



UNIVERSITA' POLITECNICA DELLE MARCHE

FACOLTA' DI INGEGNERIA

Corso di Laurea Triennale in Ingegneria Biomedica

Dipartimento di Ingegneria Industriale e Scienze Matematiche

**VALUTAZIONE DELL'ATTIVITA' INDOOR GIORNALIERA SU UTENTI
ANZIANI CON SENSORI NON INVASIVI**

**EVALUATION OF DAILY INDOOR ACTIVITY ON ELDERLY USERS
WITH NOT INVASIVE SENSORS**

Relatore:

Prof. Lorenzo Scalise

Tesi di Laurea di:

Lucana Paredes Birkie Alexandra

Correlatore:

Ph.D. Casaccia Sara

A.A. 2019 /2020

Indice

1	Introduzione:	2
1.1	Smart Homes per persone anziane	3
2	Progetto eWare	5
2.1	Programma Active and Assisted Living (AAL).....	5
2.2	Ecosistema eWare	7
3	Obbiettivo della tesi	11
4	Materiali e metodi:	12
4.1	Matlab: Organizzazione del dataset	12
4.2	Analisi statistiche.....	14
4.3	Analisi di normalità e anormalità.....	18
4.4	Machine Learning	23
4.5	Studio del significato dei cluster risultanti dall'applicazione dell'apprendimento automatico non supervisionato.	27
5	Risultati	29
5.1	Analisi statistiche.....	29
5.1.1	Analisi su tutto il periodo di campionamento.	29
5.1.2	Analisi su base mensile	31
5.1.3	Analisi su base giornaliera	32
5.1.4	Analisi sul periodo di osservazione diviso in fasce orarie	40
5.2	Studio dei casi di normalità e anormalità	41
5.3	Machine learning unsupervised	43
5.4	Studio preliminare del significato dei cluster e verifica.....	59
6	Conclusioni e sviluppi futuri	62
	Bibliografia	65

1 Introduzione:

I progressi nella scienza e nella tecnologia medica hanno avuto come conseguenza l'aumento dell'aspettativa di vita a livello globale negli ultimi decenni [1]. La popolazione di 60 anni o più cresce più velocemente di tutte le altre fasce di età. Con il rapido invecchiamento della popolazione mondiale destinato a continuare correlato al calo della mortalità degli anziani, è stato stimato che entro il 2050 tutte le regioni del mondo tranne l'Africa avranno il 25% o più della loro popolazione a partire dai 60 anni. Diversi studi specifici per paese riportano alte probabilità di disabilità nella popolazione anziana [2].

In tutto il mondo la demenza senile è una delle principali cause di disabilità e dipendenza. Per sopperire a limitazioni nelle attività quotidiane delle persone affette da questa patologia è necessaria maggiore assistenza che implica un carico di stress sia per le persone affette da questa patologia che per la famiglia [3]. Le facoltà mentali generalmente colpite dalla demenza senile sono il pensiero, la memoria (a breve e lungo termine), l'orientamento spazio-temporale, il linguaggio, la comprensione, il calcolo, l'apprendimento, la lingua e il giudizio [4].

Le Smart Homes svolgono un ruolo fondamentale per rivoluzionare il sistema sanitario per la popolazione anziana fornendo un sostanziale supporto [1].

1.1 Smart Homes per persone anziane

La Smart Home è una casa intelligente che sfrutta il concetto di Internet Of Things, quindi un impianto integrato domotico la cui unità centrale si occupa della gestione e del controllo di un determinato numero di dispositivi periferici. Si tratta di un sistema che sfrutta tecnologie moderne come sensori ambientali e medici, attuatori, processori di calcolo e piattaforme di comunicazione wireless, per offrire servizi di assistenza sanitaria, sicurezza e benessere. Di seguito una figura che illustra a tutti gli effetti una casa intelligente.

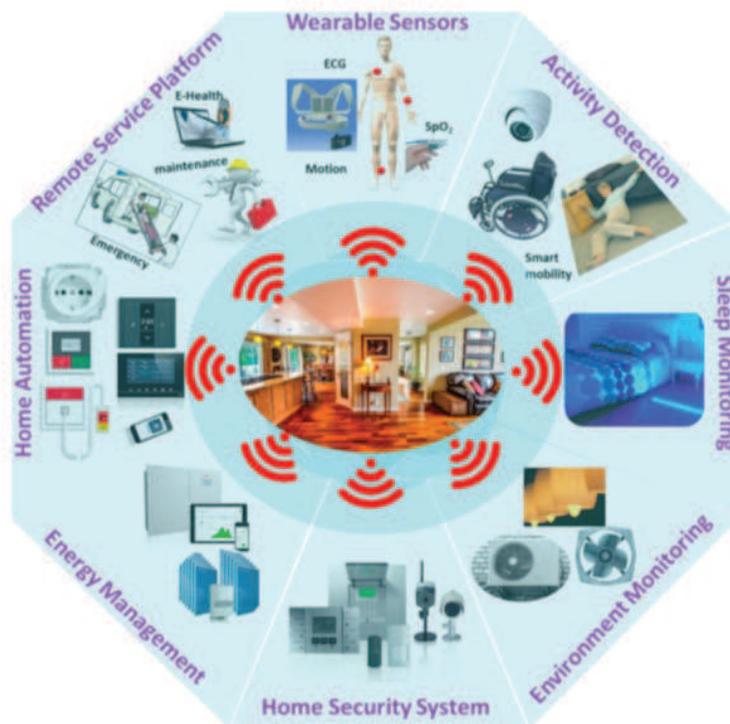


Figura 1: Smart home integrata con sistemi automatizzati per l'assistenza sanitaria agli anziani.

Uno dei principali obiettivi delle case intelligenti è quello di consentire alle persone anziane di vivere in maniera indipendente e attiva nel loro ambiente domestico e ricevere servizi di assistenza continui, non invasivi e senza interruzioni. Come conseguenza si ottiene una riduzione al minimo della frequenza delle visite o della durata in centri sanitari. Inoltre, monitorando e controllando il locale domestico è possibile ricavare modelli comportamentali di vita quotidiana dei soggetti anziani [1]. Occorre evidenziare che a partire dall'indagine dei modelli di comportamento, ricavati dall'utilizzo di sensori installati nelle abitazioni di utenti anziani, in mancanza di processi di interpretazione si ottengono informazioni che non forniscono risultati rilevanti. Risulta quindi di fondamentale importanza effettuare l'analisi dei dati per ottenere suggerimenti significativi e procedere con lo sviluppo di nuove prestazioni di assistenza destinate al cittadino.

2 Progetto eWare

2.1 Programma Active and Assisted Living (AAL)

AAL è un programma che finanzia l'innovazione tecnologica promosso dall'Unione Europea, che ha come obiettivo prolungare il tempo in cui gli anziani possono vivere all'interno del loro ambiente domestico in maniera indipendente attraverso l'impiego di tecnologie intelligenti e servizi di assistenza da remoto.

L'Università Politecnica delle Marche (UNIVPM) ha preso parte al progetto eWare (“Early Warning Accompanies Robotics Excellence”), progetto avviato nel 2017 e concluso a Maggio 2020 [5].



Figura 2: Manifesto progetto eWare.

Il progetto eWARE si occupa dello sviluppo di un servizio utile e significativo per utenti anziani. I principali insiemi target in eWare sono costituiti dalle persone affette da demenza senile dalla fase lieve a media che vivono da sole, e dai loro assistenti formali e informali. Gli utenti finali primari, come le municipalità, le organizzazioni di assistenza e i loro rappresentanti saranno in contatto con gli utenti finali secondari, che sono le organizzazioni di assistenza. Gli utenti finali terziari comprendono organizzazioni per l'assistenza domiciliare, compagnie di assicurazione e governi regionali e nazionali. Le nazioni coinvolte in questo progetto sono: Paesi Bassi, Italia, Svizzera e Norvegia.

Di seguito vengono elencati i principali obiettivi del progetto eWare:

- Il miglioramento della qualità della vita degli assistenti informali e delle persone con demenza.
- La riduzione dello stress soggettivo della comunità dei pazienti e degli assistenti formali e informali (caregiver).
- L'aumento della qualità dell'assistenza professionale [5].

2.2 Ecosistema eWare

L'ecosistema eWare è il risultato della combinazione di due tecniche eHealth:

- Monitoraggio delle attività di vita quotidiana (ADL) dell'utente a casa, attraverso una rete di sensori.
- Piccolo robot sociale che interagisce con l'anziano affetto da demenza attraverso suggerimenti rilevanti per il contesto al fine di raggiungere determinati obiettivi di "Activity of Daily Living" (ADL).

Rendere possibile l'interazione tra persone con demenza e robot sociali per fornire suggerimenti e feedback sui loro obiettivi quotidiani non solo promuove miglioramenti della qualità della vita di persone affette con demenza ma favorisce la riduzione del carico di stress comunemente affrontato dai caregiver.

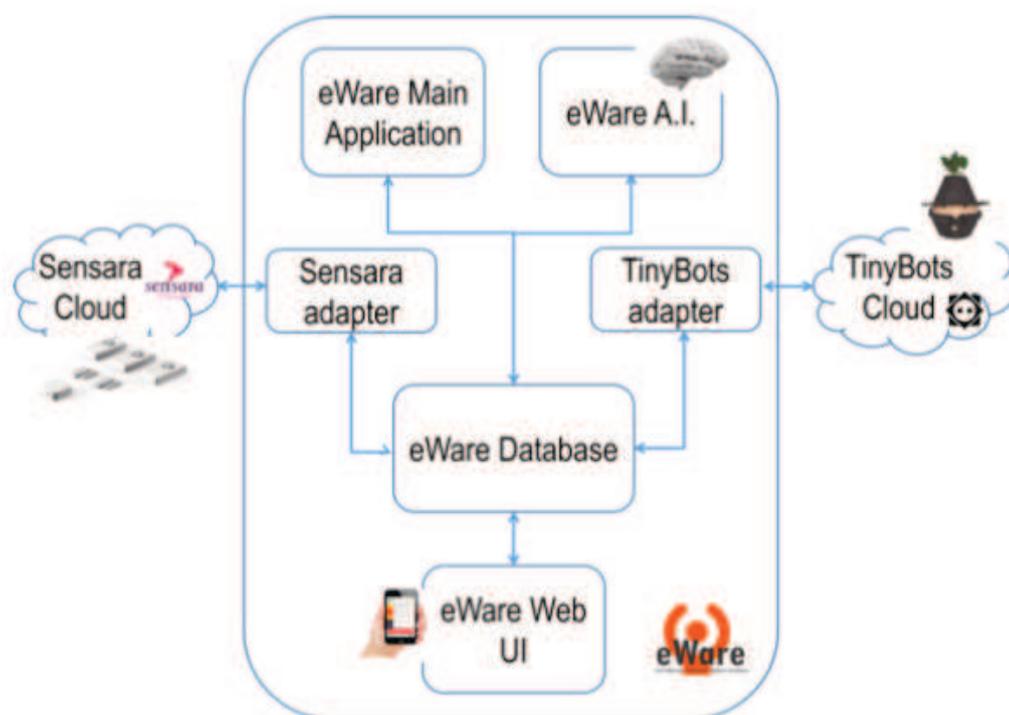


Figura 3: Architettura dell'eco-sistema eWare.

Il Sensara Lifestyle System è una nuova generazione di servizio Internet of Things (IoT) costituita da una rete di 5 sensori wireless che funzionano a batteria:

- 3 sensori a infrarossi passivi PIR, fissati a 1,40 m di altezza, installati in servizi igienici, cucina, ingresso e soggiorno.
- 2 sensori di contatto porta (sensori ON/OFF) installati in cucina, nel cassetto di uso frequente o sulla porta del frigorifero, ecc.

Un sensore ad infrarossi passivo (PIR sensor) rileva la radiazione infrarossa (IR) emessa da corpi e oggetti contenuti all'interno del suo campo visivo, il suo funzionamento non prevede l'emissione di nessuna forma di energia.

Questo è uno dei sensori più utilizzati in quanto fornisce osservazioni economiche e non invasive, oltre che efficienti dal punto di vista energetico.

Le misurazioni ottenute tramite i PIR risultano essere qualitative in quanto registrano solo lo stato di attivazione booleano dei sensori. Le valutazioni effettuate non rilevano tutte le informazioni sull'oggetto in osservazione, pertanto questa tipologia di sensori può essere utilizzata in ambienti sensibili alla privacy [6].

I sensori comunicano con il software nel "Sensara Cloud" utilizzando un ricevitore wireless con un raggio di connessione sufficiente a coprire l'appartamento sorvegliato. Questo ricevitore richiede una fonte di alimentazione cablata e una connessione Internet.

Il Sensara Lifestyle System provvede al monitoraggio delle attività della vita quotidiana delle persone anziane che vivono da sole a casa, garantendo il rispetto della loro privacy. I dati vengono raccolti dai sensori e caricati "nel cloud", motore di analisi che dopo circa due settimane dispone di informazioni sufficienti per riconoscere i modelli di vita dell'utente in casa: ora del risveglio dell'anziano in casa, per quanto tempo si usa il bagno, la frequenza dell'uso della cucina, la durata temporale delle uscite di casa, tempi di riposo, ecc.

Ricavati i modelli comportamentali possono essere rilevate e analizzate le eccezioni e anomalie, permettendo all'algoritmo di analisi dati di Sensara l'invio di una notifica allo smartphone di un familiare.

Ai fini del progetto eWare viene utilizzata Tessa, un robot sociale sviluppato da Tinybots. Attraverso Tessa è possibile fornire stimoli per gli utenti affetti da demenza senile parlando, trasmettendo suggerimenti, promemoria e riproduzione musicale. Queste funzioni favoriscono l'allontanamento della persona con demenza da uno stato passivo. L'intelligenza artificiale (AI) dei Tinybots impara nel tempo ed è in grado di supportare le esigenze specifiche degli utenti [5][7].

3 Obiettivo della tesi

Lo scopo della tesi è quello di avvalersi di analisi statistiche, studio di casi di normalità e anormalità e apprendimento automatico non supervisionato in modo da ottenere metodi utili per l'individuazione di anomalie nei modelli comportamentali degli utenti affetti da demenza senile. In particolare, vengono considerati 3 utenti, relativi alle residenze 102, 232 e 239.

Le attività quotidiane o "Activities of Daily Living" (ADLs) prese in considerazione in questo lavoro di tesi sono specificamente:

- Toileting: Visita al bagno.
- Eating: Attività di cucina (come colazione, pranzo o cena)
- Entering leaving: Apertura della porta di casa senza conseguente uscita del residente
- Inside: Il residente o un'altra persona è entrato in casa utilizzando una delle porte esterne con un sensore porta su di essa.
- Outside: Il residente o un'altra persona è uscito utilizzando una delle porte esterne e questa era l'ultima persona a casa in quel momento.
- Sleeping asleep: Il residente è andato a letto o è inattivo durante le normali ore notturne / serali.
- Sleeping awake: Il residente si è alzato dal letto o è diventato attivo durante le normali ore notturne / mattutine.

4 Materiali e metodi:

Questo capitolo presenta una descrizione delle analisi effettuate singolarmente su ognuna delle tre residenze prese in considerazione durante questo lavoro di tesi.

Vengono quindi delineati i metodi utilizzati per analizzare la variabilità delle attività degli utenti.

4.1 Matlab: Organizzazione del dataset

Il dataset cronologico di partenza per questa tesi è stato precedentemente ricavato dall'elaborazione dei segnali grezzi trasmessi al sistema dai sensori installati in ogni residenza.

I dati in esso racchiusi riguardano le attività quotidiane o “Activities of Daily Living” (ADLs) svolte dagli utenti, le componenti del dataset rappresentano le attivazioni delle attività, specificandone minuti, ora e data di registrazione nel sistema.

Il primo passo è stato quello di creare un dataset da utilizzare per le analisi sulla piattaforma di calcolo Matlab. Quindi i dati iniziali, vettori di tipo stringa, sono stati organizzati sotto forma di una matrice iniziale in cui le righe corrispondono alle attivazioni delle attività e le colonne sono:

- 1 vettore le cui componenti contengono il numero associato alla residenza.
- 1 vettore contenente il tipo di ADL svolta.
- 1 vettore di tempo (“datetime”) che individuano anno, mese, giorno, ora, minuti e secondi.

Quindi, data la richiesta di concentrarsi su tre precisi utenti, dalla matrice iniziale sono state individuate ed estrapolate 3 “sottomatrici residenze” ottenute in seguito alla selezione delle righe relative alle residenze 102, 232 e 239.

In seguito, sono stati individuati ed eliminati i dati fonte di rumore in quanto relativi ai giorni in cui sono stati riscontrate delle problematiche, dovute alle prove per l’installazione del sistema o a motivi tecnici anomali.

4.2 Analisi statistiche

Per realizzare un metodo di valutazione della variabilità delle attività sono state eseguite analisi statistiche prendendo in considerazione il periodo di campionamento prima interamente e poi gradualmente suddiviso in: mesi, giorni e infine sono state considerate 4 particolari fasce orarie.

Inizialmente per mettere a confronto i rilevamenti complessivi di ogni attività sono state calcolate le percentuali delle attivazioni delle attività sull'intero periodo di campionamento.

Sistematicamente per ogni attività sono state selezionate le attivazioni e, attraverso operazioni iterative, ricavata la percentuale rispetto al numero complessivo di attivazioni registrate in tutto il periodo di analisi. Per intero periodo di campionamento si intende l'intervallo di tempo tra la prima e l'ultima attivazione quindi numero di giorni di osservazione che è un parametro diverso per ogni residenza.

Inizio e fine dei periodi di osservazione sono diversi per ogni residenza variano per le 3 residenze:

- 102: da dicembre 2018 ad aprile 2020.
- 232: da ottobre 2019 ad aprile 2020.
- 239: da novembre 2019 ad aprile 2020.

I risultati ottenuti dal calcolo delle percentuali sono stati collezionati in un vettore e, per facilitarne la lettura, sono stati realizzati:

- Grafico a torta, attraverso la funzione matlab “pie3”, in cui le sezioni del grafico rappresentano le percentuali delle attività, la percentuale maggiore è stata opportunamente evidenziata attraverso il vettore “explode”. La legenda favorisce la comprensione associando a ciascun colore la percentuale di ogni attività.
- Istogramma, attraverso la funzione matlab “hist”, in cui lungo l’asse delle ascisse troviamo le attività e lungo quello delle ordinate le percentuali, scritte opportunamente all’interno di ogni rettangolo che costituisce l’istogramma.

Procedendo con lo studio della variabilità mensile il periodo di analisi è stato suddiviso in mesi, per ognuno dei quali sono stati iterativamente ricavati i valori percentuali corrispondenti alle attivazioni di ogni attività rispetto al numero di attivazioni globali rilevate durante quel mese.

I risultati ottenuti sono stati ordinati in modo da poter visualizzare per tutte le attività e per tutti i mesi di riferimento la variabilità di quell’attività. È stata quindi creata una tabella le cui righe indicano i mesi e le colonne le attività.

Per approfondire progressivamente lo studio della variabilità delle attività si è deciso di procedere cercando di mettere in evidenza per ogni attività le variazioni che avvengono quotidianamente durante le 24 ore che compongono un giorno. A questo scopo sono state create delle “Matrici orarie”.

Queste matrici sono composte da 24 colonne, una per ogni ora del giorno, e un numero di righe pari al numero di giorni di osservazione.

Per la loro costruzione è stato ricavato per ogni ora di registrazione il numero di occorrenze delle attivazioni ordinatamente per ogni attività, sfruttando il vettore colonna di tempo, che ad ogni attivazione associa l’istante preciso della sua registrazione. È stato ottenuto un numero di “Matrici orarie” pari al numero delle attività prese in considerazione.

Per rendere facilmente interpretabili le variazioni che quotidianamente avvengono sono state create delle immagini attraverso la funzione “imagesc” a partire dalle matrici orarie.

Per ogni matrice oraria ogni elemento, che contiene il numero di ripetizioni delle attivazioni, specifica il colore per un pixel della corrispondente immagine.

L’immagine risultante si presenta come una griglia di pixel m per n dove m è il numero di righe quindi 24 ed n è il numero delle colonne quindi dei giorni di osservazione, la gamma di colori utilizzata è quella presente nella mappa dei colori [8].

Successivamente seguendo il procedimento descritto in precedenza si è considerato il periodo di osservazione suddiviso in mesi quindi sono state costruite delle “Matrici orarie mensili” ognuna delle quali è stata opportunamente visualizzata attraverso un’immagine che permette di osservare le variazioni che avvengono quotidianamente all’interno di ogni mese.

Infine, la variabilità delle attività è stata studiata suddividendo le ore del giorno in 4 fasce orarie:

- Notte: dalle 00:00 alle 6:00.
- Mattina: dalle 7:00 alle 12:00.
- Pomeriggio: dalle 13:00 alle 18:00.
- Sera: dalle 19:00 alle 24:00.

Partendo dalla suddivisione del periodo di osservazione nelle fasce descritte in precedenza, per ognuna di esse è stato calcolato il valore percentuale di ogni attività all’interno di ogni mese.

I risultati ottenuti da queste operazioni sono stati opportunamente collezionati in tabelle denominate con i mesi che compongono il periodo di osservazione.

4.3 Analisi di normalità e anormalità

Il metodo di individuazione dei giorni normali e anormali è stato ottenuto stabilendo un determinato intervallo di normalità sulla base di strumenti statistici quali media e scarto interquartile.

La media, in inglese “Median”, si calcola sommando tutti i valori a disposizione e dividendo il risultato per il numero complessivo dei dati. È stato preferibile fare riferimento alla media anziché alla mediana poiché è stato osservato che per il dataset a disposizione, ampio, il valore di media si avvicina a quello di mediana.

Lo scarto interquartile o differenza interquartile, in inglese “Interquartile range” o IQR, è la differenza tra terzo e primo quartile in altre parole l’ampiezza della fascia di valori che contiene la metà “centrale” dei valori osservati. Si tratta di un indice di dispersione quindi di una misura di quanto si allontanano i valori da un determinato valore centrale.

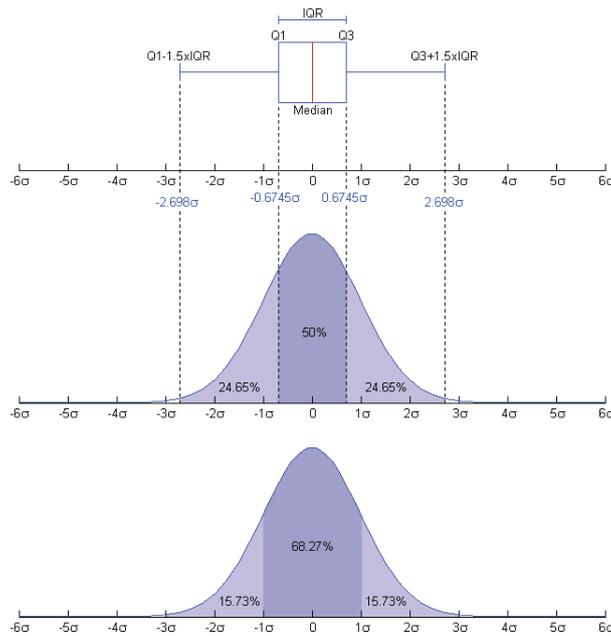


Figura 4: Concetto di media e scarto interquartile.

Le operazioni che di seguito verranno descritte sono state eseguite separatamente per ogni residenza.

Per ognuna delle attività considerate sono state individuate le attivazioni giornaliere e sulla base di queste sono stati creati vettori le cui componenti contengono i numeri delle attivazioni giornaliere. Per ogni vettore è stato isolato il primo mese di osservazione del quale sono stati calcolati media e IQR, valori utilizzati per la definizione di due intervalli di riferimento:

- Range di normalità, definito come l'intervallo che va dal valore della media a cui viene sommato lo scarto interquartile, al valore della loro differenza, comprese estremità.

- Range di anormalità, al quale appartengono tutti quei valori non rientranti nel range di normalità.

In seguito ad operazioni di confronto tra le componenti dei vettori delle attivazioni giornaliere con i corrispettivi valori di riferimento, sono stati prodotti vettori le cui componenti contengono:

- “0” se il numero di attivazioni rientra nel range di normalità.
- “1” se invece rientra nel range di anormalità.

Quindi i vettori sono stati riordinati in una matrice di 0 e 1 le cui righe sono le attività e le colonne i giorni a partire dal trentunesimo giorno di osservazione. Per facilitare la comprensione di questi risultati è stata utilizzata la funzione matlab “heatmap”, attraverso la quale è possibile visualizzare una mappa di colori che associa ai valori normali il colore blu, mentre a quelli anormali il bianco.

Denominante le righe con le corrispondenti attività è possibile individuare in quale giorno e per quale attività è stata registrata una anomalia.

Successivamente sono stati distinti due casi di anormalità.

- Il primo si rileva per quei valori maggiori della somma tra media e IQR.
- il secondo si riscontra per i valori minori della differenza tra media e IQR.

In questo modo seguendo le operazioni di confronto precedentemente descritte tra i vettori delle attivazioni giornaliere e gli intervalli di riferimento si ottengono vettori contenenti:

- “1” se i valori di attività sono inferiori alla differenza dei due indicatori.
- “2” se il numero di attivazioni rientra nel range di normalità.
- “3” se i valori di attività superano la somma di media e scarto interquartile.

Come prima attraverso i vettori risultanti è stata prodotta una matrice opportunamente raffigurata attraverso la funzione Matlab “heatmap” che ha generato una mappa di colori:

- Blu per i valori corrispondenti a 1.
- Azzurro per i valori pari a 2.
- Bianco per i valori equivalenti a 3.

In questo modo si è reso possibile distinguere il tipo di anomalia registrata per ogni attività per ogni giorno di registrazione.

Infine, scorrendo la matrice ottenuta distinguendo i due tipi di anormalità, distinguendo i giorni anomali come quelli in cui risulta almeno un'attività anomala, attraverso operazioni iterative di confronto è stato prodotto un vettore di 0 e 1. Il vettore risultante è stato quindi rappresentato con una mappa di colori che associa al bianco i giorni anomali mentre e il blu ai giorni nomali.

4.4 Machine Learning

L'apprendimento automatico è una branca dell'intelligenza artificiale, si basa sul concetto che i sistemi intelligenti possono fare esperienza dai dati, identificare modelli autonomamente e determinare decisioni attraverso un intervento dell'uomo ridotto al minimo.

Le tecniche di apprendimento automatico includono procedimenti computazionali che combinano concetti fondamentali in informatica con concetti di teoria statistica nella costruzione di modelli matematici, probabilità e ottimizzazione, poiché l'obiettivo pratico principale consiste nel ricavare inferenze da un campione, cioè previsioni per campioni, attraverso la progettazione di algoritmi di previsione efficienti e robusti. I procedimenti computazionali fanno uso dell'esperienza per migliorare prestazioni o fornire previsioni accurate.

I ruoli dell'informatica sono:

- Soddisfare la necessità di algoritmi efficienti allo scopo di risolvere il problema dell'ottimizzazione, così come quello di valutare ed elaborare un'enorme quantità di dati. L'efficienza dell'algoritmo di apprendimento o inferenza può essere importante quanto la sua accuratezza predittiva.
- Appreso il modello, ne garantisce rappresentazione e soluzione algoritmica efficienti per l'inferenza. Il modello può essere predittivo per fare previsioni in futuro, o descrittivo per la rappresentazione dai dati, o entrambi. [9]

Di seguito vengono elencate alcune attività di apprendimento automatico standard:

- **Classificazione:** Assegna una particolare categoria a ciascun campione.
- **Regressione:** Prevede un valore reale per ogni campione. La penalità per una previsione inesatta dipende dall'ampiezza della differenza tra valori veri e valori previsti, diversamente dal problema della classificazione, dove generalmente non esiste nozione di vicinanza tra le categorie assegnate.
- **Ranking:** Ordina i campioni secondo un preciso criterio.
- **Clustering:** Partiziona un insieme di elementi in sottoinsiemi omogenei.
- **Apprendimento molteplice:** Trasforma una rappresentazione iniziale di campioni in una di dimensione inferiore mantenendo determinate proprietà della rappresentazione di partenza [10].

Le tipologie dominanti di apprendimento automatico sono:

- Apprendimento automatico supervisionato.
- Apprendimento automatico non supervisionato.

Per procedere con l'individuazione di giorni anormali e normali delle attività studiate, in questo lavoro di tesi si è deciso di procedere con l'utilizzo del Machine Learning non supervisionato. Poiché questo approccio permette di raggruppare elementi in base ad una loro somiglianza [11].

È stata effettuata la suddivisione del periodo di analisi in 8 fasce orarie, ciascuna di tre ore. Procedendo ordinatamente, singolarmente per ogni attività, sono stati creati otto “vettori attivazioni” di lunghezza pari al numero di giorni di osservazione, uno per ogni fascia. In questo modo ogni componente di ciascun vettore contiene il numero di attivazioni rilevato all’interno di una fascia oraria di ciascun giorno.

Per poter distinguere i giorni anormali da quelli normali si è messa in pratica l’attività di “Clustering” in modo da identificare per ciascun vettore attivazione i gruppi omogenei al loro interno. Quindi i “vettori attivazioni” sono stati iterativamente e ordinatamente dati in entrata alla funzione matlab “K-means”. Questo algoritmo è disponibile nell’ambiente Matlab attraverso la funzione “kmeans(X,k)”, che esegue il clustering k-mean per partizionare l’oggetto in entrata X, matrice n per p, in k cluster restituendo un vettore n per 1 contenente gli indici dei cluster di ogni osservazione [8][12].

Il risultato dell’applicazione della funzione kmeans su ogni vettore attivazione è un insieme di otto vettori denominati “vettori cluster”, ciascuno dei quali è composto da componenti contenenti gli indici dei cluster ottenuti.

In seguito a numerose prove con diversi valori di k si è giunti alla conclusione che per k uguale a 3 e a 4 si ottiene una buona correlazione tra vettori cluster e vettori attivazioni corrispondenti alle medesime fasce orarie. Per individuare la correlazione si è fatto ricorso all’uso della funzione matlab “corr” che dati due elementi ne restituisce un coefficiente di correlazione.

Per verificare ulteriormente la correlazione tra vettori cluster e vettori attivazioni è stata applicata la funzione matlab di sintassi “plotregression(targets, outputs)”. Questo tipo di analisi ha permesso di verificare la relazione lineare tra variabili indipendenti (vettori attivazioni) e variabili dipendenti (vettori cluster) o variabili di risposta.

Lo scopo di questo studio è quello di misurare gli effetti delle variabili indipendenti sulla variabile dipendente in funzione di un determinato modello teorico.

Il risultato dell'applicazione della funzione “plotregression” è la traccia della regressione lineare tra vettori attivazione e vettori cluster.

4.5 Studio del significato dei cluster risultanti dall'applicazione dell'apprendimento automatico non supervisionato.

È stato effettuato uno studio preliminare per ricavare il significato degli indici dei cluster ottenuti dall'applicazione dell'apprendimento automatico non supervisionato.

Lo studio si è basato sull'osservazione diretta della sovrapposizione della rappresentazione degli indici di dispersione ottenuti da analisi statistiche alla rappresentazione dei cluster.

Le analisi statistiche effettuate si basano su particolari indici di dispersione statistici, che sono:

- Media.
- Scarto interquartile.
- Deviazione standard.

Dato un insieme di campioni la deviazione standard ne misura la dispersione rispetto alla media. Quanto più i valori sono lontani dalla media tanto più la deviazione standard è grande.

Per ogni vettore attivazione sono stati calcolati gli indici di dispersione elencati precedentemente attraverso operazioni iterative nell'ambiente Matlab. I valori dei risultati sono stati quindi ordinatamente inseriti all'interno di una tabella le cui righe indicano le otto fasce orarie, mentre per le colonne abbiamo media, deviazione standard e scarto interquartile.

A questo punto iterativamente per ogni fascia oraria all'interno dell'ambiente di calcolo matlab sono stati plottati su una stessa figura tutti i dati corrispondenti necessari alla comprensione del significato degli indici dei cluster, quindi:

- Vettore attivazione.
- Vettore cluster.
- Media.
- Somma e differenza tra media e scarto interquartile.

In seguito ad un attento controllo di ogni figura prodotta sono stati ordinatamente annotati i significati trovati per ogni indice cluster, all'interno di due tabelle generali, una per ognuno dei due valori k considerati. Le colonne indicano gli indici cluster mentre le righe le 8 fasce orarie.

I risultati ottenuti da questo studio preliminare risultano poco consistenti poiché in generale i significati ricavati non si confermano omogeneamente per ogni fascia, impedendo di stabilire una espressione globale per ogni indice cluster.

Per verificare i significati trovati per ogni indice si è svolto un particolare controllo: sono stati individuati i giorni corrispondenti ad un determinato cluster, per tali giorni è stato ricavato il numero di attivazioni in modo tale da poter calcolare la media delle attivazioni. Tale procedimento è stato iterato per ogni fascia oraria. I risultati così ottenuti sono stato ordinatamente collezionati all'interno di una tabella le cui righe indicano le fasce orarie mentre le colonne gli indici cluster.

5 Risultati

Di seguito vengono esposti i risultati riguardanti la sola residenza 239.

5.1 Analisi statistiche

5.1.1 Analisi su tutto il periodo di campionamento.

Il periodo di campionamento va da novembre 2019 ad aprile 2020, per un totale di 140 giorni.

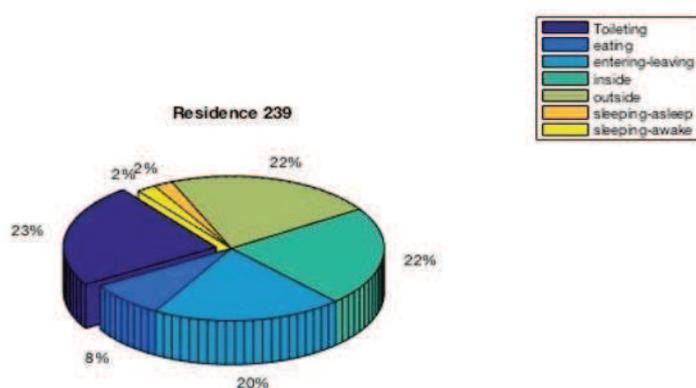


Figura 5: Grafico a torta dei valori percentuali delle attività.

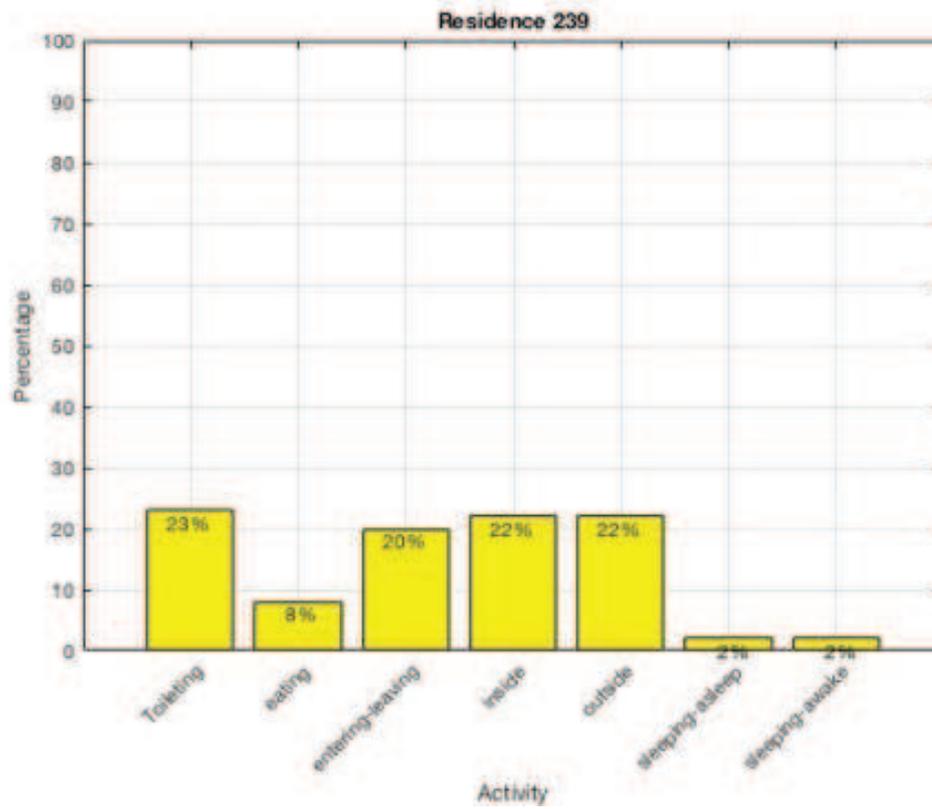


Figura 6: Istogramma dei valori percentuali delle attività.

5.1.2 Analisi su base mensile

Di seguito i risultati dello studio della variabilità mensile ovvero i valori percentuali di ogni attività ottenuti in seguito alla suddivisione in mesi del periodo di analisi.

	Toileting [%]	Eating [%]	Entering_ Leaving [%]	Inside [%]	Outside [%]	Sleeping_ Asleep [%]	Sleeping_ Awake [%]
November 2019	17	7	23	25	25	2	2
December 2019	25	8	20	20	20	3	3
January 2020	36	13	12	15	15	5	4
February 2020	22	7	20	23	23	3	3
March 2020	20	6	22	24	24	2	2
April 2020	20	6	22	25	25	1	1

Tabella 1: Tabella contenente i valori percentuali su base mensile delle attività.

5.1.3 Analisi su base giornaliera

Per ogni attività vengono di seguito esposte le variazioni che avvengono quotidianamente durante le 24 ore che compongono ognuno dei 140 giorni di osservazione della residenza valutata. Così come vengono mostrati i risultati ottenuti dalla suddivisione del periodo di campionamento in mesi per poter meglio individuare i cambiamenti che avvengono giorno per giorno.

- Attività “Toileting”.

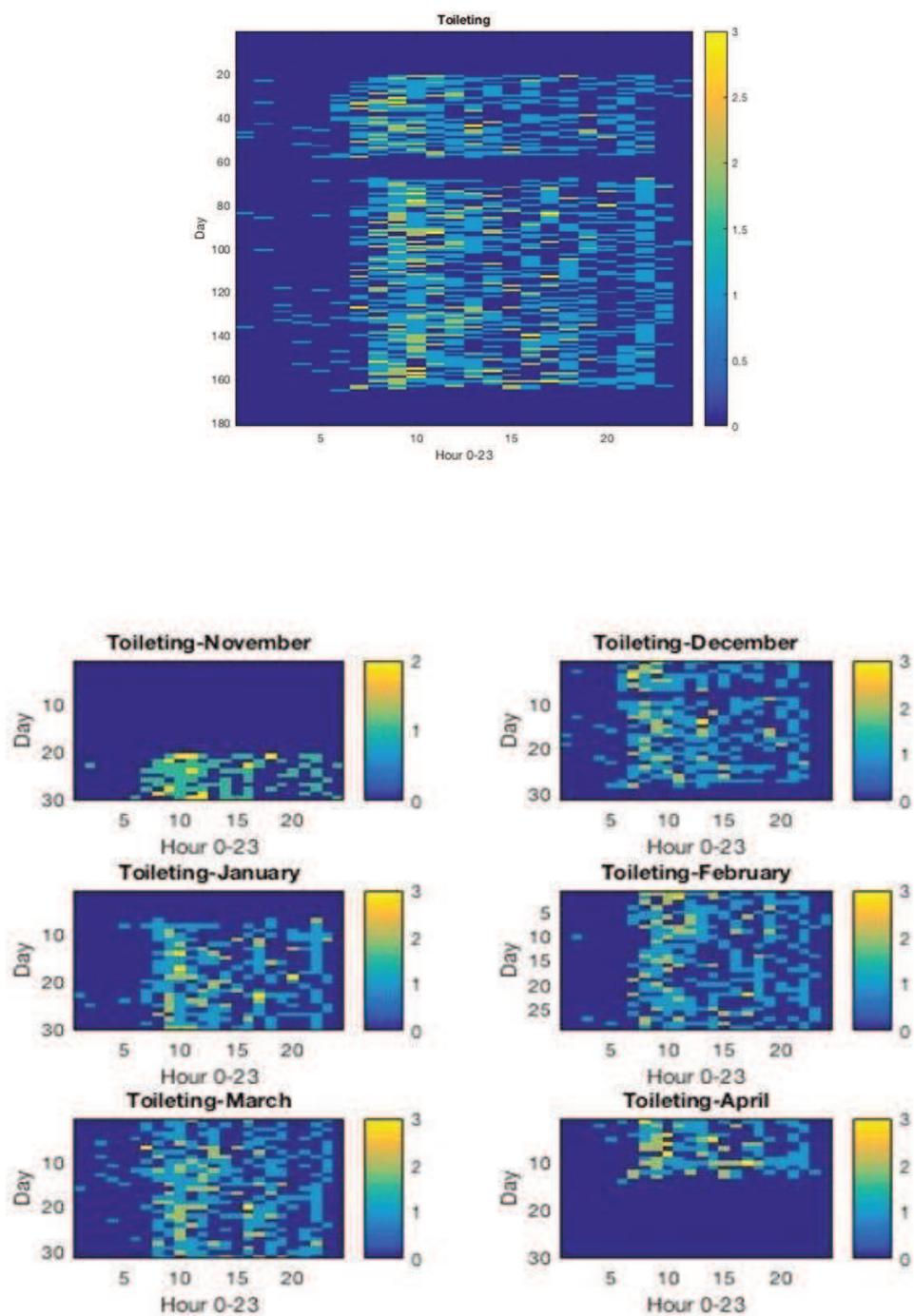


Figura 7: Grafici della variabilità dell'attività Toileting.

- Attività “Eating”

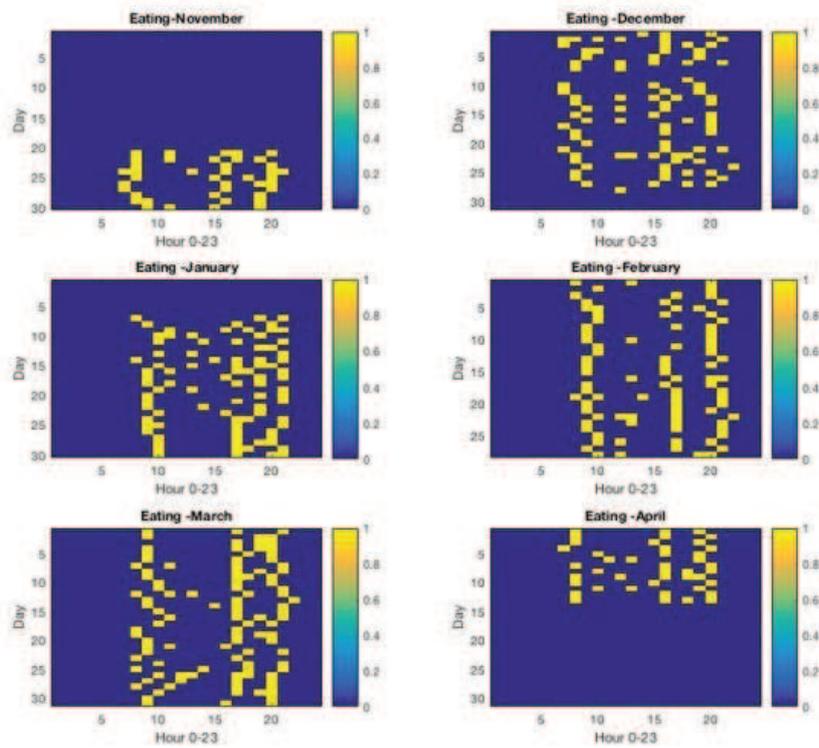
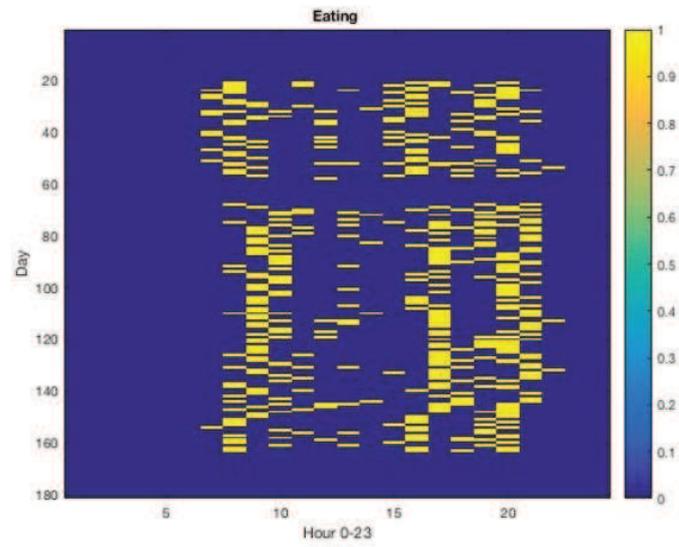


Figura 8: Grafici della variabilità dell'attività Eating.

- Attività “Entering Leaving”

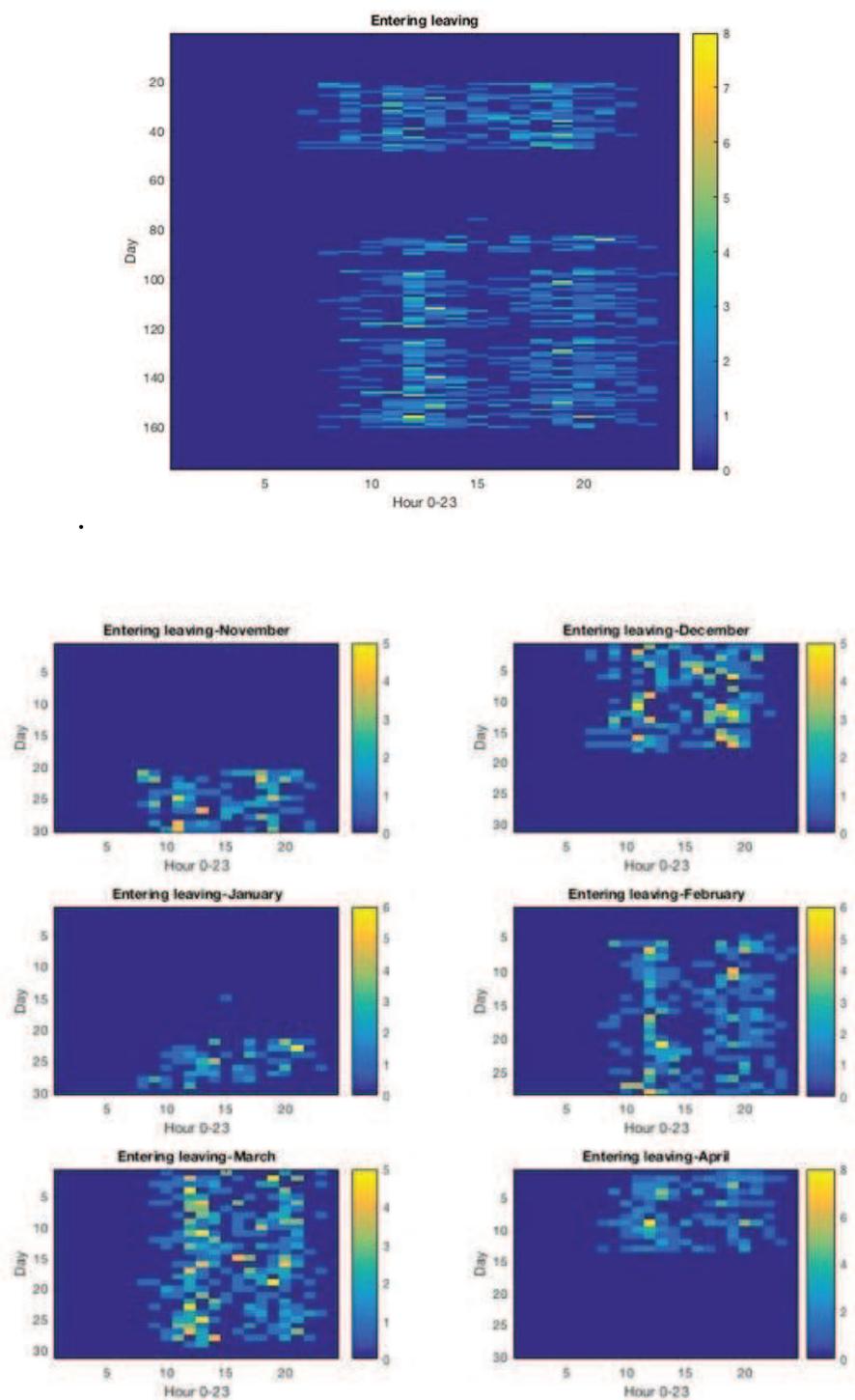


Figura 9: Grafici della variabilità dell'attività Entering Leaving.

- Attività “Inside”.

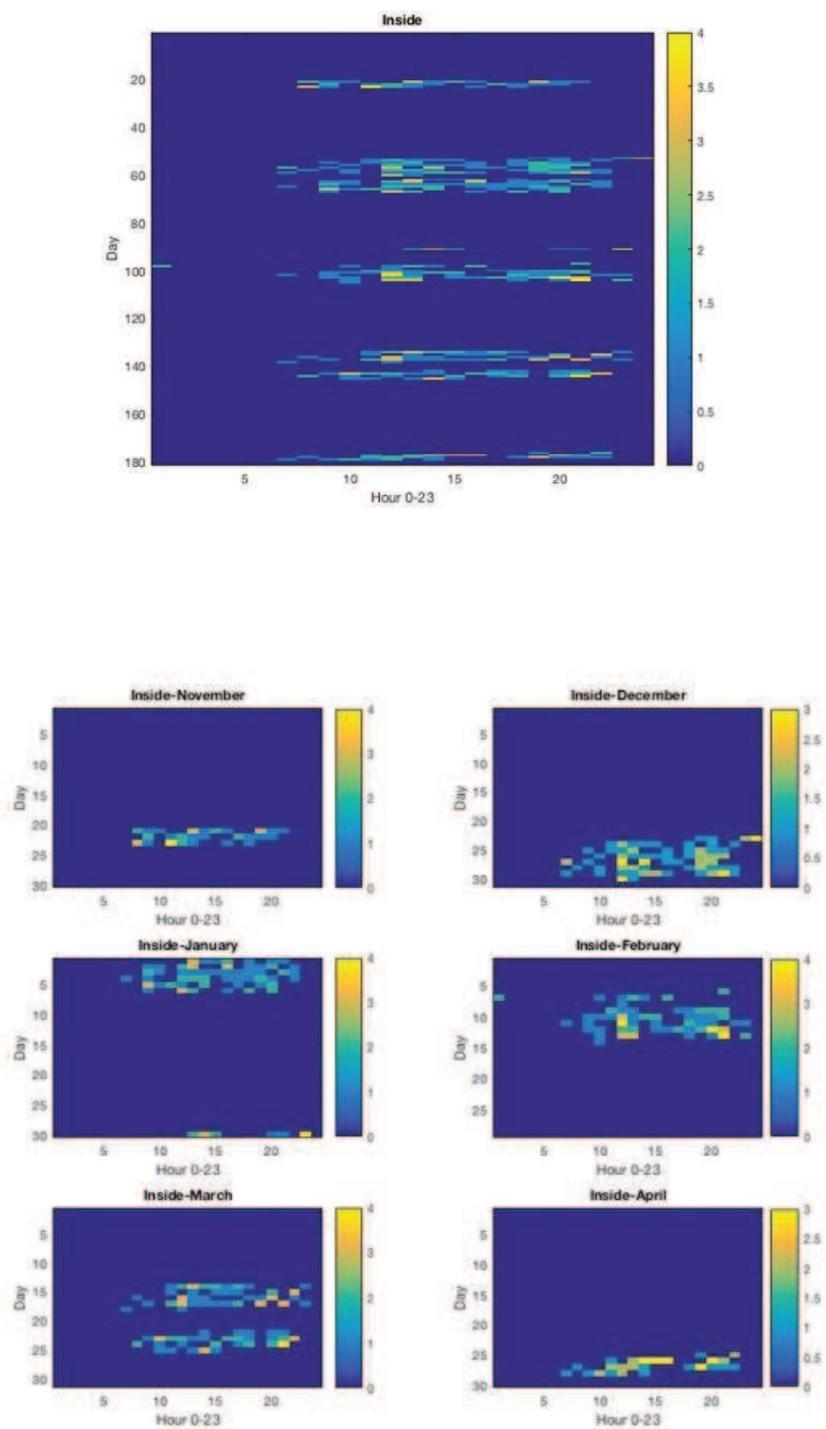


Figura 10: Grafici della variabilità dell'attività Inside.

- Attività “Outside”.

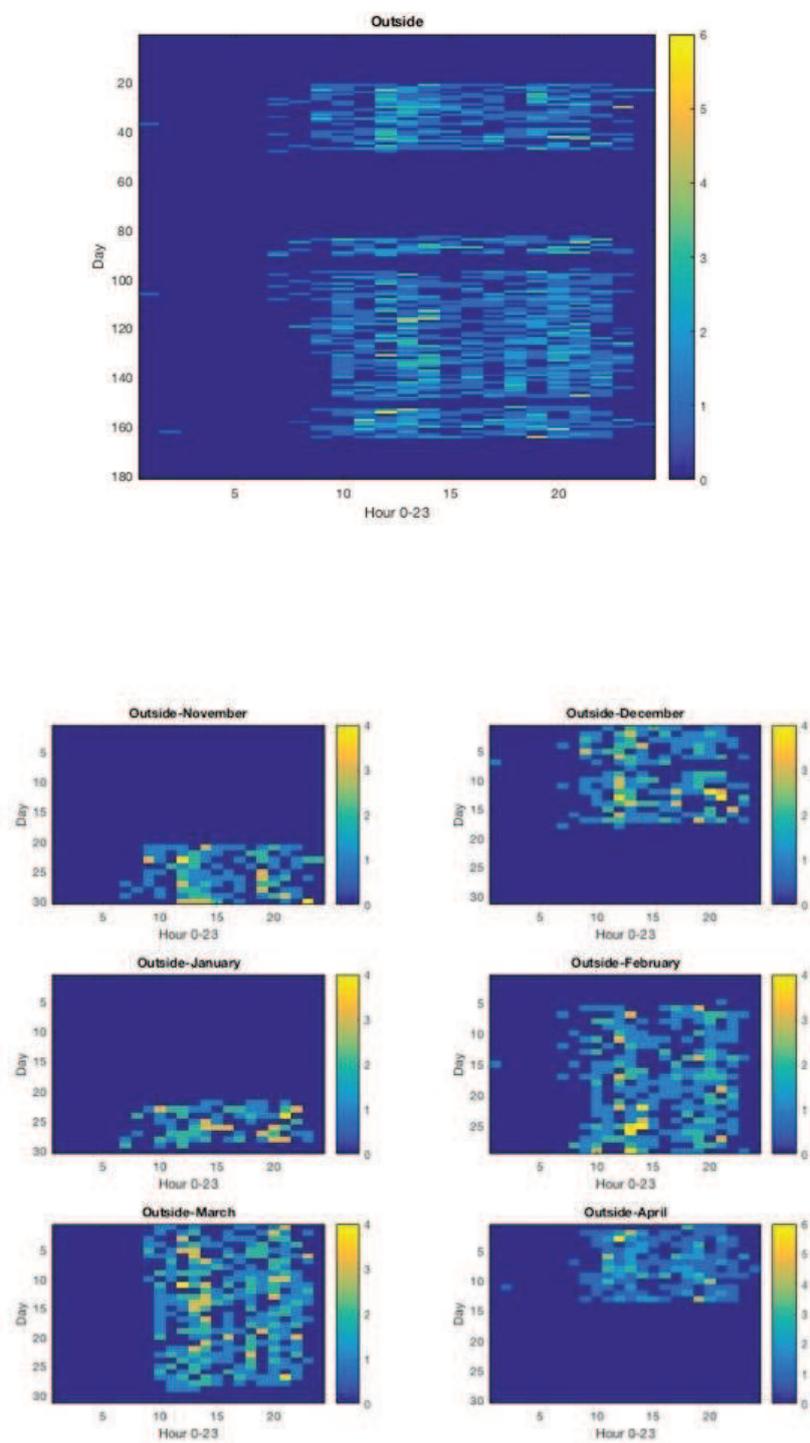


Figura 11: Grafici della variabilità dell'attività Outside

- Attività “Sleeping Asleep”.

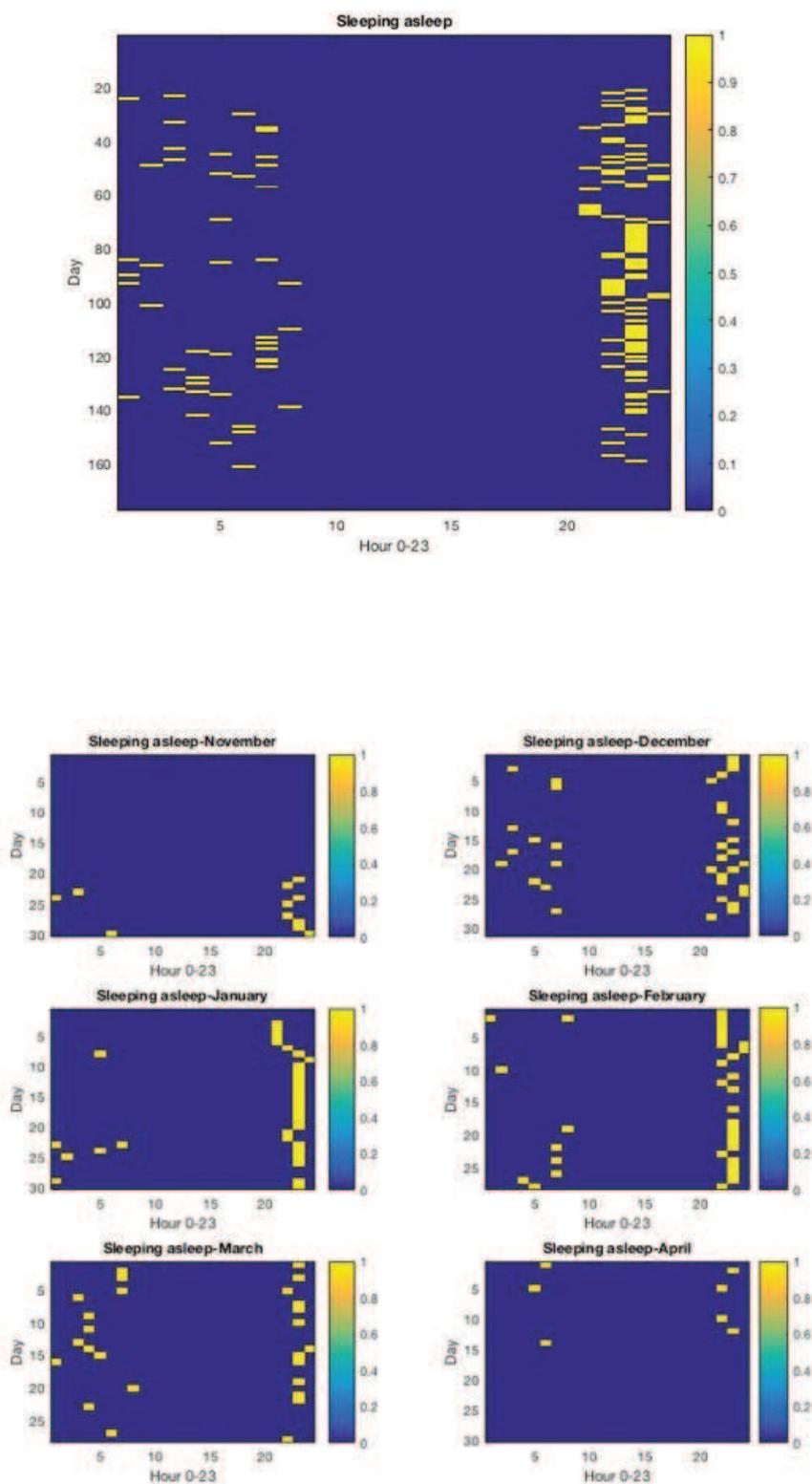


Figura 12: Grafici della variabilità dell'attività Sleeping asleep.

- Attività “Sleeping Awake”.

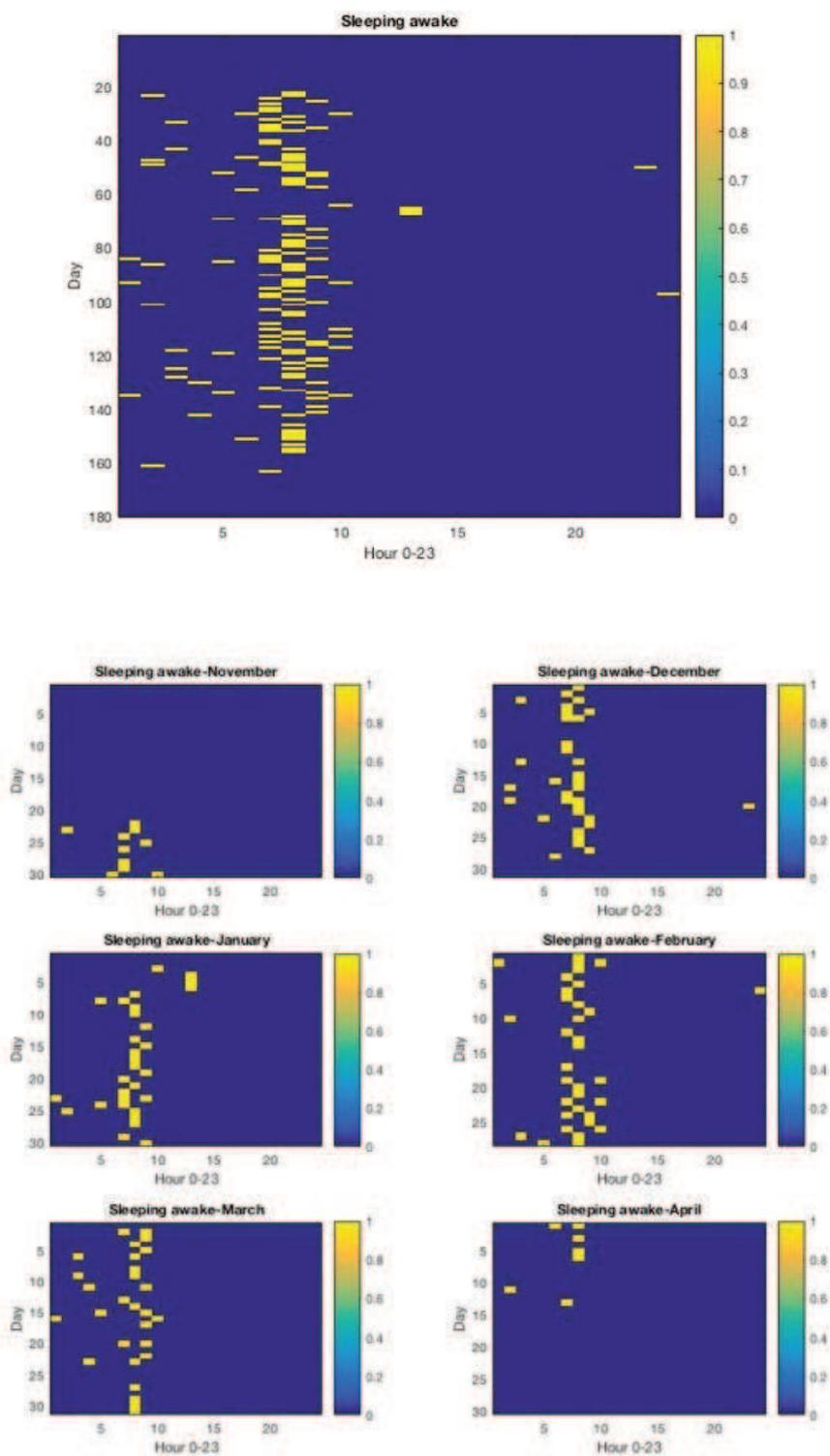


Figura 13: Grafici della variabilità dell'attività Sleeping awake.

5.1.4 Analisi sul periodo di osservazione diviso in fasce orarie

Di seguito vengono espone le tabelle contenenti i risultati dello studio della variabilità delle attività in seguito alla suddivisione delle ore del giorno in 4 fasce orarie:

- Notte: dalle 00:00 alle 6:00.
- Mattina: dalle 7:00 alle 12:00.
- Pomeriggio: dalle 13:00 alle 18:00.
- Sera: dalle 19:00 alle 24:00.

NOVEMBER 2018	Morning (%)	Afternoon (%)	Evening, Evening (%)	Night (%)	Outside (%)	Evening, Outside (%)	Evening, Inside (%)
NIGHT	26	0	0	16	11	16	32
MORNING	23	7	22	23	24	0	2
AFTERNOON	14	7	24	27	28	0	0
EVENING	11	9	24	27	24	6	0
DECEMBER 2018	Morning (%)	Afternoon (%)	Evening, Evening (%)	Night (%)	Outside (%)	Evening, Outside (%)	Evening, Inside (%)
NIGHT	33	0	0	9	7	22	28
MORNING	31	8	17	20	20	0	4
AFTERNOON	23	9	25	21	22	0	0
EVENING	18	8	23	21	20	9	0
JANUARY 2019	Morning (%)	Afternoon (%)	Evening, Evening (%)	Night (%)	Outside (%)	Evening, Outside (%)	Evening, Inside (%)
NIGHT	31	0	0	9	9	20	31
MORNING	46	11	10	13	13	0	7
AFTERNOON	31	17	16	18	19	0	0
EVENING	26	13	12	17	16	15	0
FEBRUARY 2019	Morning (%)	Afternoon (%)	Evening, Evening (%)	Night (%)	Outside (%)	Evening, Outside (%)	Evening, Inside (%)
NIGHT	31	0	0	9	9	20	31
MORNING	28	7	21	21	21	0	4
AFTERNOON	23	6	19	25	27	0	0
EVENING	14	9	20	26	23	8	0
MARCH 2019	Morning (%)	Afternoon (%)	Evening, Evening (%)	Night (%)	Outside (%)	Evening, Outside (%)	Evening, Inside (%)
NIGHT	31	0	0	8	8	20	31
MORNING	26	6	24	20	20	0	3
AFTERNOON	19	7	22	26	27	0	0
EVENING	15	7	20	30	28	4	0
APRIL 2019	Morning (%)	Afternoon (%)	Evening, Evening (%)	Night (%)	Outside (%)	Evening, Outside (%)	Evening, Inside (%)
NIGHT	30	0	0	9	9	20	31
MORNING	24	5	25	23	23	0	1
AFTERNOON	17	7	21	27	28	0	0
EVENING	15	5	22	29	27	3	0

Tabella 2: Contiene i valori percentuali ottenuti in seguito alla divisione del periodo di osservazione in 4 fasce orarie.

5.2 Studio dei casi di normalità e anormalità

L'intervallo di normalità contiene i valori che ricadono tra media e scarto interquartile mentre l'intervallo di anormalità comprende i valori che non rientrano nel range di normalità.

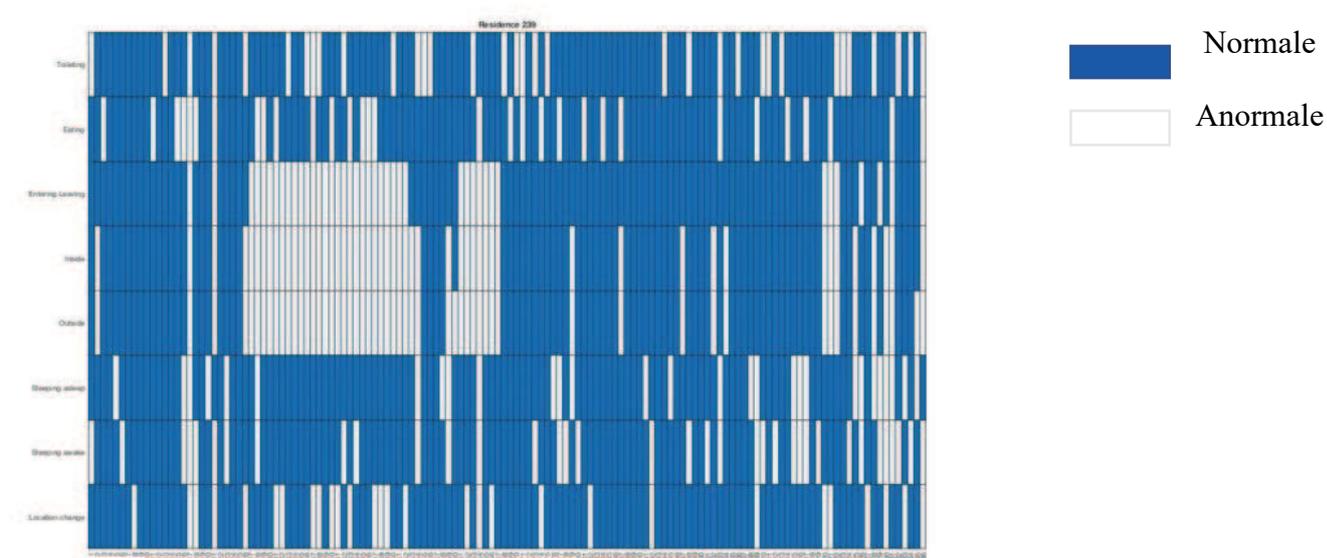


Figura 14: Grafico normalità e anormalità delle attività.

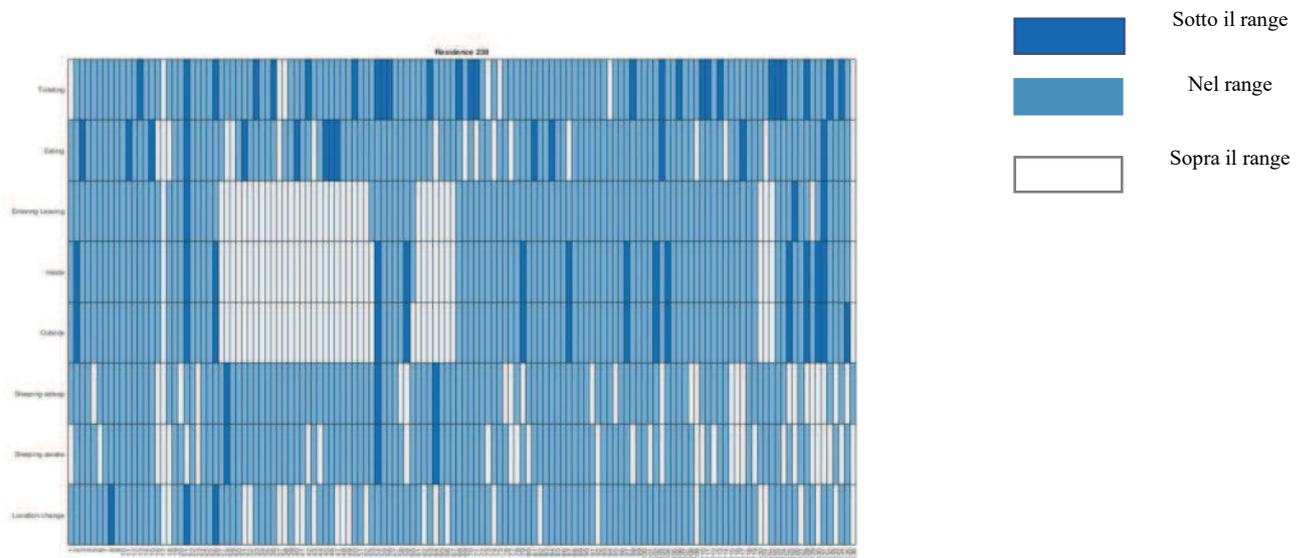


Figura 15: Grafico normalità e due tipi di anomalità delle attività.

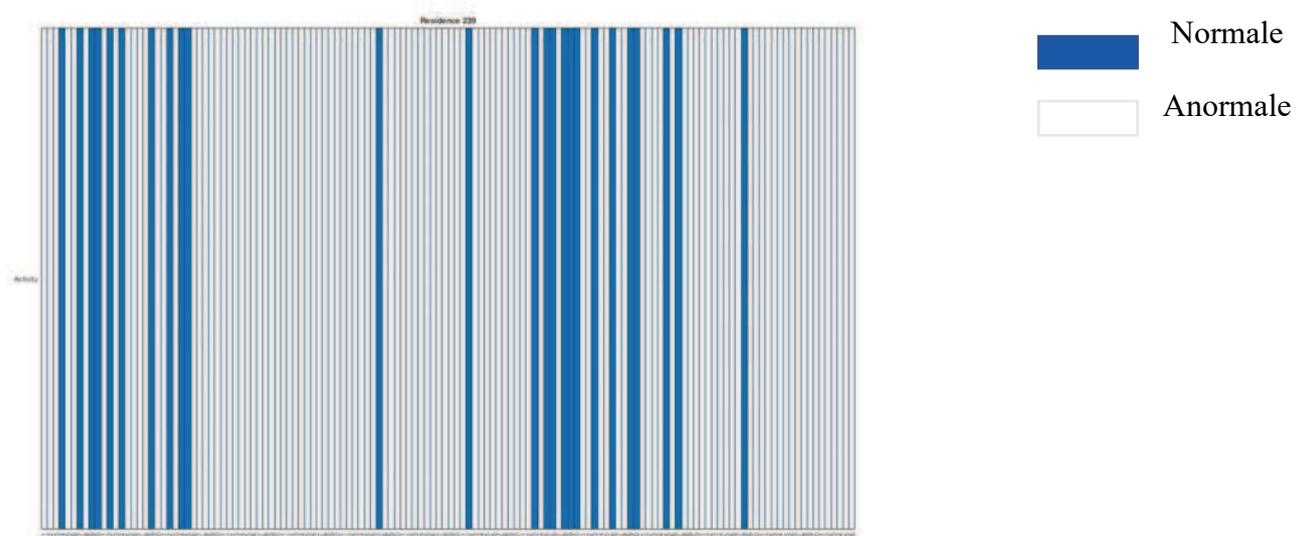


Figura 16: Grafico dei giorni normali e dei giorni anormali della residenza 239.

5.3 Machine learning unsupervised

I grafici a seguire sono stati ottenuti in seguito alla suddivisione del periodo di osservazione in 8 fasce orarie, sono il risultato del plot degli indici cluster, ottenuti dall'applicazione della funzione kmeans prima con k uguale a 3 e poi uguale a 4, sovrapposti ai plot dei vettori delle attivazioni delle attività.

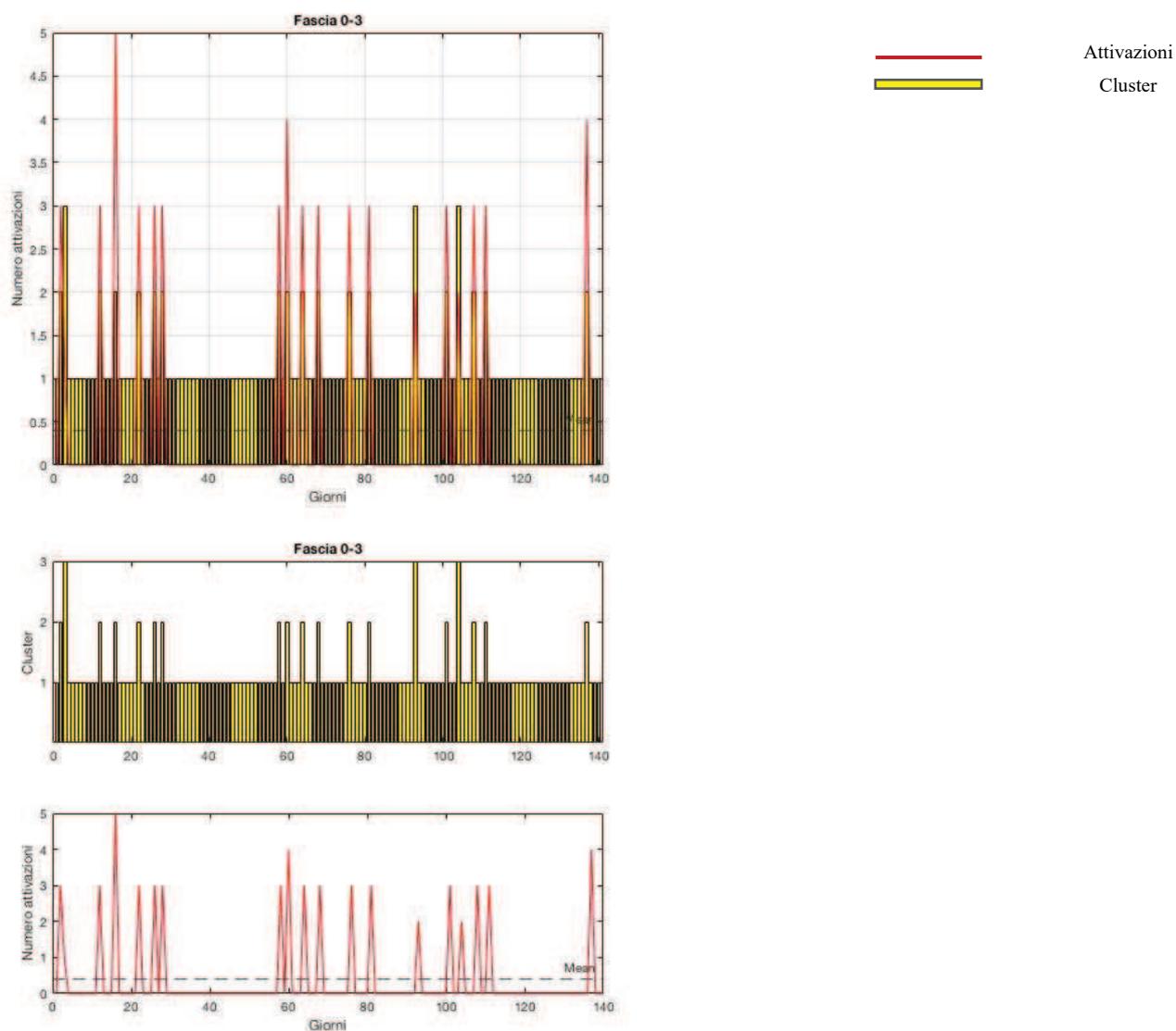


Figura 17: Grafici attivazioni e cluster con k=3 (fascia 0-3).

— Attivazioni
— Cluster

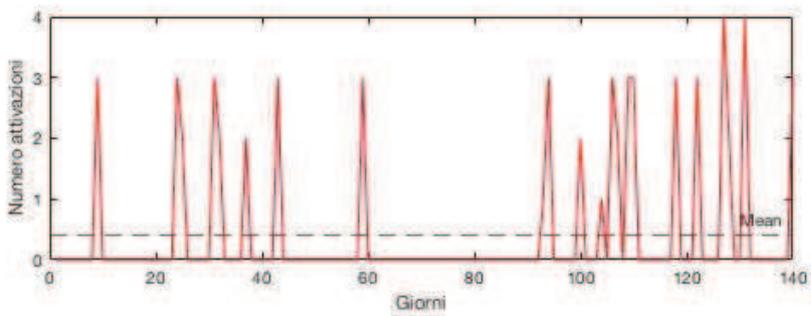
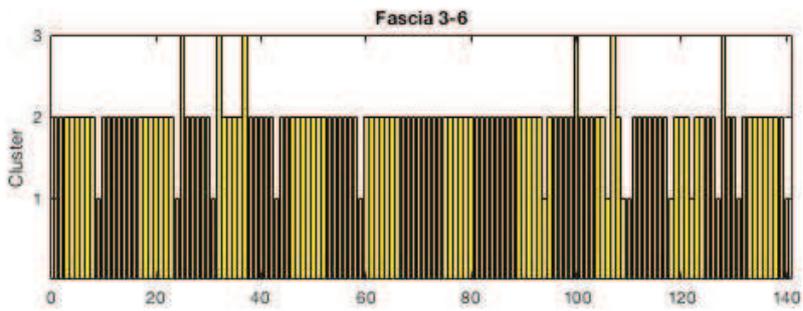
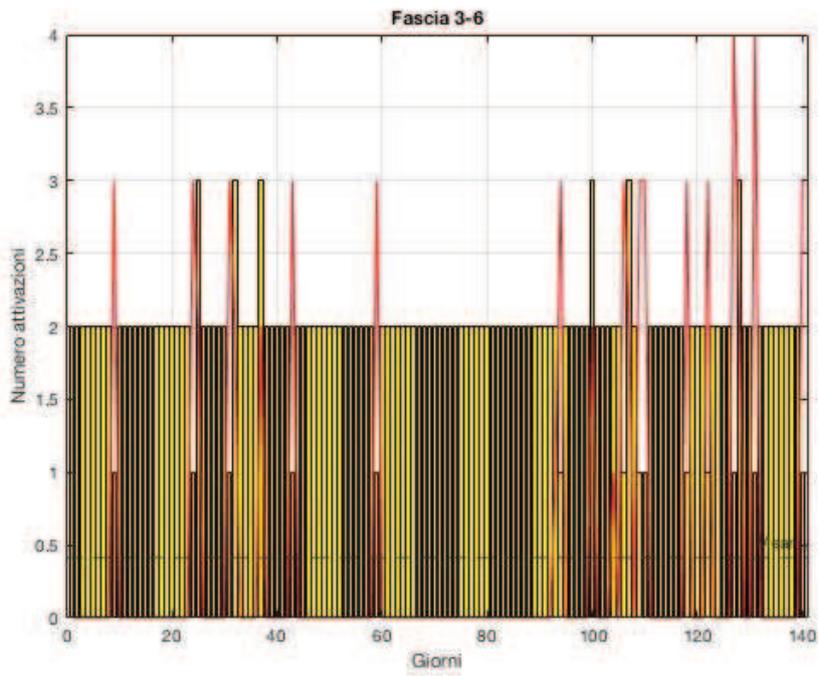


Figura 18: Grafici attivazioni e cluster con $k=3$ (fascia 3-6).

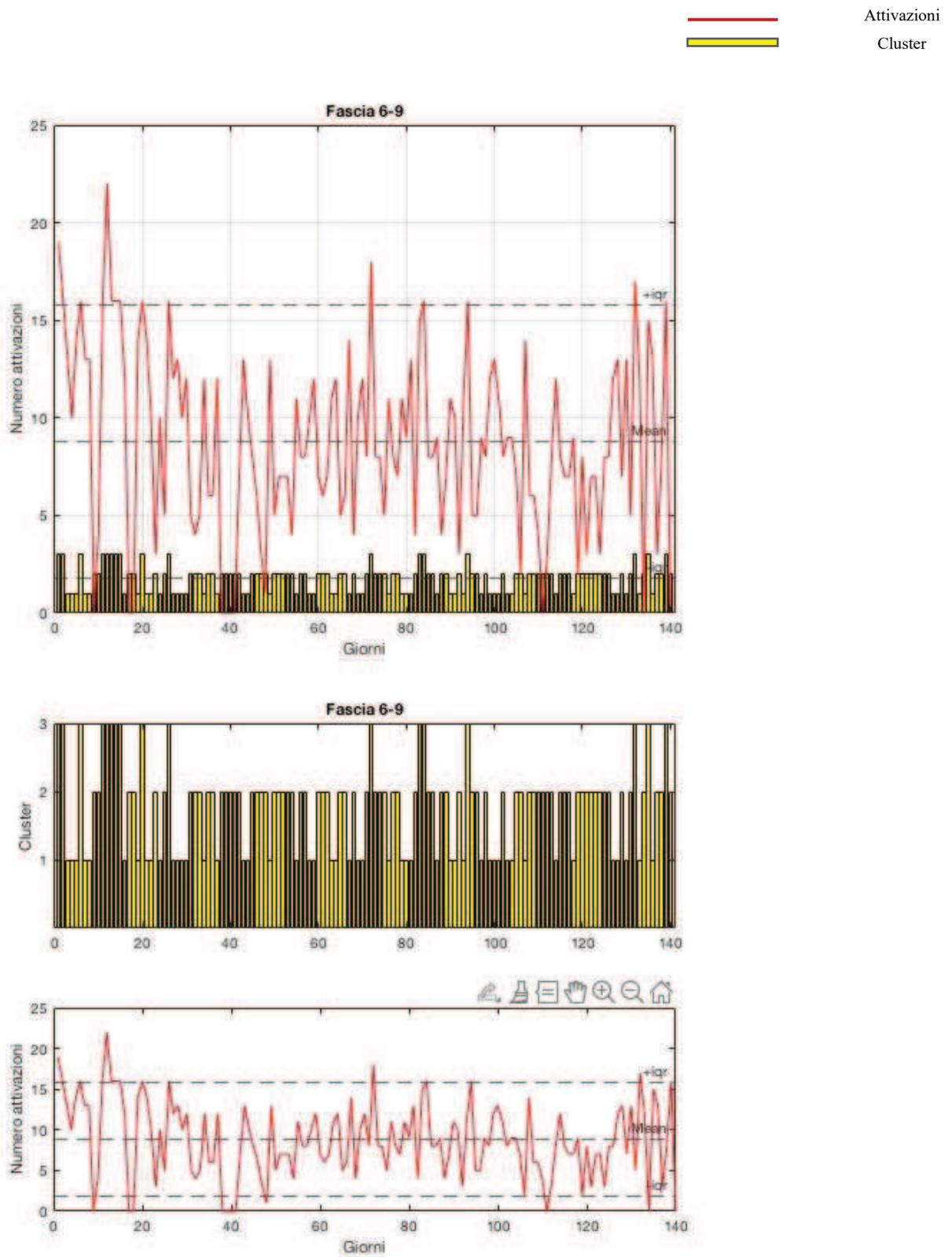


Figura 19: Grafici attivazioni e cluster con $k=3$ (fascia 6-9).

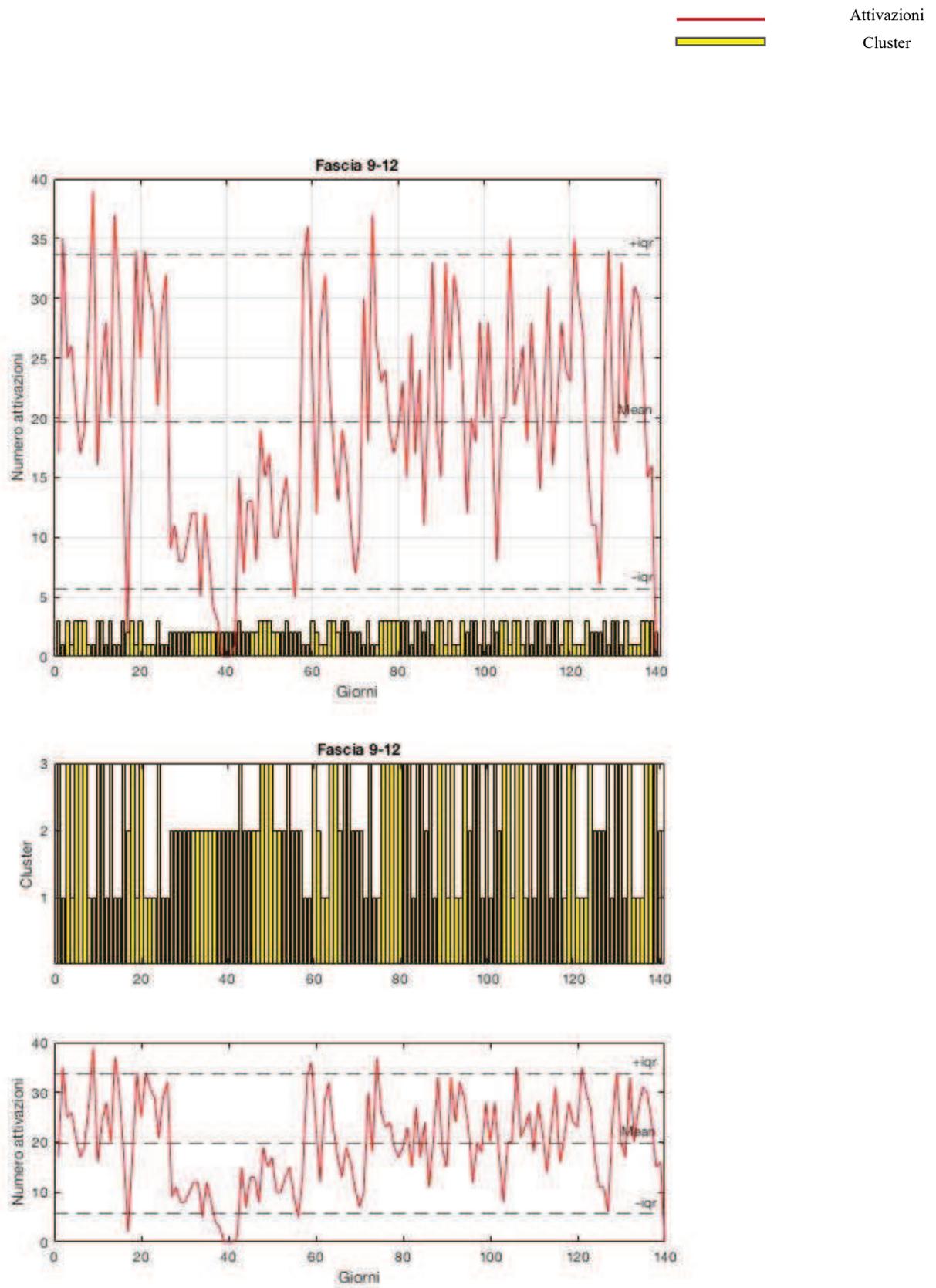


Figura 20: Grafici attivazioni e cluster con k=3 (fascia 9-12).

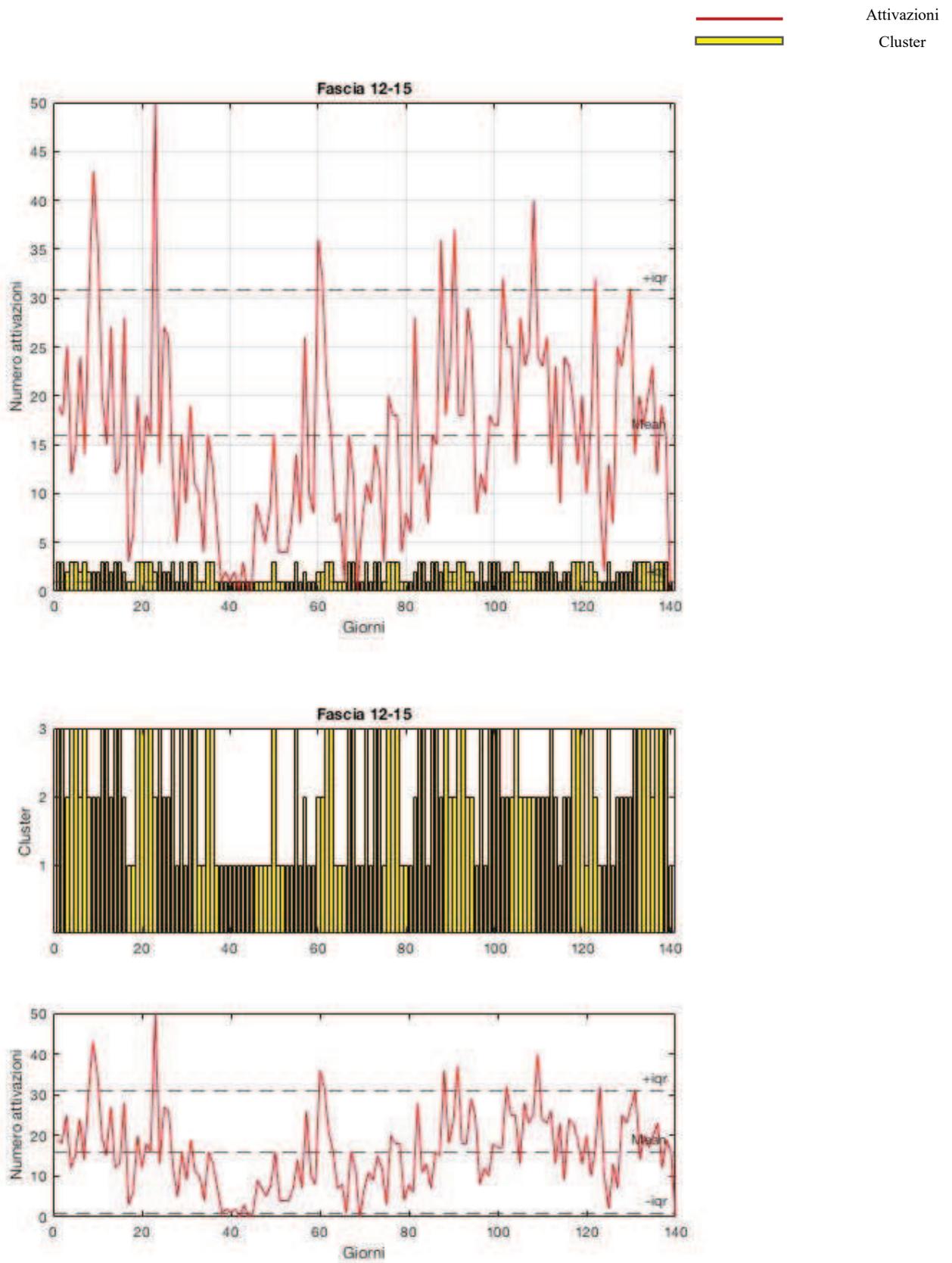


Figura 21: Grafici attivazioni e cluster con k=3 (fascia 12-15).

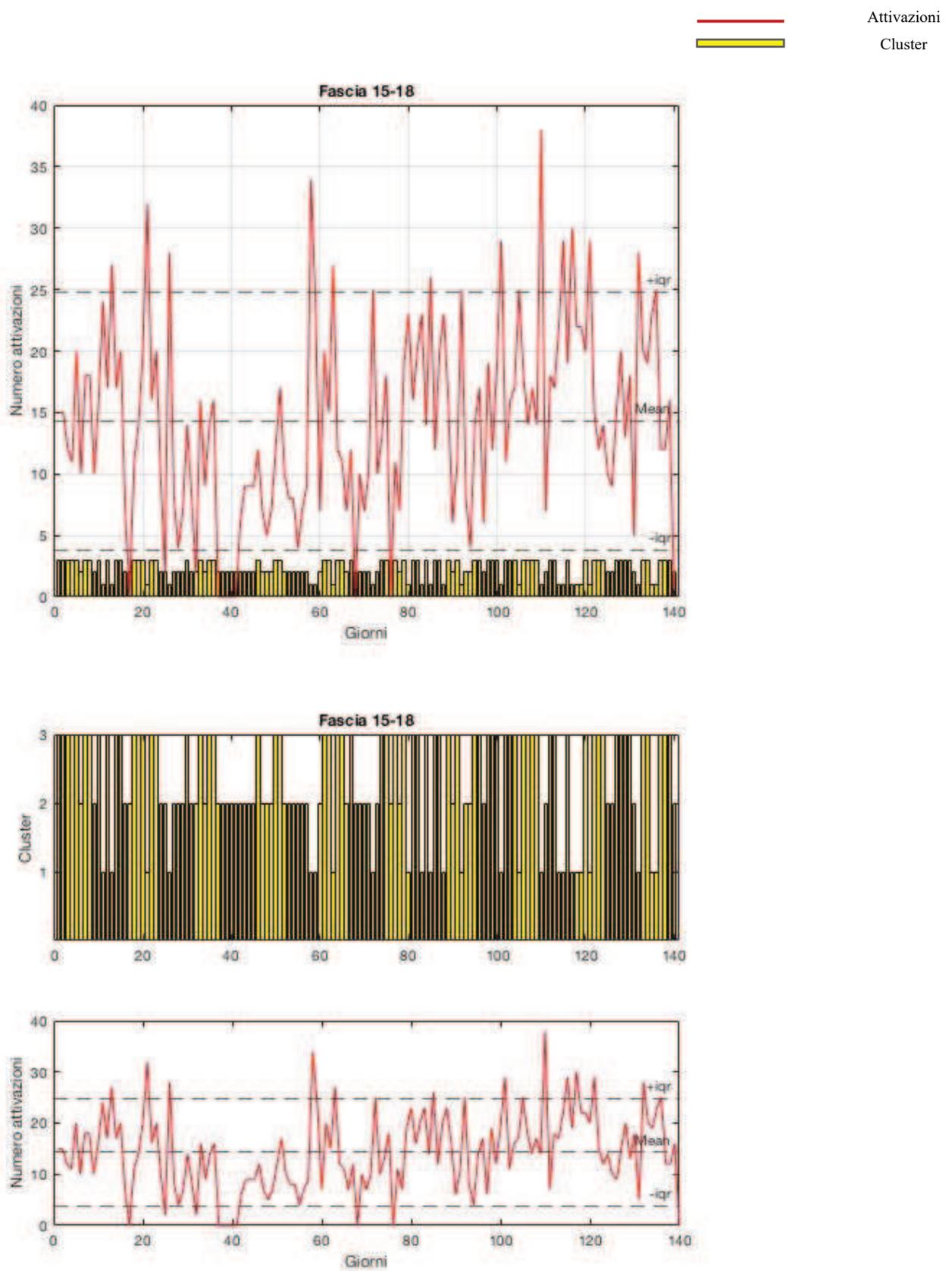


Figura 22: Grafici attivazioni e cluster con $k=3$ (fascia 15-18).

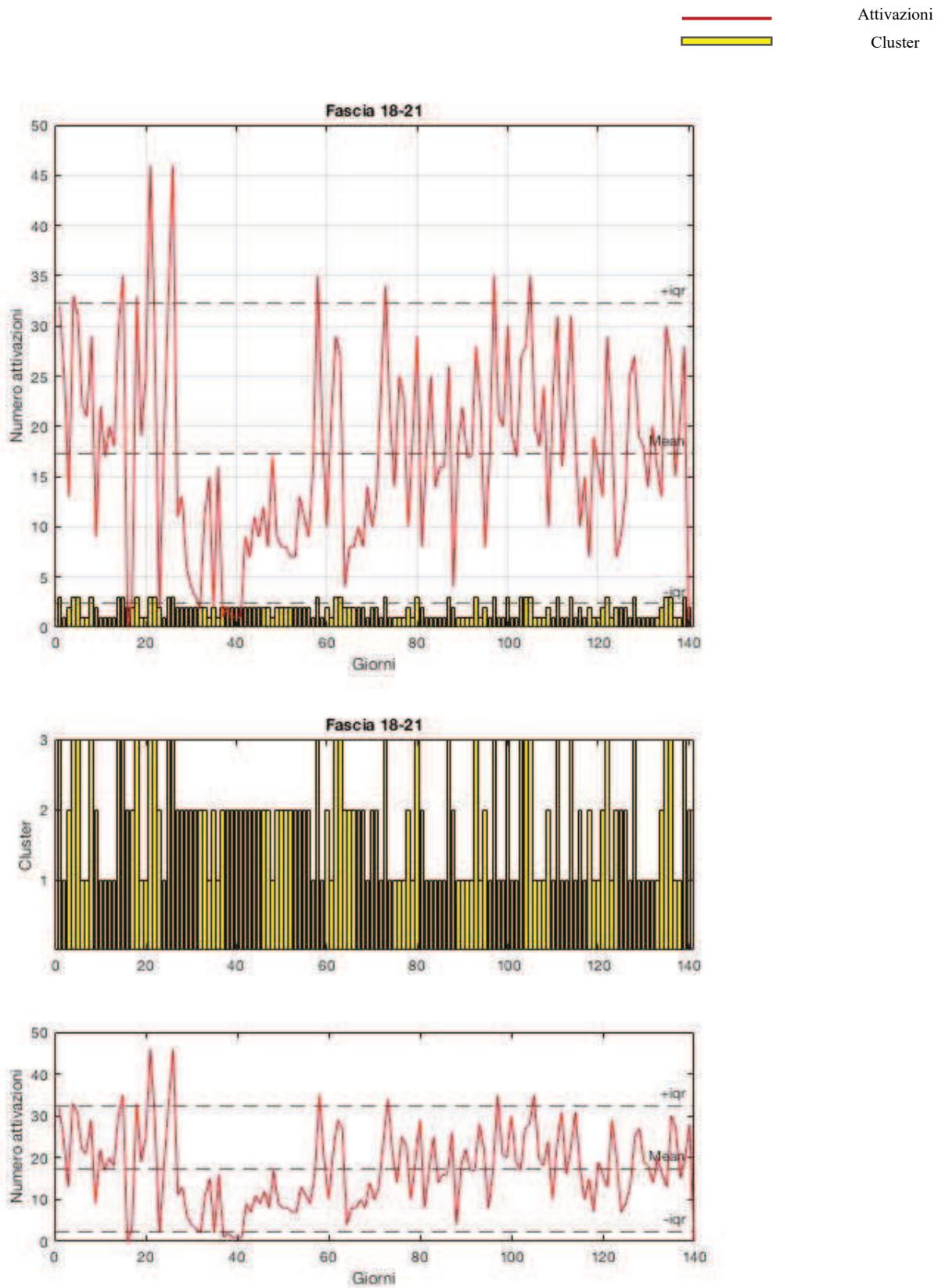


Figura 23: Grafici attivazioni e cluster con k=3 (fascia 18-21).

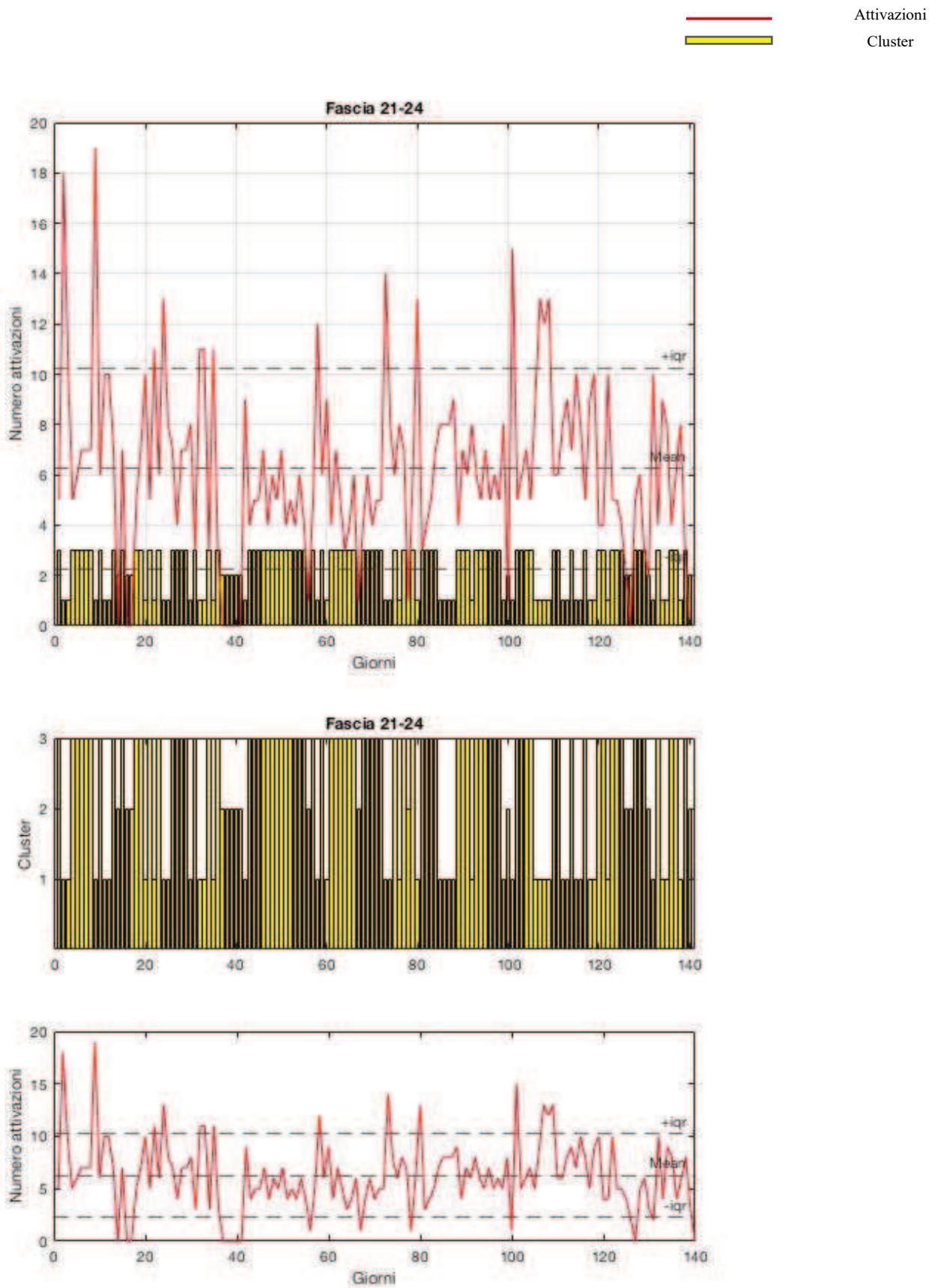


Figura 24: Grafici attivazioni e cluster con $k=3$ (fascia 21-24).

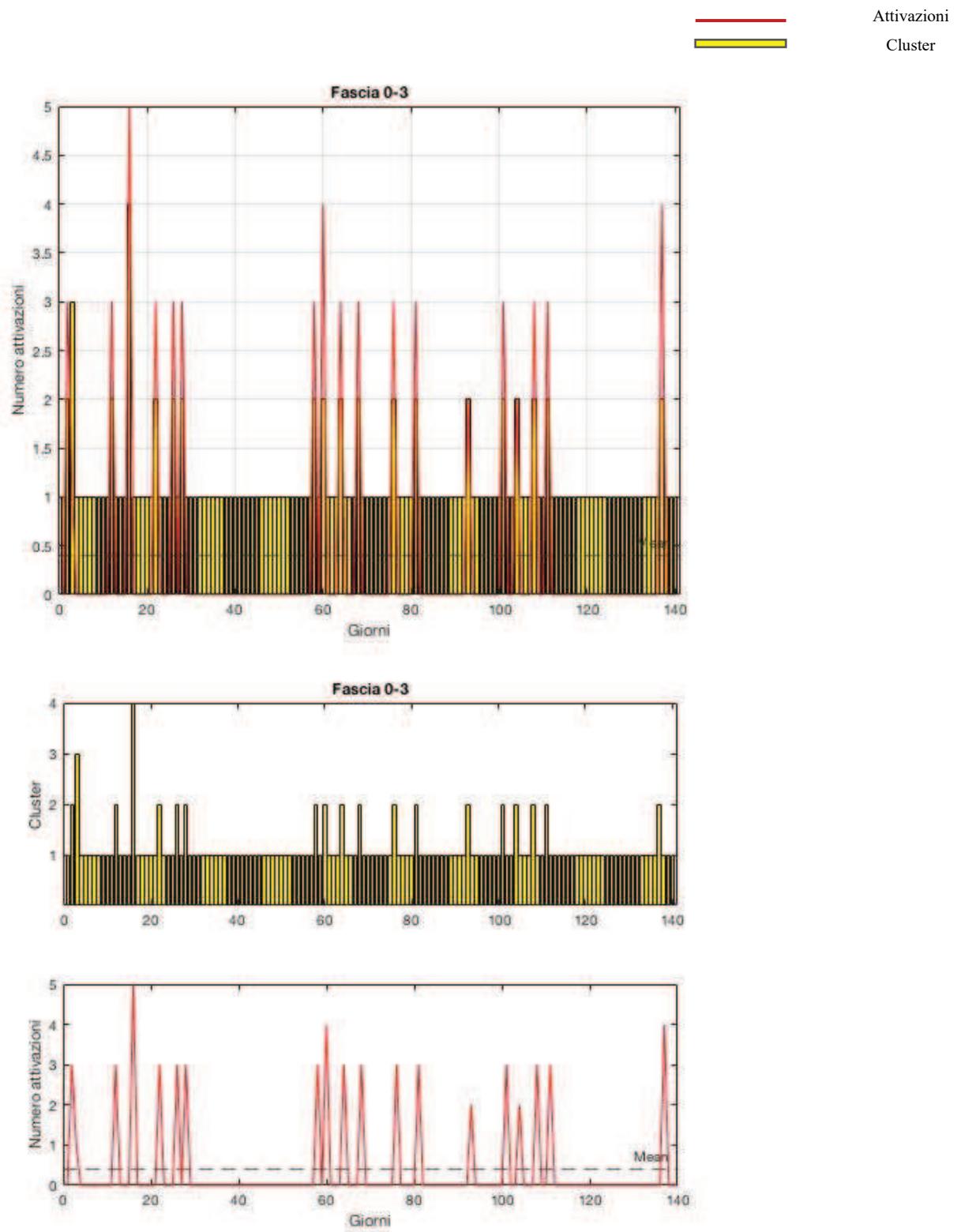


Figura 25: Grafici attivazioni e cluster con k=4 (fascia 0-3).

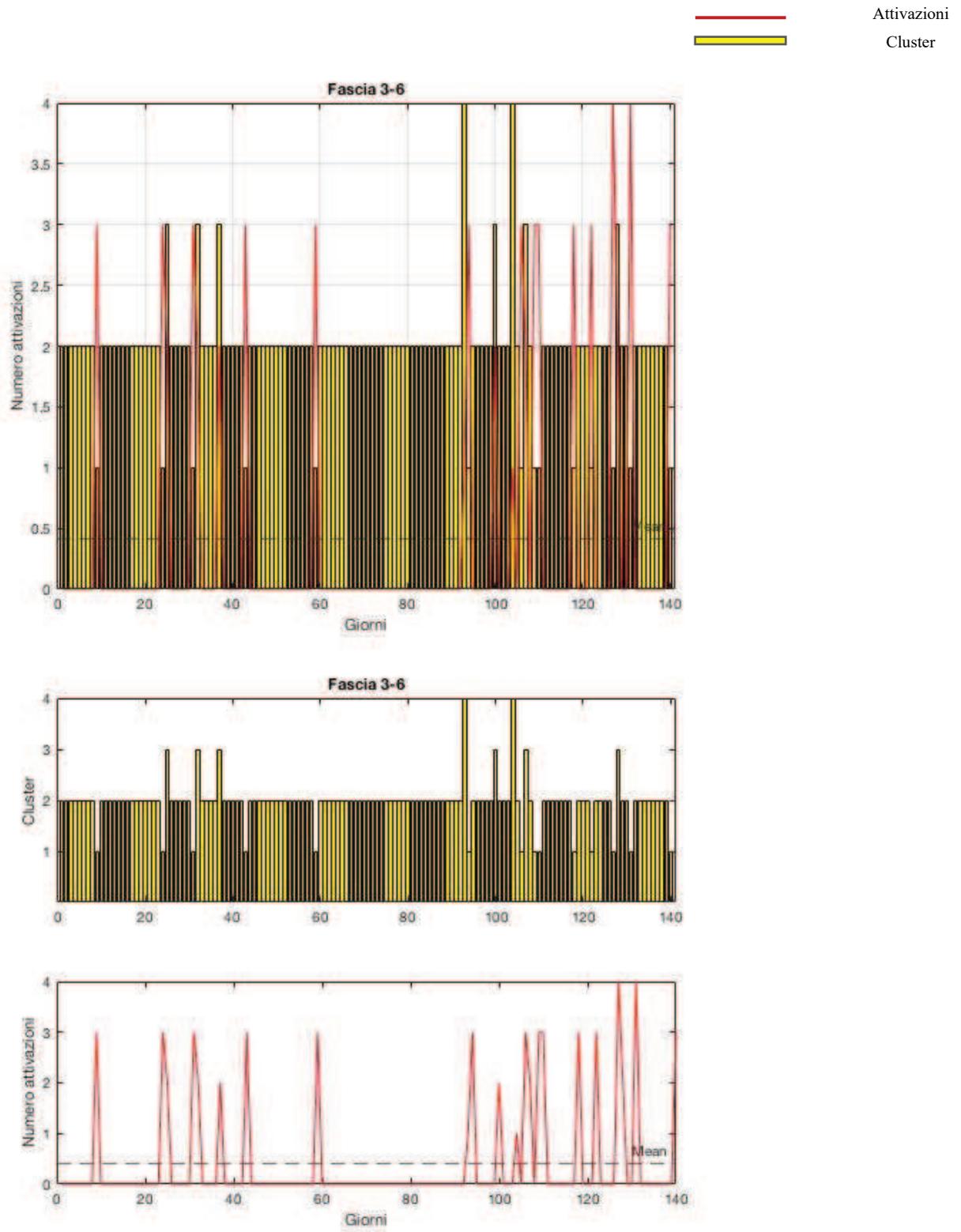


Figura 26: Grafici attivazioni e cluster con $k=4$ (fascia 3-6).

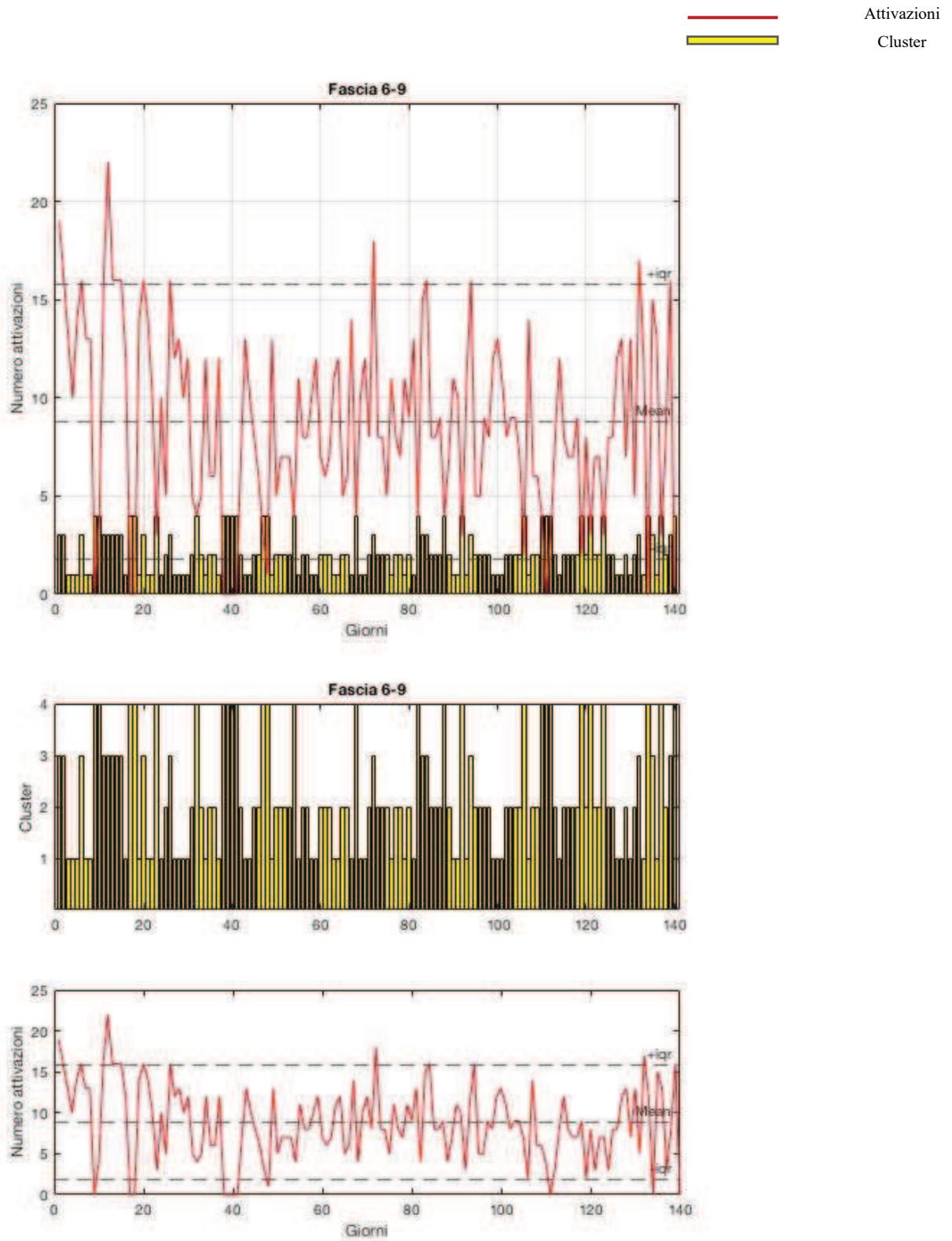


Figura 27: Grafici attivazioni e cluster con k=4 (fascia 6-9).

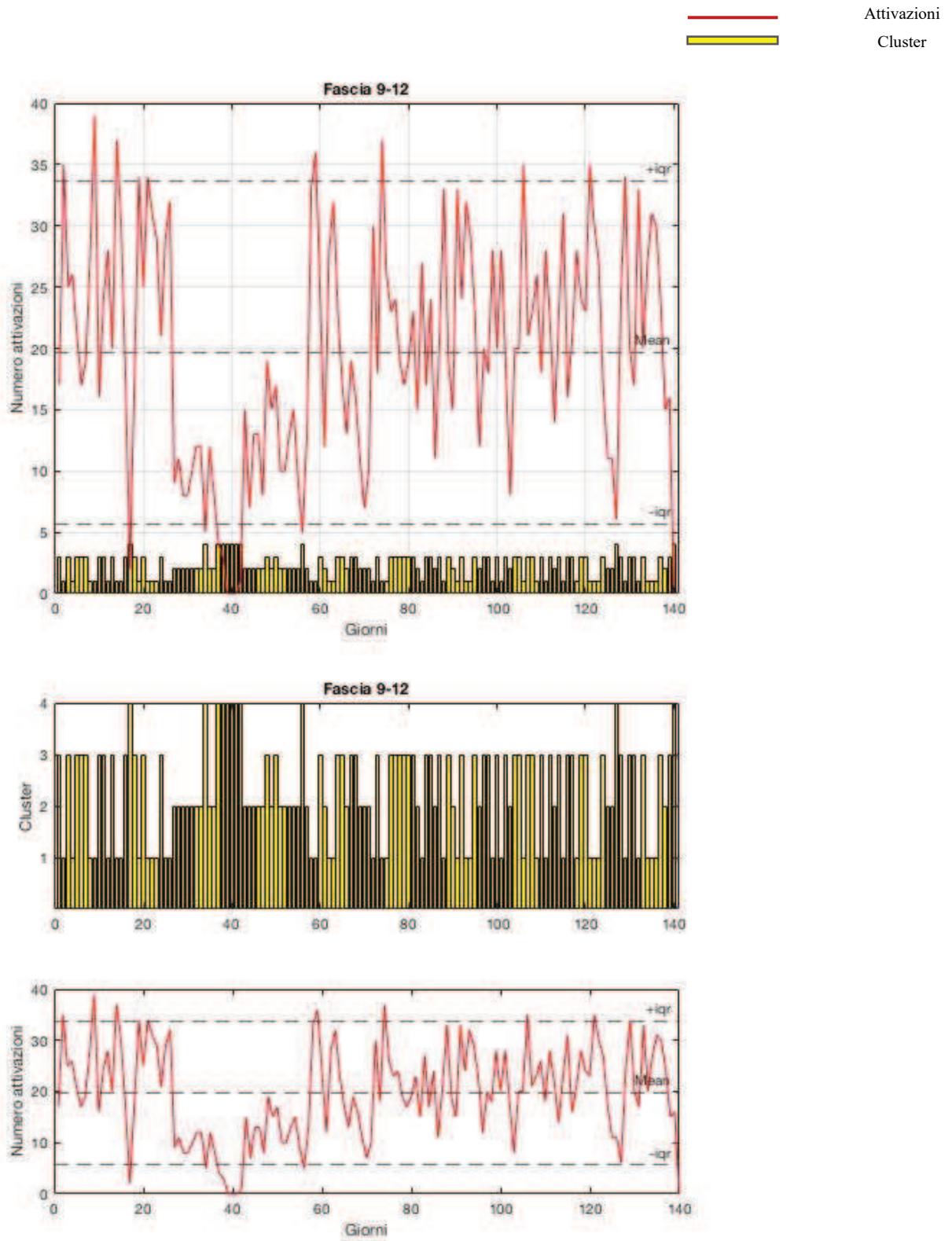


Figura 28: Grafici attivazioni e cluster con $k=4$ (fascia 9-12).

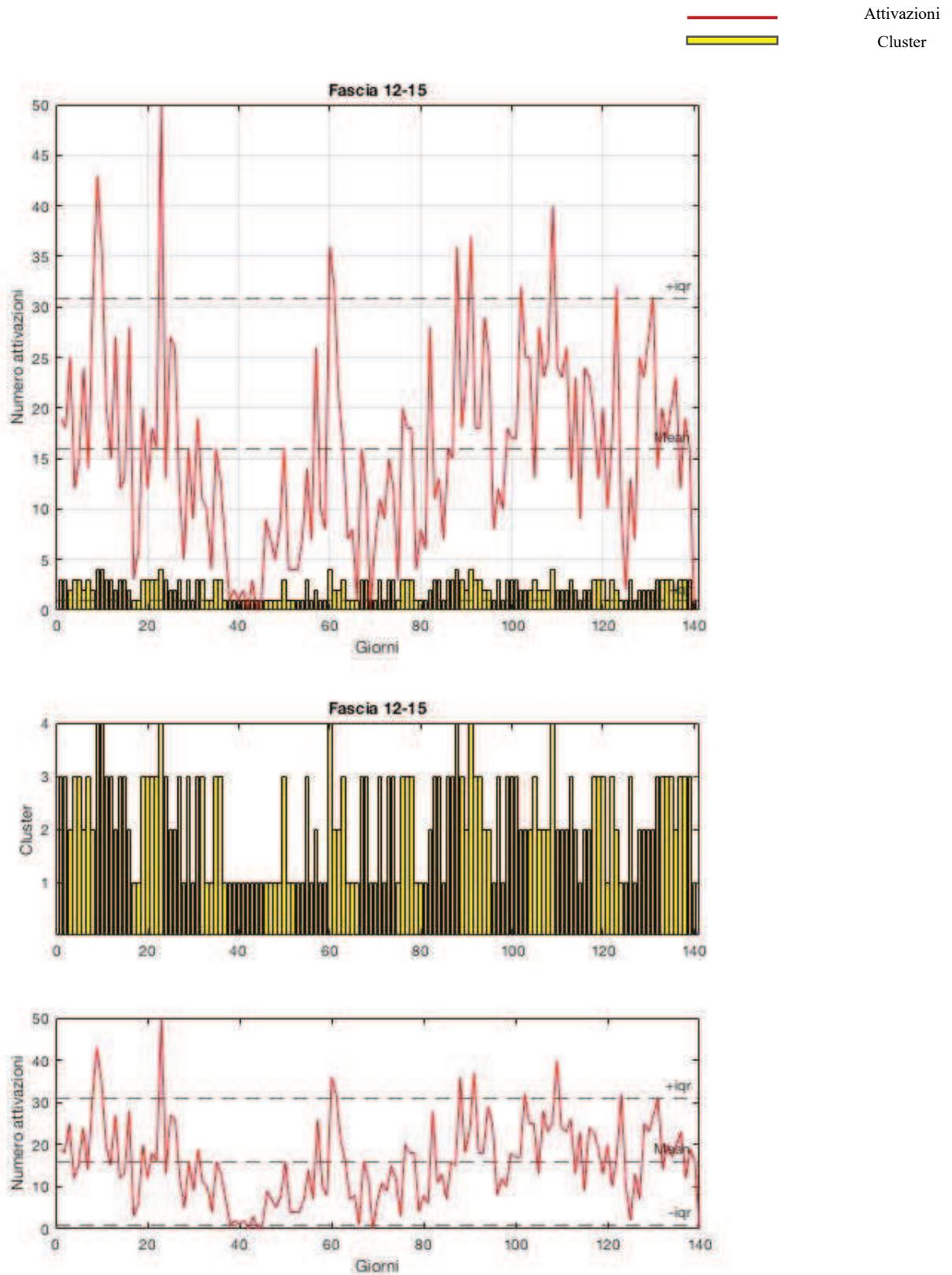


Figura 29: Grafici attivazioni e cluster con $k=4$ (fascia 12-15).

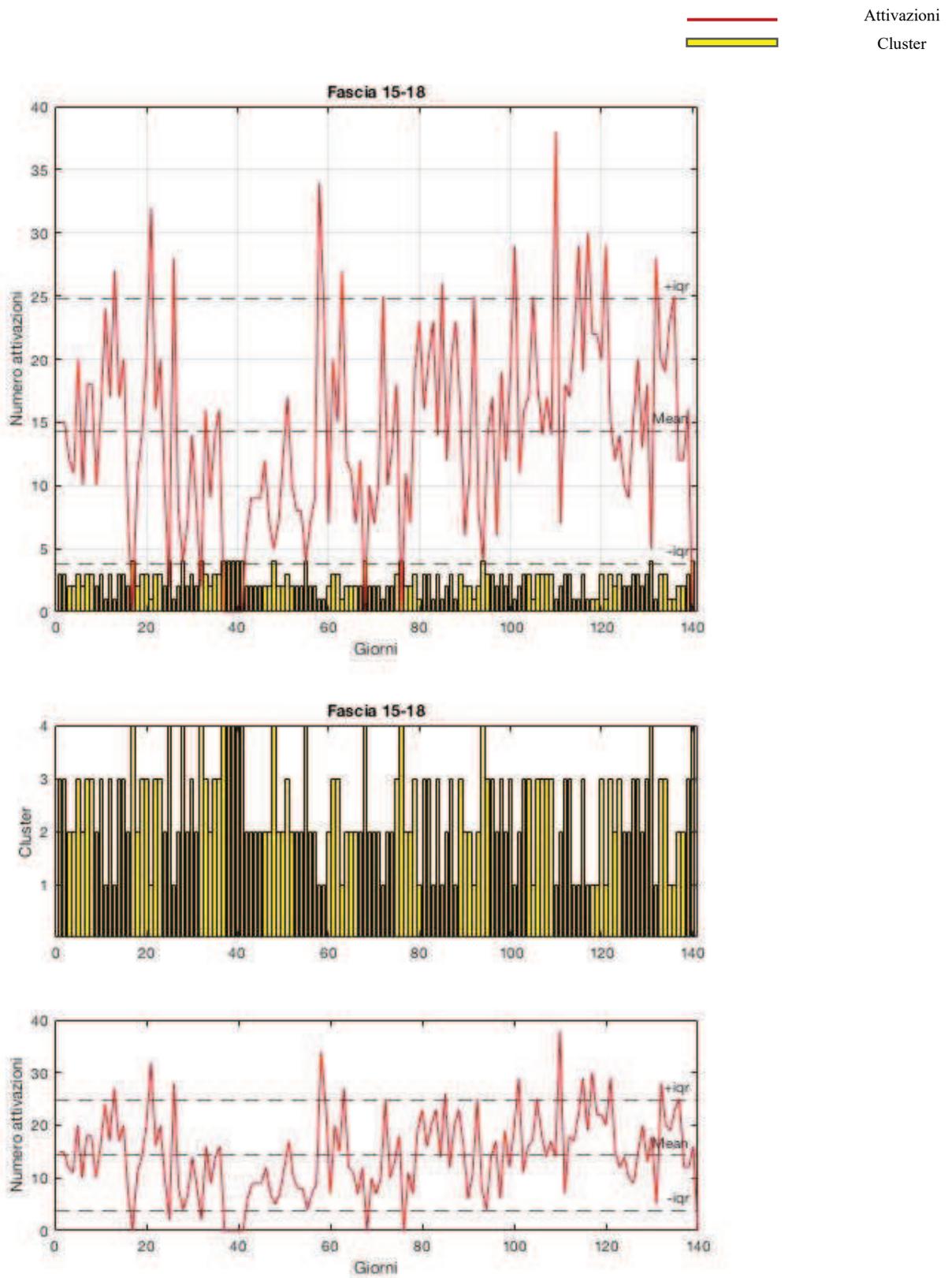


Figura 30: Grafici attivazioni e cluster con $k=4$ (fascia 15-18).

— Attivazioni
— Cluster

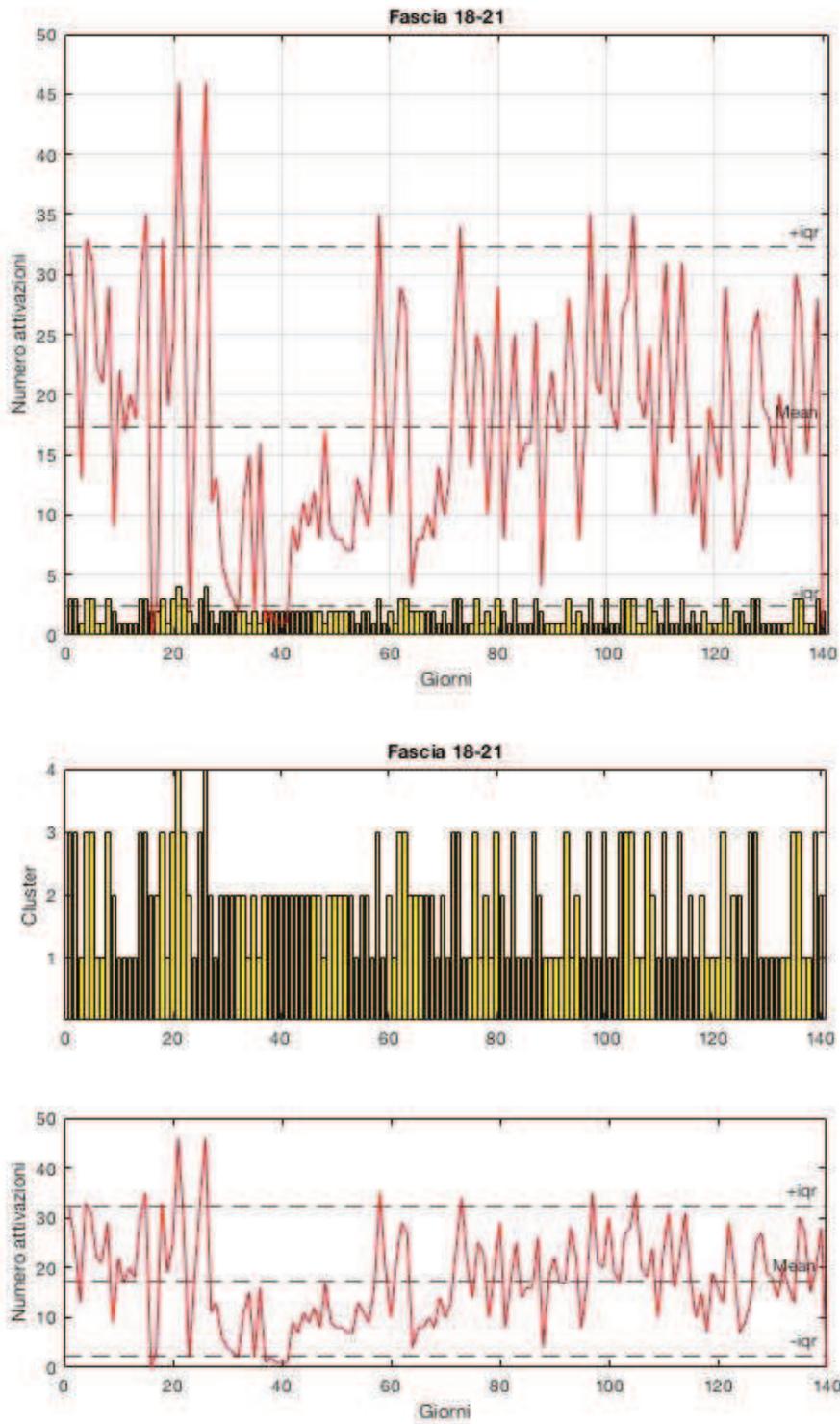


Figura 31: Grafici attivazioni e cluster con k=4 (fascia 18-21).

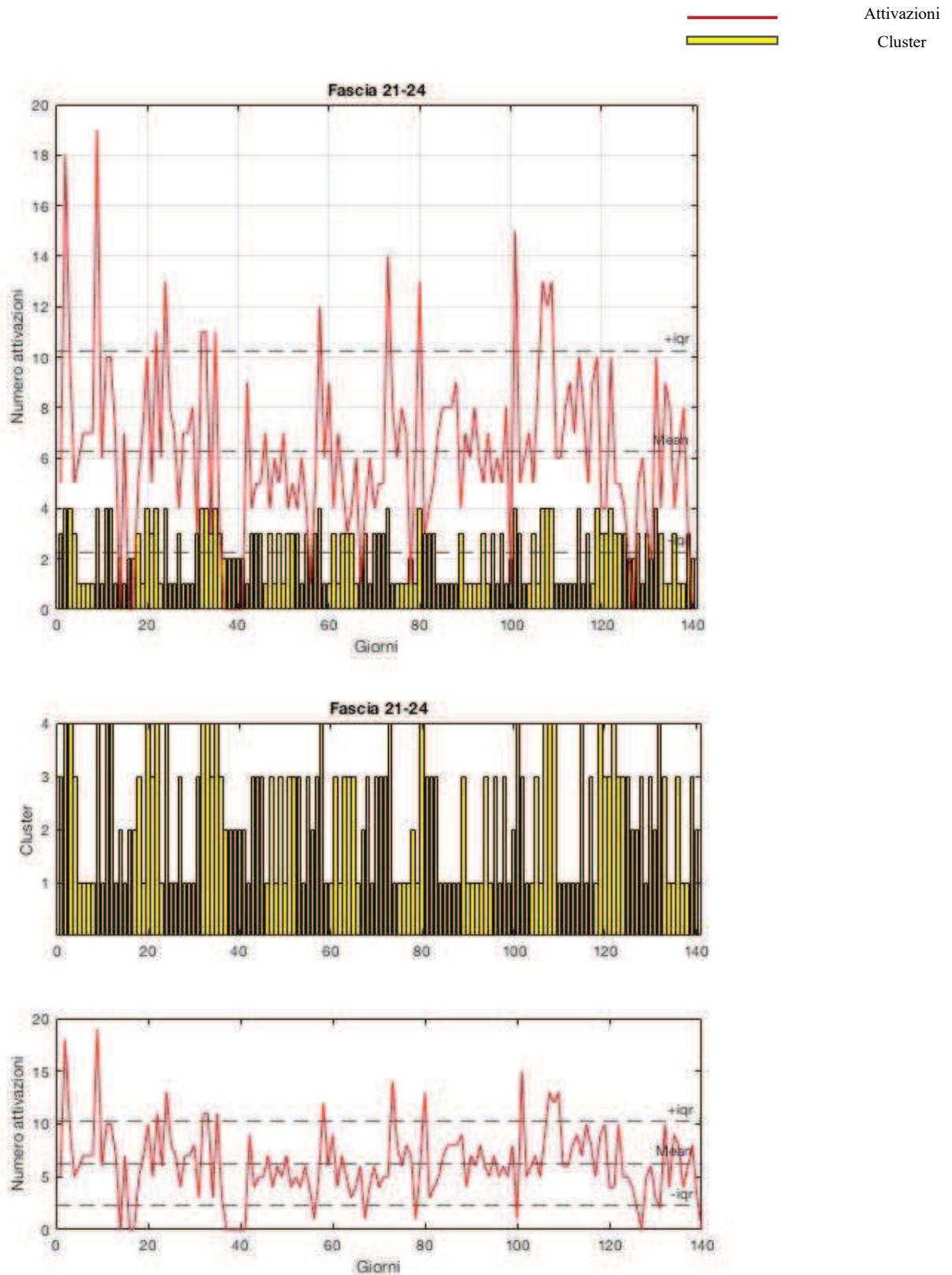


Figura 32: Grafici attivazioni e cluster con $k=4$ (fascia 21-24).

5.4 Studio preliminare del significato dei cluster e verifica

In questo paragrafo vengono riportate tabelle che contengono quelli che sono i significati ricavati dall'osservazione dei risultati dell'apprendimento automatico non supervisionato combinati con gli indici di dispersione.

Inoltre, vengono esposte le tabelle necessarie tanto alla determinazione dei concetti alla base di ogni cluster quanto alla loro verifica.

FASCE ORARIE	MEDIA	STANDARD DEVIATION	SCARTO INTERQUARTILE
0-3	0.4	1.0	0
3-6	0.4	1.0	0
6-9	8.8	4.8	7
9-12	19.7	9.2	14
12-15	15.9	9.9	15
15-18	14.3	7.9	10.5
18-21	17.3	9.8	15
21-24	6.2	3.5	4

Tabella 3: Indici di dispersione per ogni fascia.

FASCE ORARIE	CLUSTER 1	CLUSTER 2	CLUSTER 3
0-3	<media	>media	>>media
3-6	>>media	<media	>media
6-9	<media+iqr, >media	<media, <media-iqr	>media+ iqr
9-12	Nel range	<media, <media-iqr	>media+iqr, >media& <media+iqr
12-15	<<media	Intorno media+iqr	Nel range
15-18	>media+iqr	<media	Nel range
18-21	Nel range	<media	Intorno media+iqr
21-24	Intorno media+iqr	<media-iqr	Nel range

Tabella 4: Significati dei cluster per ogni fascia con k=3.

MEDIE FASCE ORARIE	CLUSTER 1	CLUSTER 2	CLUSTER 3
0-3	0	3.3	1.7
3-6	3.1	0,1	2
6-9	11.6	5.9	16.6
9-12	30.9	9	19.6
12-15	5.9	28.7	15.6
15-18	26.5	7.3	15.6
18-21	19	7.7	31.4
21-24	10.3	1.3	5.3

Tabella 5: Medie del numero delle attivazioni all'interno di ogni fascia, calcolata sulla base del cluster di appartenenza (k=3).

FASCE ORARIE	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
0-3	<media	>>media	>>media	>>>media
3-6	>>>media	<media	>>media	>media
6-9	<media+iqr&>media	>media-iqr&<media	>media+iqr	<media
9-12	Intorno media+iqr	<media& >media-iqr	>media& <media+iqr	<media-iqr
12-15	<media	<media+iqr&>media	>media-iqr &<media	>media+iqr
15-18	>media+iqr	>media-iqr &<media	<media+iqr&>media	<media-iqr
18-21	Nel range	<media	>media	>media+iqr
21-24	<media+iqr&>media	<media-iqr	<media& >media-iqr	>media+iqr

Tabella 6: Significati dei cluster per ogni fascia con k=4.

MEDIE FASCE ORARIE	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
0-3	0	3	1	5
3-6	3.1	0	2	1
6-9	11.9	7	16.6	3.1
9-12	30.9	11.3	20.2	3.7
12-15	5.9	26	15.5	39.5
15-18	26.5	9.5	17	3.7
18-21	17.6	6.9	29.2	46
21-24	7.1	1.3	4.3	12

Tabella 7: Medie del numero delle attivazioni all'interno di ogni fascia, calcolata sulla base del cluster di appartenenza (k=4).

6 Conclusioni e sviluppi futuri

In questo lavoro di tesi si è cercato di produrre metodi per la valutazione delle attività indoor di utenti anziani attraverso l'individuazione di anomalie nei modelli comportamentali presi in considerazione, avvalendosi di analisi di tipo statistico e dell'apprendimento automatico non supervisionato all'interno della piattaforma di calcolo Matlab.

Dai risultati delle analisi statistiche, condotte sui dati della residenza 239, e dalla loro rappresentazione grafica si possono trarre le seguenti osservazioni:

- Le attività legate ai bisogni fisiologici dell'utente risultano le più rilevanti dal confronto dei valori percentuali calcolati. Si verificano con una leggera regolarità durante le giornate del periodo di osservazione, si nota una maggiore concentrazione all'interno delle fasce orarie "Mattina" e "Pomeriggio".
- Le attività legate all'apertura della porta di casa che eventualmente indicano anche visite di persone esterne si riscontrano in maniera quasi costante durante il periodo di osservazione, in particolare, all'interno delle fasce "Mattina" e "Pomeriggio". Risultato che potrebbe essere collegato al fatto che l'utente abita nello stesso palazzo della figlia e spesso vanno nel suo appartamento.

Dallo studio dei casi di normalità e anormalità, focalizzandosi sul tratto finale del periodo di osservazione (marzo, aprile), si riscontrano risultati che potrebbero essere collegabili alle restrizioni conseguenti al lockdown per la pandemia di COVID-19:

- Per le attività legate all'uscita/entrata del soggetto si evidenziano valori sotto al range di normalità.
- Per le attività legate al sonno dell'utente si verificano valori sopra l'intervallo di normalità.

Il lavoro svolto non ha tenuto conto dei margini di errore associabili alla base di partenza statistica e alla variabilità delle attività rilevate. Pertanto, questo implica un limite che potrebbe essere superato dallo sviluppo di lavori futuri.

Dai risultati delle analisi statistiche si evince che la progressiva suddivisione del periodo di campionamento da intero a fasce orarie ha permesso di individuare in maniera gradualmente dettagliata i livelli di attività dell'utente durante il periodo di osservazione. Tale suddivisione potrebbe essere applicata da approfondimenti futuri per l'analisi di normalità e anormalità.

Inoltre, occorre rimarcare che lo studio dei risultati dell'apprendimento automatico non supervisionato deve essere condotto da sviluppi futuri, allo scopo di determinare in maniera più accurata e consistente i significati degli indici dei cluster.

Per concludere, i metodi ottenuti da questo lavoro di tesi vanno inseriti nell'ambito del monitoraggio del comportamento dei soggetti anziani. Approcci di questo tipo mirano a supportare il lavoro degli assistenti informali e formali e del personale sanitario, che si occupano di soggetti anziani affetti da demenza senile, attraverso la produzione di suggerimenti significativi al contesto in seguito all'individuazione di anomalie dei modelli comportamentali degli utenti.

Bibliografia

- [1] Majumder S, Aghayi E, Noferesti M, et al. *Smart Homes for Elderly Healthcare—Recent Advances and Research Challenges*. Published online 201.
- [2] Jinkook Lee, Samuel Lau, Erik Meijer, Peifeng Hu. *Living longer, with or without disability? A Global and Longitudinal Perspective*.
- [3] (World Health Organization, 2019).
- [4] Arshad Yahya, Mina Chandra*, Kuljeet Singh Anand, Jyoti Garg. Behavioral and psychological symptoms in dementia and caregiver burden. *Behavioral and psychological symptoms in dementia and caregiver burden*. Published online February 6, 2015.
- [5] Henk Herman Nap PhD. AAL-2016-eWARE-Part-B.
- [6] Shengjun Xiao, Linwang Yuan, Wen Luo, Dongshuang Li 4, Chunye Zhou and Zhaoyuan Yu. *Recovering Human Motion Patterns from Passive Infrared Sensors: A Geometric-Algebra Based Generation-Template-Matching Approach*. Published online December 3, 2019.
- [7] Casaccia S, Revel GM, Scalise L, et al. *Social Robot and Sensor Network in Support of Activity of Daily Living for People with Dementia*. In: Brankaert R,

IJsselsteijn W, eds. *Dementia Lab 2019. Making Design Work: Engaging with Dementia in Context*. Vol 1117. Communications in Computer and Information Science. Springer International Publishing; 2019:128-135. doi:[10.1007/978-3-030-33540-3_12](https://doi.org/10.1007/978-3-030-33540-3_12)

[8] MathWorks, "MATLAB," [Online]

[9] Ethem Alpaydin. *Introduction to Machine Learning*. Cambridge, Massachusetts: The MIT Press, 2020. (pp. 1-2).

[10]. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, Second Edition, 2018. (pp. 3-12).

[11]. Giuseppe Bonaccorso. *Machine Learning Algorithms*. Packt Publishing Ltd, 24 lug 2017. Packt Publishing Ltd, 24 lug 2017 (pp. 59-67).

[12] Joaquín Pérez Ortega¹, Ma. Del Rocío Boone Rojas,^{1,2} María J. Somodevilla García². *Research issues on K-means Algorithm: An Experimental Trial Using Matlab*.