

Analisi *real-time* dei trend di difettosità nei sistemi di avvitatura semi-automatica

Statistica & Data Science

Leonardo Livi

11 Aprile, 2024



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Table of Contents

1 Background ed obiettivi

► Background ed obiettivi

► Dati e metodi

► Risultati

Pipeline di processazione del dato

Dashboard interattiva

Modellazione statistica degli errori

► Conclusioni

Introduzione

1 Background ed obiettivi

Nel contesto industriale contemporaneo, la necessità di ottimizzare i processi produttivi e garantire la qualità dei prodotti è fondamentale per la competitività delle aziende. In questo scenario, la raccolta e l'analisi dei dati rappresentano un elemento cruciale per identificare aree di miglioramento e ottimizzazione nel campo della mobilità elettrica.





Introduzione /2

1 Background ed obiettivi

Nel flusso produttivo, l'azienda coinvolta nell'elaborato dispone di una quantità dati importante: dalla raccolta dati per quanto riguarda le istruzioni di montaggio fino alle misure dei collaudi. Attraverso delle centraline presenti nei vari banchi di lavoro, esistono dei sistemi che vanno a verificare quante volte, con che tempi ed in che modo avvengono le avvitature delle viti all'interno dei prodotti.

La necessità principale è sempre stata quella di costruire una base dati sulla quale verificare la correttezza delle informazioni e monitorare l'andamento nel tempo delle operazioni.



Procedure di sviluppo

1 Background ed obiettivi

Sviluppo algoritmo



Obiettivi:

- Estrazione dati dai file di log
- Modellazione struttura dati

Controllo statistico di processo



Obiettivi:

- Monitoraggio dei tempi medi di avvitatura
- Individuazione Seriali 'out of control'

Modellazione statistica



Obiettivi:

- Inferenza sui parametri che governano i processi
- Previsione delle numerosità dei difetti

Table of Contents

2 Dati e metodi

► Background ed obiettivi

► **Dati e metodi**

► Risultati

Pipeline di processazione del dato

Dashboard interattiva

Modellazione statistica degli errori

► Conclusioni

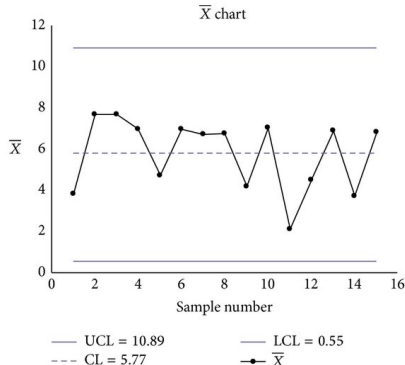
I file di *log* sono dei file di testo che vengono generati automaticamente nel momento in cui si conclude la fase finale di completamento del prodotto ed al loro interno vengono visualizzate tutte le procedure eseguite. Nel nostro caso di studio l'interesse si rivolge solamente in quelle istruzioni che contengono la stringa 'W - Ps:', un particolare formato utilizzato per esprimere che in quell'operazione è avvenuta un'avvitatura in una specifica componente.

```
14/02/2023 12:55:20 ; 24 W - Ps: 040-3 n°/Batch: 01/08 R: 0 Tq: 000037-0 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:14 ; 24 W - Ps: 040-3 n°/Batch: 01/08 R: 1 Tq: 000074-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:17 ; 24 W - Ps: 040-3 n°/Batch: 02/08 R: 1 Tq: 000071-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:20 ; 24 W - Ps: 040-3 n°/Batch: 03/08 R: 0 Tq: 000081-2 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:21 ; 24 W - Ps: 040-3 n°/Batch: 03/08 R: 1 Tq: 000070-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:23 ; 24 W - Ps: 040-3 n°/Batch: 04/08 R: 1 Tq: 000079-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:26 ; 24 W - Ps: 040-3 n°/Batch: 05/08 R: 1 Tq: 000077-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:28 ; 24 W - Ps: 040-3 n°/Batch: 06/08 R: 0 Tq: 000082-2 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:30 ; 24 W - Ps: 040-3 n°/Batch: 06/08 R: 1 Tq: 000072-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:32 ; 24 W - Ps: 040-3 n°/Batch: 07/08 R: 1 Tq: 000072-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
14/02/2023 12:56:33 ; 24 W - Ps: 040-1 n°/Batch: 08/08 R: 1 Tq: 000078-1 Ag: 00000-1 Rd: 00000-1 Tool: A3610740
```

Control Chart

2 Dati e metodi

Il Controllo Statistico del Processo (SPC) è un insieme di strumenti statistici e di gestione che, attraverso la sorveglianza della distribuzione delle caratteristiche di qualità, permettono ai processi di funzionare in modo più prevedibile e più redditizio.



Formule

$$CL = \bar{\bar{X}}$$

$$\text{dove } \bar{\bar{X}} = \sum_{i=1}^k \bar{x}_i$$

$$UCL_{\bar{x}} = \bar{x} + 3 \frac{\sigma}{\sqrt{n}}$$

$$LCL_{\bar{x}} = \bar{x} - 3 \frac{\sigma}{\sqrt{n}}$$

Modelli per dati di conteggio

2 Dati e metodi

Tipo	Distribuzione	Metodo	Descrizione
GLM	Poisson	ML quasi adjusted	Poisson regression Quasi-Poisson regression Adjusted Poisson regression
	NB	ML	Negative Binomial regression
zero-augmented	Poisson	ML	zero-inflated Poisson (ZIP), hurdle Poisson
	NB	ML	zero-inflated NB (ZINB), hurdle NB

Modellazione statistica degli errori

2 Dati e metodi

L'obiettivo è modellare il numero di errori commessi durante i processi di lavorazione, dove questi vengono identificati come il numero di avvitature non andate a buon fine all'interno di una procedura univoca, secondo le covariate disponibili per le varie operazioni:

- $\beta_1 \Rightarrow$ *Fase* : indicatore della fase di lavoro
- $\beta_2 \Rightarrow$ *Batch* : numero di avvitature attese
- $\beta_3 \Rightarrow$ *Duration* : durata totale della procedura (in secondi)
- $\beta_4 \Rightarrow$ *Derive* : indicatore di procedura in deriva (T/F)

Sovradispersione nei dati di conteggio

2 Dati e metodi

Si verifica il fenomeno della sovradispersione quando la vera varianza condizionata $V(Y | X_1, \dots, X_k)$ è più grande della varianza condizionata basata sul modello di regressione.

⇒ una soluzione a questo problema è l'utilizzo di una distribuzione che rompa la relazione deterministica che c'è tra media e varianza e che quest'ultima sia maggiore della media, ovvero la distribuzione **Negative Binomial**.

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

Stessa equazione specificata nel caso Poisson ma con un vincolo in più:

$$V(Y_i) = \mu_i + \frac{\mu_i^2}{\alpha}$$

dove α è il parametro di sovradispersione

Eccesso di zeri

2 Dati e metodi

Modellare dati di conteggio richiede particolare attenzione riguardo ai conteggi zero ($Y = 0$), spesso il modello ha un cattivo adattamento solo perché le frequenze osservate e previste per $Y = 0$ sono marcatamente differenti. Per ovviare a questo problema vi sono:

- i modelli **Hurdle** che combinano una componente di conteggio troncata a sinistra con una componente di ostacolo censurata a destra
- i modelli **Zero-inflated** che adottano un approccio un po' diverso: sono modelli misti che combinano una componente di conteggio e un punto massa in zero

Modelli per eccesso di zeri

2 Dati e metodi

Hurdle:

$$\begin{cases} \text{logit}(\pi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} & \text{modello binario per la prob. di avere un conteggio positivo} \\ \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} & \text{componente di conteggio troncata a sinistra (in } \gamma = 1) \end{cases}$$

Zero-inflated:

$$\begin{cases} \text{logit}(\pi_i) = \mathbf{z}_i^T \boldsymbol{\gamma} & \text{modello binario per la prob. di appartenere al gruppo 'Sempre zero'} \\ \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} & \text{componente di conteggio per il gruppo 'Non sempre zero'} \end{cases}$$

Table of Contents

3 Risultati

► Background ed obiettivi

► Dati e metodi

► **Risultati**

Pipeline di processazione del dato

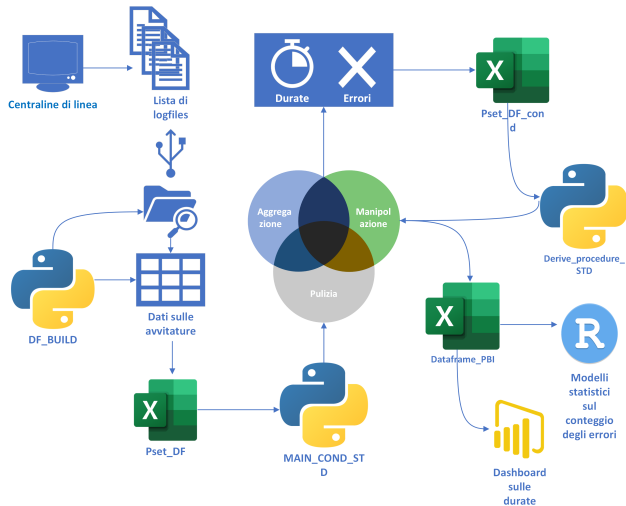
Dashboard interattiva

Modellazione statistica degli errori

► Conclusioni

Pipeline di processazione del dato

3 Risultati

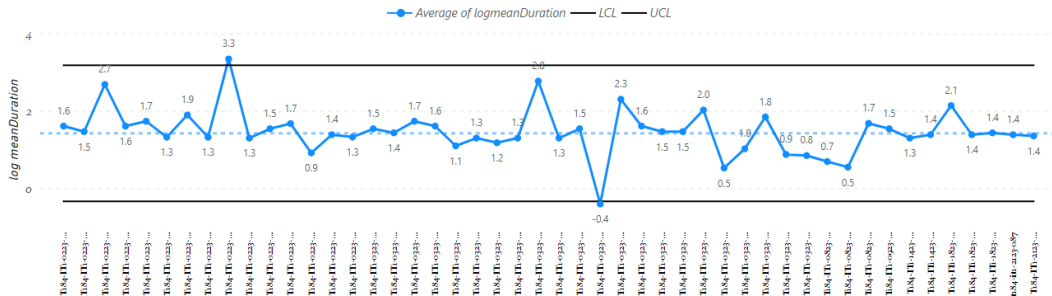


Dashboard interattiva

3 Risultati

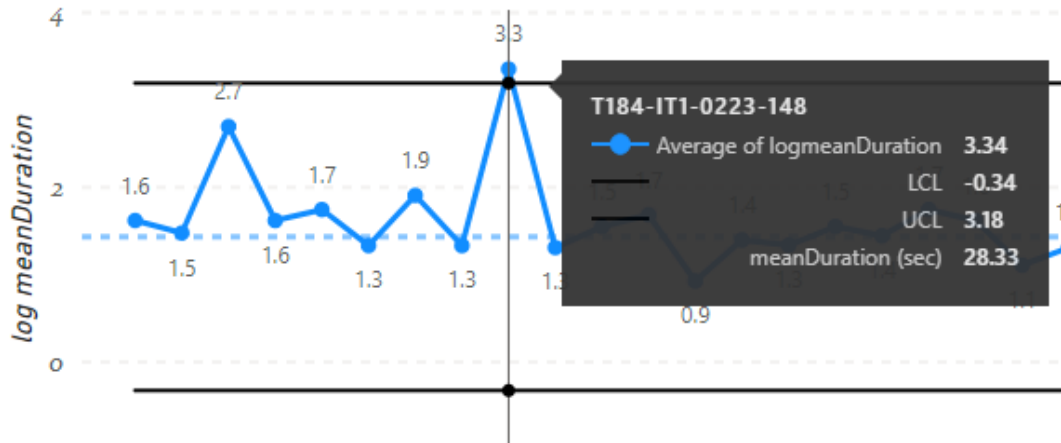
Serial Series

XBar Chart



Seriale fuori controllo statistico

3 Risultati



Informazioni aggiuntive

3 Risultati

Global_ID	SerialNumber	Tool_ID	Operatore	Pset_ID	SerialDuration
6AGC084099	T184-IT1-0223-042	A3511732	F1333	61	5.00
6AGC084099	T184-IT1-0223-063	A3610740	F1199	61	4.33
6AGC084099	T184-IT1-0223-074	A3511732	35001	61	14.67
6AGC084099	T184-IT1-0223-110	A3610740	F1309	61	5.00
6AGC084099	T184-IT1-0223-121	A3511732	35001	61	5.67
6AGC084099	T184-IT1-0223-124	A3610740	F1309	61	3.75
6AGC084099	T184-IT1-0223-126	A3511732	35001	61	6.67
6AGC084099	T184-IT1-0223-141	A3610740	F1369	61	3.75
6AGC084099	T184-IT1-0223-148	A3511732	F1199	61	28.33
6AGC084099	T184-IT1-0223-151	A3610740	F1199 F1371	61	3.67
6AGC084099	T184-IT1-0223-155	A3610740	F1309	61	4.67
6AGC084099	T184-IT1-0223-174	A3610740	F1199	61	5.33
6AGC084099	T184-IT1-0223-184	A3511732	F1369	61	2.50

L'implementazione di questo tipo di infografica in azienda e quindi nelle linee di produzione, ha portato diversi benefici:

- **Monitoraggio in tempo reale**
- **Miglioramento continuo**
- **Comunicazione efficace**

Risultati regressione Poisson

3 Risultati

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.1771	0.0250	7.08	0.0000***
FASE2	0.0039	0.0292	0.13	0.8943
FASE3	-0.1167	0.0303	-3.85	0.0001***
FASE4	-1.4869	0.1002	-14.84	0.0000***
Batch	0.0383	0.0035	10.91	0.0000***
Duration	0.0054	0.0001	37.45	0.0000***
Derive	1.5481	0.1849	8.37	0.0000***

Categorie di riferimento: Fase = *FASE1*, Derive = *False*

Esempio: $\exp(\beta_{Batch}) = \exp(0.0383) = 1.039$

Al netto della fase, della durata, dell'essere un osservazione in deriva o no, all'aumentare del numero di viti da avvitare il rischio di commettere un errore cresce del 3.9%

Risultati regressione Negative Binomial

3 Risultati

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.0764	0.0494	1.546	0.122***
FASE2	-0.0088	0.06	-0.148	0.882
FASE3	-0.0880	0.0589	-1.495	0.135
FASE4	-1.2084	0.138	-8.753	0.0000***
Batch	0.006	0.08	0.771	0.440
Duration	0.0112	0.0005	21.626	0.0000***
Derive	1.747	0.2945	5.936	0.0000***

Parametro di dispersione $\hat{\alpha} = 0.5766$, $SE(\hat{\alpha}) = 0.0215$

Risultati regressione Hurdle e Zero-inflated

3 Risultati

	Hurdle models		Zero-inflated models	
	Truncated Negbin	Binomial	Negbin	Binomial
(Intercept)	-0.014	-0.139 *	0.328***	0.131
FASE2	0.130 .	-0.154 .	0.127 *	0.603 *
FASE3	0.160 *	-0.446***	-0.079	-0.342
FASE4	-0.711 **	-1.435***	-0.851***	1.466 *
Batch	0.020 *	-0.041 **	-0.003	0.302***
Duration	0.006***	0.022***	0.008***	-0.368 **
Derive	2.677***	1.528***	1.765***	18.561

Per il modello hurdle, la componente binomiale descrive la probabilità di osservare un conteggio positivo mentre, per il modello ZINB, la componente binomiale predice la probabilità di osservare un conteggio nullo dalla componente di massa del punto 0

Valutazione performance dei modelli

3 Risultati

	ML-Pois	NB	Hurdle-NB	ZINB
n. parametri	7	8	15	15
logLik	-9512	-7168	-7073	-7045
AIC	19037.98	14351.2	14175.76	14119.67
BIC	19082.78	14402.41	14271.76	14215.67
$\sum_i \hat{f}_i(0)$	1124	2230	2232	2257

Il numero di zeri osservato nei dati è 2232 in un campione di 4449 processi di avvitatura

Table of Contents

4 Conclusioni

- ▶ Background ed obiettivi
- ▶ Dati e metodi
- ▶ Risultati
 - Pipeline di processazione del dato
 - Dashboard interattiva
 - Modellazione statistica degli errori
- ▶ Conclusioni



Conclusioni

4 Conclusioni

- La comparazione dei diversi modelli utilizzati per l'analisi dei dati di difettosità nei sistemi di avvitatura semi-automatica ha fornito risultati interessanti e utili per il miglioramento della qualità del processo produttivo. Alla luce delle valutazioni basate sugli indici di AIC e BIC, nonché sulla verosimiglianza e sulla capacità predittiva di zeri nei dati, emerge chiaramente che il modello *hurdle* risulta essere il più adatto per modellare efficacemente i dati raccolti.
- Il modello *zero-inflated*, rispetto al modello *hurdle* e agli altri modelli considerati, ha dimostrato di avere un AIC e BIC inferiori, indicando una migliore aderenza ai dati e una maggiore parsimonia. Inoltre, la verosimiglianza del modello *zero-inflated* è risultata significativamente più alta rispetto agli altri modelli, suggerendo che questo modello sia in grado di catturare più accuratamente la struttura sottostante dei dati di difettosità nel caso dei processi di avvitatura.



Conclusioni

4 Conclusioni

- Nonostante questo, il modello *hurdle* mantiene un'interpretazione maggiormente comprensibile e gradevole alla luce del contesto aziendale e viste le differenze quasi nulle di adattabilità secondo i criteri di informazione, si può presumibilmente adottare quest'ultimo modello per analizzare la tendenza degli errori al variare delle caratteristiche inserite nel modello.
- In conclusione si può affermare con fiducia che il modello *hurdle* rappresenta la scelta ottimale per l'analisi dei trend di difettosità nei sistemi di avvitatura semi-automatica. La sua capacità di adattarsi ai dati, di massimizzare la verosimiglianza e di predire con precisione la presenza di zeri, insieme alla chiarezza e all'efficacia della dashboard interattiva, lo rende uno strumento potente e affidabile per migliorare la qualità del processo produttivo, per identificare tempestivamente eventuali anomalie o difetti e per conoscere quali sono le componenti del processo che maggiormente influenzano il tasso di errore.

Analisi *real-time* dei trend di difettosità nei sistemi di avvitatura semi-automatica

Thank you for listening!
Any questions?