.820729e-05
##                               variance = 0.989123
##                               coef. of skewness = 0.06343127
##                               coef. of kurtosis = 2.666096
## Filliben correlation coefficient = 0.9986754
## *****
##
##      Shapiro-Wilk normality test
##
## data:  resid(m2.2)
## W = 0.9971, p-value = 0.03012
```

### 6.3.3 Modello $m_3$

```
## *****
## Family:  c("BB", "Beta Binomial")
##
## Call:  gamlss(formula = cbind(Cleaning, 200 - Cleaning) ~
##      Rolling + Soaking + Assembly, sigma.formula = ~Rolling,
##      family = BB, data = data)
##
## Fitting method: RS()
##
## -----
## Mu link function:  logit
## Mu Coefficients:
##      Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -3.685954   0.034581 -106.590 < 2e-16 ***
## Rolling      0.034042   0.008053   4.227 2.55e-05 ***
## Soaking      0.031298   0.006766   4.626 4.15e-06 ***
## Assembly     0.050219   0.003459  14.517 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## Sigma link function:  log
## Sigma Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.72435    0.16626 -34.429 <2e-16 ***
## Rolling      0.10153    0.04824   2.105  0.0355 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
## No. of observations in the fit:  1180
## Degrees of Freedom for the fit:  6
##      Residual Deg. of Freedom:  1174
##      at cycle:  8
##
## Global Deviance:      6439.203
##      AIC:      6451.203
##      SBC:      6481.643
## *****
```



```
## *****
## Summary of the Randomised Quantile Residuals
##               mean    = 0.004107116
##               variance = 0.9737252
##               coef. of skewness = 0.2686725
##               coef. of kurtosis = 3.546523
## Filliben correlation coefficient = 0.9966977
## *****

##
## Shapiro-Wilk normality test
##
## data: resid(m3.2)
## W = 0.99339, p-value = 4.185e-05
```

## 7 Appendice B: Metodo dei momenti

Nel seguito vengono riportate alcune proprietà della distribuzione Beta\_binomiale utilizzata nei diversi modelli come distribuzione marginale del numero di difettosi nelle fase. Sia  $X \sim BetaBin(n, \alpha, \beta)$ :

- la funzione di probabilità è:

$$f_X(x|n, \alpha, \beta) = \binom{n}{x} \frac{B(x + \alpha, n - x + \beta)}{B(\alpha + \beta)}$$

con  $n \in N_0$ ,  $\alpha > 0$  e  $\beta > 0$ .

- i primi due momenti sono rispettivamente:

$$E[X] = n \frac{\alpha}{\alpha + \beta} \quad E[X^2] = \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

- una stima per  $\alpha$  e  $\beta$  si può ottenere con il metodo dei momenti:

$$\hat{\alpha} = \frac{nm_1 - m_2}{n(\frac{m_2}{m_1} - m_1 + 1) + m_1} \quad \hat{\beta} = \frac{(n - m_1)(n - \frac{m_2}{m_1})}{n(\frac{m_2}{m_1} - m_1 + 1) + m_1}$$

dove  $m_1$  e  $m_2$  sono rispettivamente il primo e il secondo momento campionario.

- il parametro di dispersione si può ottenere come:

$$\hat{\rho} = \frac{1}{1 + \hat{\alpha} + \hat{\beta}}$$

- una distribuzione Beta Binomiale ha una varianza pari a  $1 + \hat{\rho}(n - 1)$  volte quella di una distribuzione binomiale.