UNIONE EUROPEA

FONDO EUROPEO DI
SVILUPPO REGIONALE

Ministero Sviluppo Economico

REGIONE PUGLIA
DIPARTIMENTO SVILUPPO ECONOMICO,
INNOVAZIONE, ISTRUZIONE,
FORMAZIONE E LAVORO

TR 5.3

ANALISI, INTERPRETAZIONE DEI DATI SPERIMENTALI E MIGLIORAMENTO

INDICE

INDICE 2

1. STORIA DEL DOCUMENTO 3
2. INTRODUZIONE 4
3. MANUTENZIONE RAPIDA ED AFFIDABILE 6
   - 3.1 ANALISI PER L'INDAGINE PER LA MANUTENZIONE RAPIDA ED AFFIDABILE 6
   - 3.2 CONCLUSIONI E RIFERIMENTI LETTERARI 8
4. COMPOSIZIONE DEI SERVIZI 10
   - 4.1 ANALISI TPR – RUN 20 10
5. SPECIFICAZIONE DEI SERVIZI 17
   - 5.1 ANALISI TPR RUN 30 – RUN 31 – RUN 32 18
6. RICERCA DI UN SERVIZIO 29
   - 6.1 ANALISI TPR – RUN 40 – RUN 41 – RUN 42 30
7. ANALISI CES 39
8. USABILITÀ E UX DEL PORTALE 41
   - 8.1 ANALISI PER L'INDAGINE DI USABILITÀ DEL PORTALE 41
     - *8.1.1 Usabilità della Composizione dei Servizi 41*
     - *8.1.2 Usabilità della Specificazione dei Servizi 45*
     - *8.1.3 Usabilità della Ricerca dei Servizi 49*
     - *8.1.4 Analisi dell'indice SUS nelle tre sperimentazioni 52*
   - 8.2 ANALISI PER L'INDAGINE DI UX DEL PORTALE 54
9. SECURITY TESTING 60
   - 9.1 DISTRIBUZIONE DELLE VULNERABILITÀ PER AGING 60
   - 9.2 DISTRIBUZIONE DELLE VULNERABILITÀ CATALOGATE PER CLASSE 63
   - 9.3 DISTRIBUZIONE DELLE VULNERABILITÀ PER LAYER 66
   - 9.4 INIZIATIVE DI MIGLIORAMENTO 67
10. SICUREZZA CON LA BLOCKCHAIN 68
    - 10.1 RISULTATI 68
    - 10.2 ANALISI DEI RISULTATI 69

1. STORIA DEL DOCUMENTO

DATA	AUTORE	VERSIONE	CONTENUTI & MODIFICHE

2. INTRODUZIONE

L'analisi dei dati ottenuti durante l'esperimento suggerisce che la piattaforma web in questione potrebbe beneficiare di miglioramenti per migliorare la fruizione del servizio e ridurre la variabilità nei tempi di completamento del task.

Tra le varie repliche sperimentali, quindi, sono stati più uprade alla piattaforma I-Mole.

A titolo esemplificativo si segnala:

- La progettazione ed implementazione della Dashboard per il Consumer e per il Vendor (vedi Modifica 7 del TR. 5.1), con cui accedendo alla sezione Dashboard, è possibile avere accesso a report real-time raggruppati in "Ordini" e in "Servizi" di trasporto, magazzino e digitali.
- La progettazione ed implementazione servizio che consente la visibilità, in tempo reale, dello stato per ogni tipo di servizio. Ogni servizio ha una dashboard dedicata (vedi Modifica 7 del TR. 5.1).
- La progettazione ed implementazione della sezione "Consulenza" al fine di permettere, sia al Consumer che al Vendor, di richiedere una consulenza diretta con un'Analyst (vedi Modifica 13 del TR 5.1).
- Che la dashboard del punto precedente è stata ulteriormente ampliata introducendo l'integrazione tramite API della piattaforma Exentriq con cui è possibile avere accesso a report real-time raggruppati in Dashboard. Il vantaggio derivato è la visibilità su OnTime, KPI, Sentiment Analysis (vedi Modifica 8 del TR. 5.1).
- Che la dashboard dei punti precedenti è stata ulteriormente ampliata introducendo un algoritmo per l'analisi della distribuzione dei dati fornita dai KPI della piattaforma I-Mole (vedi Modifica 9 del TR. 5.1).
- Che durante la sperimentazione i tutor hanno potuto constatare che il processo di sottoscrizione dei contratti e di validazione degli ordini rappresentava un passaggio migliorabile, a detta dei soggetti sperimentali. A tal proposito è stato progettato ed implementato un meccanismo di autenticazione basato su OTP (vedi Modifiche 10 e 11 del TR 5.1).

Visti i risultati quantitativi della sperimentazione migliorati nelle varie repliche sperimentali anche in seguito ai miglioramenti citati e raccolti ulteriori feedback da parte dei soggetti sperimentali, sono in fase di valutazione ulteriori funzionalità da integrare in futuro nella piattaforma I-Mole e che allo stato attuale sono inserite nel backlog.

Alcune di queste funzionalità, di seguito indicate, sono attualmente in fase di valutazione, progettazione al fine di una loro implementazione del breve periodo.

Altre estensioni, sicuramente più impattanti, si è deciso di rimandarle nel medio periodo, ovvero nella fase di industrializzazione del progetto I-Mole. Tra le più significative si segnalano le seguenti:

- introduzione di una dashboard di Business Intelligence che andrebbe ad affiancare quelle implementate durante questo progetto. Tale sistema di Decision Support System diventerebbe un plus per utenti premium. La stella piattaforma sarà implementata con PowerBi di Microsoft oppure con tool similari, magari open source, come Knime.
- semplificazione del workflow ordini che non necessitano di personalizzazioni ad hoc.
- Ottimizzazione dell'interfaccia utente per rendere più intuitiva e user-friendly la navigazione sulla piattaforma.

Possibili azioni correttive potrebbero includere:

- Fornire training o supporto agli utenti per migliorare la loro familiarità con la piattaforma e aumentare l'efficienza nell'esecuzione dei compiti.

3. MANUTENZIONE RAPIDA ED AFFIDABILE

3.1 Analisi per l'indagine per la Manutenzione Rapida ed Affidabile

La tabella seguente riassume le metriche del debito tecnico per diverse versioni della piattaforma Imole, analizzando sia il front-end che il back-end. Le metriche considerate sono:

- **Bug**: Numero di errori rilevati nel codice
- **Vulnerabilità**: Presenza di potenziali falle di sicurezza.
- **Security Hotspots**: Punti critici che potrebbero essere sfruttati per attacchi informatici.
- **Code Smells**: Pratiche di codifica non ottimali che possono rendere il codice più difficile da leggere e mantenere.
- **Duplicazione**: Percentuale di codice duplicato.
- **Loc (K)**: Numero di migliaia di righe di codice.
- **Debito**: Tempo stimato per risolvere il debito tecnico.
- **Maintainability**: Manutenibilità del codice (da A a E, dove A è la migliore).
- **Security**: Sicurezza del codice (da A a E, dove A è la migliore).
- **Testbook Runned**: Numero di test eseguiti.
- **Test Cases**: Numero di casi di test (eseguiti/totali).
- **Time Spent**: Tempo impiegato per la realizzazione dei task previsti dalla release

	v0.6.6		v0.7.0		v0.7.5		v0.8.0		v0.8.1		v0.8.3	
	Backend	Frontend	Backend	Frontend	Backend	Frontend	Backend	Frontend	Backend	Frontend	Backend	Frontend
Bugs	67	26	27	28	27	29	44	29	54	32	**0**	**0**
Vulnerabilities	0	0	0	0	0	0	0	0	0	0	0	0
Security Hotspots	4	5	5	6	16	3	30	5	29	5	13	4
Code Smells	564	139	688	159	392	157	586	232	647	236	259	143
Duplications	5.4%	6.3%	4.5%	8.8%	4.7%	10.5%	3.8%	10.9%	3.7%	11.1%	3.7%	11.3%
Loc (K)	11	19	13	20	15	19	23	26	25	29	25	29
Debt	5d 3h	1d 5h	6d 4h	2d 2h	4d 7h	2d	7d 1h	2d 7h	7d 7h	2d 7h	**3d 4h**	**2d 0h**
Maintainability	A	A	A	A	A	A	A	A	A	A	A	A
Security	A	A	A	A	A	A	A	A	A	A	A	A
Testbook Runned	3		3		2		4		2		3	
Test Cases (Runned/Total)	287/219		324/285		307/298		434/307		318/312		376/324	
Test Time Spent	548h		1002h		1296h		1040h		400h		440h	

Figura 3.1 - Metriche debito tecnico

3.1.1 Analisi dell'andamento

- Bug: Il numero di bug varia considerevolmente tra le versioni: nelle versioni più vecchie erano in numero piuttosto elevato, salvo diminuire con le release successive. L'ultima release oggetto di analisi, vede un totale azzeramento dei bugs rilevati.
- Vulnerabilità: Non sono state identificate vulnerabilità in nessuna versione, suggerendo un buon controllo sulla sicurezza del codice.
- Security Hotspots: Un aumento significativo si osserva nella v0.8.0 dovuto all'integrazione di un sistema esterno, seguito da un calo nelle versioni successive.
- **Code Smells**: Il numero di code smells è generalmente alto, ma una drasitca riduzione si è messa in atto nella versione v0.8.3.
- **Duplicazione**: La duplicazione di codice si mantiene relativamente stabile, con un leggero aumento nelle ultime versioni.
- **Loc (K)**: Il numero di righe di codice cresce con l'avanzamento delle versioni, sintomo di una complessità crescente della piattaforma.
- **Debito**: Il tempo stimato per risolvere il debito tecnico è stato in generale in calo, salvo subire una inversione a ridosso della versione v0.8.0 e successiva, imputabile all'integrazione di componenti esterne (Exentriq e Blockchain) che hanno segnato cambiamenti tecnologici importanti. Le versioni successive hanno visto un importante consolidamento (in positivo) del debito.
- **Maintainability**: La manutenibilità del codice rimane alta (A) per tutte le versioni.
- **Security**: La sicurezza del codice rimane alta (A) per tutte le versioni.
- **Testbook Runned**: Il numero di test eseguiti varia, con un picco nella v0.8.0 che ha richiesto affinamenti ulteriori prima di raggiungere una condizione stabile.
- Test Cases: Il rapporto tra test eseguiti e totali è generalmente buono, con un leggero calo nelle ultime versioni, sintomo della maggiore stabilità del prodotto.
- Test Time Spent: Il tempo impiegato per eseguire i test aumenta con l'aumentare del numero di test.

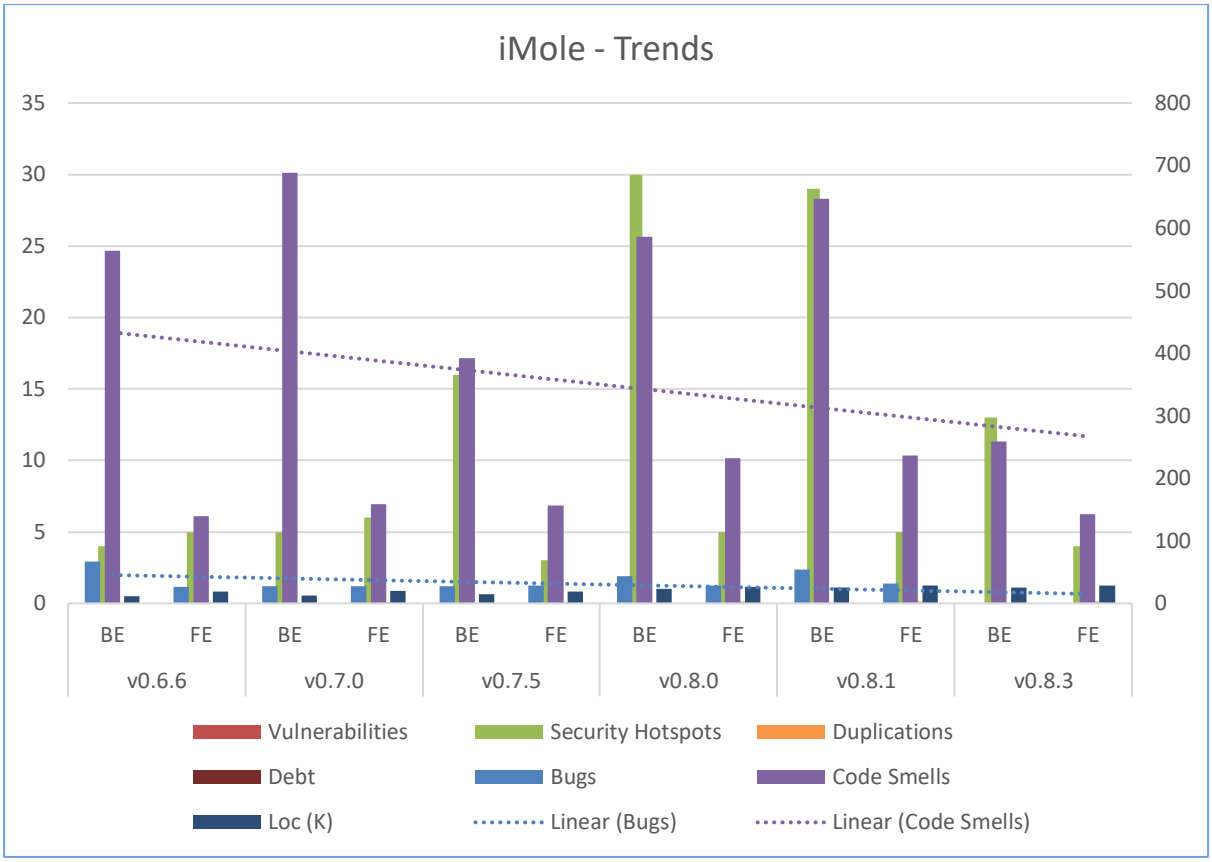


Figura 3.2 - I-Mole Trends

3.2 Conclusioni e Riferimenti Letterari

L'analisi del debito tecnico condotta evidenzia un quadro complesso, con alcune metriche in aumento (come il numero di bug e code smells) e altre stabili o in leggero calo (come la duplicazione di codice). In letteratura, l'ingegneria del software offre diversi spunti per interpretare questa situazione.

3.2.1 L'impatto dei cambiamenti tecnologici

Secondo studi come "**Technical Debt in Agile Software Development**" di Adam S. Oentaryo et al., l'introduzione di cambiamenti tecnologici significativi può rendere difficoltoso il controllo delle metriche del debito tecnico. L'aumento di bug e code smells osservato nelle ultime versioni potrebbe essere associato all'adozione di nuove tecnologie o all'evoluzione del framework utilizzato.

3.2.2 Risanamento del debito tecnico

Tuttavia, la letteratura evidenzia anche la possibilità di risanare il debito tecnico in release successive. Il libro "**Managing Technical Debt**" di Thomas A. Limon et al. descrive diverse strategie per affrontare il debito tecnico, come la refactoring del codice e l'implementazione di test automatici. L'andamento del debito tecnico nel caso analizzato suggerisce che il team di sviluppo sta adottando alcune di queste strategie, con un conseguente miglioramento della stabilità e manutenibilità del codice nelle release successive.

3.2.3 In ottica futura:

Per mantenere il debito tecnico sotto controllo, è fondamentale:

- **Monitorare costantemente le metriche** del debito tecnico e identificare le aree di attenzione.
- **Implementare best practice di sviluppo software** per ridurre la generazione di nuovo debito tecnico.
- **Pianificare attività di refactoring** per risanare il debito tecnico esistente.
- **Coinvolgere il team di sviluppo** in discussioni sul debito tecnico e sulle strategie per ridurlo.

L'analisi del debito tecnico è un processo continuo che richiede attenzione e impegno da parte di tutto il team di sviluppo. Investendo tempo e risorse per ridurre il debito tecnico, si otterranno benefici a lungo termine in termini di manutenibilità, sicurezza e prestazioni dell'applicazione.

4. COMPOSIZIONE DEI SERVIZI

Così come definitivo e progettato nell'OR 4, l'obiettivo di ricerca di questo caso di studio è il seguente:

> Analizzare Piattaforma I-Mole al fine di valutarla rispetto alla complessità di combinare servizi primi già esistenti dal punto di vista della produzione di un servizio da proporre a catalogo

L'indagine è stata eseguita da 15 Soggetti Sperimentali attraverso l'esecuzione del RUN20.

Durante questa indagine sono state rilevate le seguenti misure:

- **TPR** Tempo Persona Rilevato durante l'esecuzione del task assegnato;
- **CES** Indice del questionario CES sottoposto ai soggetti sperimentali alla fine dell'esecuzione del task ed i cui risultati sono mostrati nell'OR 5.2.

L'analisi mira a esaminare i risultati emersi dall'esperimento condotto al fine di valutare la piattaforma I-Mole. Partendo dai dati raccolti, si offre un'analisi dettagliata delle principali osservazioni emerse durante il corso della sperimentazione.

4.1 Analisi TPR – RUN 20

L'obiettivo di questa macroarea è stato quello di validare la complessità di combinare servizi già esistenti per produrre un servizio da proporre a catalogo.

Durante l'esperimento un tutor indicato da Assioma ha rilevato il valore del **TPR (Tempo Persona Rilevato)** per ogni soggetto sperimentale impegnato nell'esecuzione del task assegnato.

Inoltre, a priori è stato definito il **Tempo Persona di Confronto (TPC)** che rappresenta il valore soglia che gli esperti di dominio (esperti di Assioma che padroneggiano la piattaforma IMOLE nonché il dominio della logistica) hanno

stabilito come accettabile per completare il task. Per questo esperimento, il TPC è pari a 350 secondi, che corrisponde a 5 minuti e 50 secondi.

A fine esperimento sono stati determinati i seguenti indici:

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 347 secondi, che corrisponde a 5 minuti e 47 secondi.

Scarto percentuale: Lo scarto percentuale indica quanto i tempi persona rilevati si discostano dal valore di soglia. Esso si determina con la seguente formula (TPC - TPR) / TPC) e nell'esperimento RUN20 è risultato pari a 0,86%. Questo indica che la media dei tempi persona rilevati è leggermente inferiore al valore soglia, il che suggerisce che in media i soggetti hanno completato il task in un tempo leggermente più breve rispetto a quanto previsto dagli esperti.

Deviazione standard: La deviazione standard è una misura di dispersione che indica quanto i dati sono distribuiti intorno alla media; ovvero, rappresenta la variabilità dei dati rispetto alla media. Una deviazione standard più alta indica una maggiore dispersione dei dati rispetto alla media, mentre una deviazione standard più bassa indica una maggiore coerenza o omogeneità nei dati. Una deviazione standard relativamente alta può avere implicazioni pratiche importanti. Ad esempio, potrebbe indicare che alcuni soggetti completano il task in tempi molto più brevi rispetto ad altri, suggerendo la presenza di potenziali fattori di confusione, in senso negativo, o differenze individuali nella capacità di eseguire il task. In questo esperimento, la deviazione standard è risultata pari a 32,92 secondi; tale valore, non certamente alto, indica, dunque, che i tempi di completamento del task tra i soggetti sperimentali variano di circa 32,92 secondi rispetto alla media di 347 secondi (5 minuti e 47 secondi). Tale indice è ritenuto positivo dagli sperimentatori visto che in letteratura è indicato il 20% come valore di accettazione.

Varianza: La varianza è il quadrato della deviazione standard e rappresenta la media dei quadrati delle differenze tra ogni valore e la media. Una varianza maggiore indica una maggiore dispersione dei dati rispetto alla media. Essa è risultata pari a 1161 secondi al quadrato.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il CV si calcola come il rapporto tra la deviazione standard

e la media dei TPR. Un CV basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare. Nel RUN 20 tale valore è pari a 9,49.

I risultati, dunque, mostrano che, in media, i soggetti sperimentali hanno completato il task entro il tempo soglia stabilito dagli esperti di dominio; infatti, è emerso un tempo medio leggermente inferiore rispetto al valore previsto. Tuttavia, c'è una certa variabilità nei tempi di completamento, come indicato dalla deviazione standard e dalla varianza, il che suggerisce che alcuni soggetti hanno impiegato più tempo rispetto ad altri per completare il task.

In tali casi è opportuno valutare se e quali classi di soggetti sperimentali hanno evidenziato eventuali problemi specifici che potrebbero influenzare il buon funzionamento della piattaforma web.

Come si può notare dai grafici che seguono, in questo risultato incide sicuramente la classe dei soggetti sperimentali “Principianti” e “Principianti Mobile” che mediamente hanno impiegato un tempo superiore.

In tali casi, inoltre, la letteratura consiglia di verificare l’Usabilità e User Experience (UX) per valutare l'usabilità e l'esperienza utente con la piattaforma web per identificare eventuali ostacoli o difficoltà che potrebbero rallentare i soggetti nel completare il task; nonché, raccogliere il feedback dei partecipanti al fine di ottenere il parere dei soggetti sperimentali sulla loro esperienza durante il completamento del task ed identificare, così, eventuali problemi specifici che potrebbero influenzare i tempi di completamento e guidare gli interventi di miglioramento. Cosa puntualmente fatta durante questa sperimentazione ed i cui risultati sono pubblicati nel proseguo di questo documento.

Il risultato ottenuto nel RUN 20 suggerisce che i dati sono relativamente consistenti e ciò soddisfa appieno il gruppo impegnato nello stesso.

A rafforzare tale convinzione contribuiscono anche i 2 grafici che seguono, fortemente consigliati dalla letteratura, per la validazione dei risultati degli esperimenti di questo tipo; ovvero, il grafico a dispersione dei tempi rilevati durante l’esecuzione dei task assegnati ai 15 soggetti sperimentali e l’istogramma che illustra la media dei tempi per classe di soggetti sperimentali.

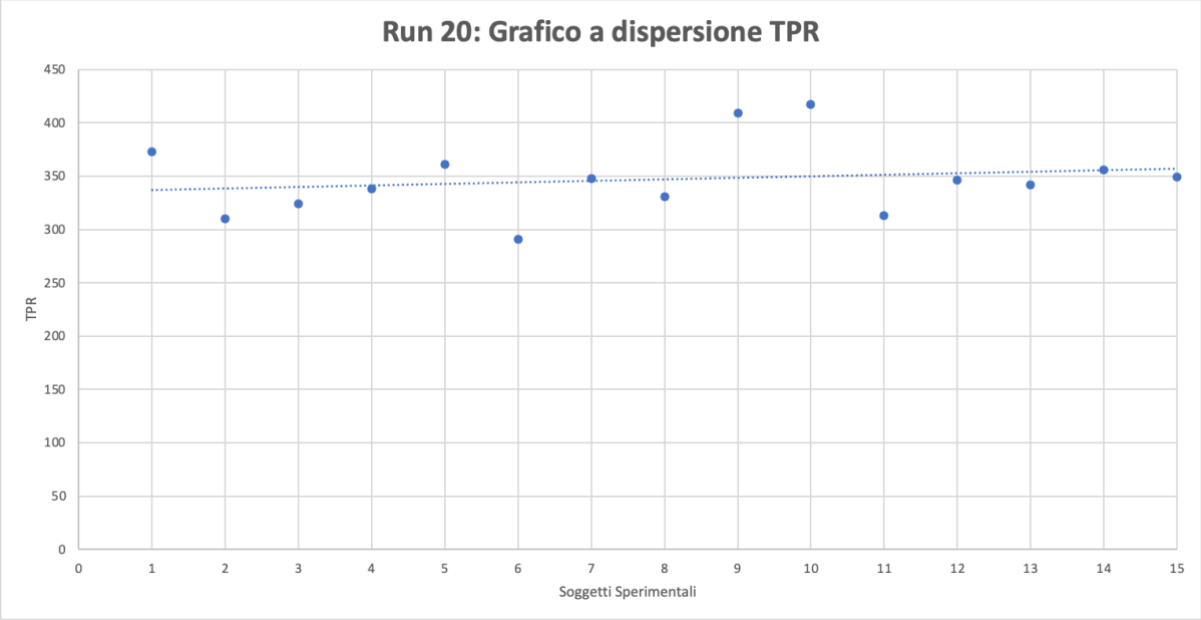


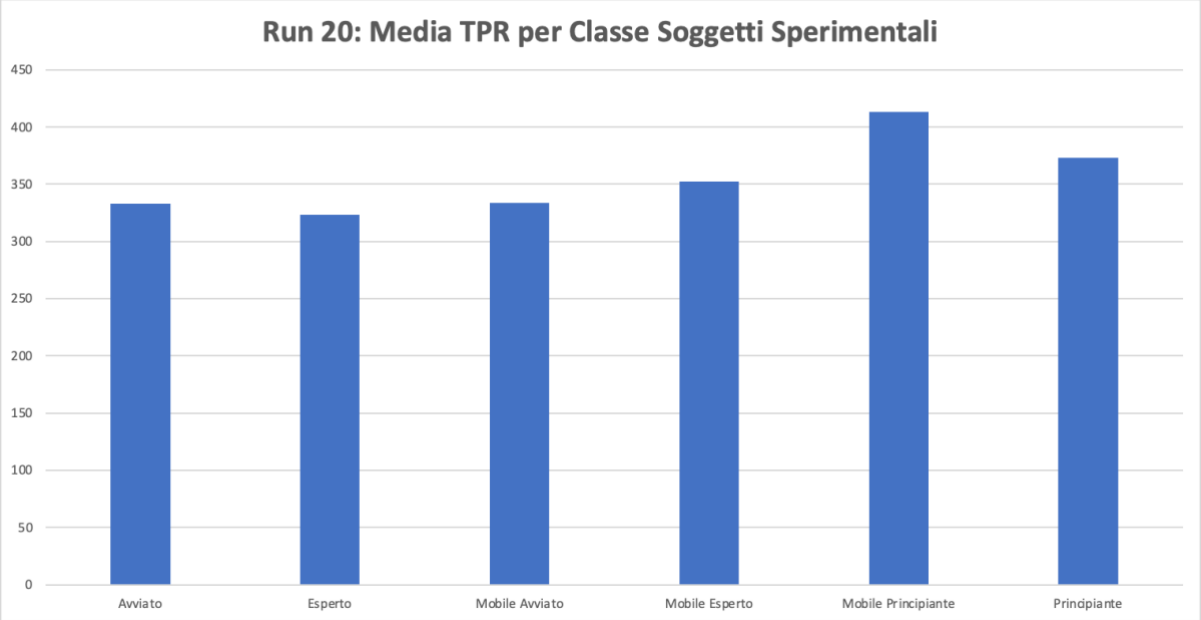
Figura 4.1 – Grafico a dispersione TPR

Dall'analisi del grafico a dispersione dei TPR emerge in tutta evidenza che i risultati si trovano nell'intorno del valore soglia (TPC). Tale indicazione è assolutamente positiva e suggerisce che, in media, i soggetti sperimentali hanno completato il task con successo ed in tempi coerenti con le aspettative della sperimentazione.

Si può notare, inoltre, che la linea di tendenza del grafico quasi si sovrappone alla linea retta orizzontale del TPC (valore 350) e ciò suggerisce che c'è una buona coerenza nei tempi di completamento del task tra i diversi soggetti sperimentali. Questo indica che la maggior parte dei partecipanti ha completato il task entro un intervallo di tempo simile e vicino al valore soglia.

Nonostante la conformità generale ai tempi soglia, è importante esaminare la dispersione dei dati intorno alla linea di tendenza. Una maggiore dispersione potrebbe indicare una variabilità nei tempi di completamento tra i partecipanti, che potrebbe essere dovuta a differenze individuali nelle capacità cognitive o nelle strategie utilizzate per completare il task. A tal fine è sicuramente utile esaminare le tendenze o pattern nel grafico a dispersione; ovvero, far emergere classi di soggetti sperimentali con prestazioni consistentemente migliori o peggiori rispetto alla media, il che potrebbe suggerire la presenza di fattori o variabili aggiuntive che influenzano i tempi di completamento.

Per tal motivo si è prodotto, dunque, il grafico che segue.



L'analisi dell'istogramma delle medie dei TPR per classe dei soggetti sperimentali evidenzia differenze nei tempi di completamento tra le classi di partecipanti; è di tutta evidenza che i "principianti" hanno impiegato mediamente più tempo nell'esecuzione del task assegnato rispetto agli altri gruppi. Questo suggerisce che ci sono differenze, anche se non sostanziali, nelle prestazioni tra i partecipanti in base al loro livello di esperienza o competenza nel contesto del task. Ciò è dovuto, probabilmente, alla poca familiarità con una piattaforma web come quella di I-Mole e, dunque, richiedono più tempo per comprendere le istruzioni o completare le attività richieste.

Al fine di esaminare eventuali tendenze o pattern all'interno delle varie classi di soggetti sperimentali per far emergere la presenza di ulteriori fattori o variabili che influenzano i tempi di completamento, si è deciso di realizzare un ulteriore istogramma, che segue, dove i soggetti sono stati raggruppati in base al tipo di dispositivo utilizzato durante l'esperimento.

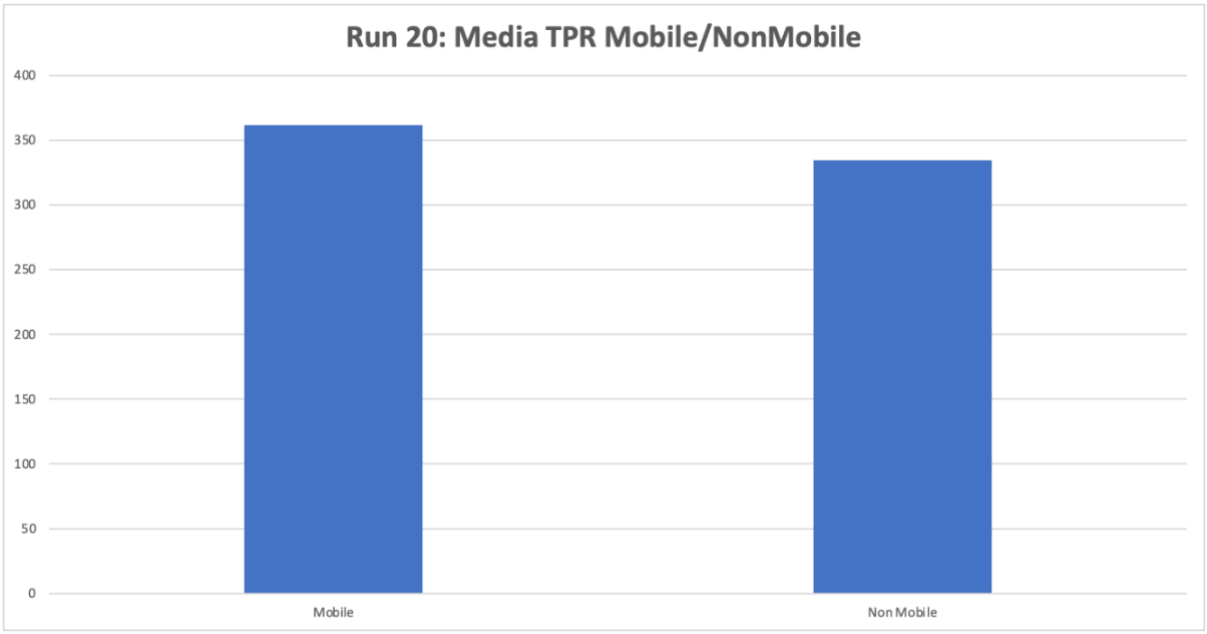


Figura 4.2 – Media TPR Mobile/ Non mobile

L'analisi di quest'ultimo istogramma induce all'osservazione che i soggetti sperimentali che hanno utilizzato un dispositivo "mobile" hanno impiegato mediamente più tempo nell'esecuzione del task assegnato. Questo suggerisce che ci potrebbero essere differenze sostanziali nelle prestazioni tra le varie tipologie di dispositivi utilizzati.

Da ulteriori incontri con il team di Assioma è emersa la natura "responsive" della piattaforma I-Mole, confermata, tra l'altro, dal fatto che tutti i soggetti sperimentali che hanno utilizzato dispositivi "mobile" hanno completato il task assegnatogli raggiungendo l'obiettivo prefissato (anche se tre soggetti sperimentali su sette lo hanno fatto oltre il tempo TPC); dunque, si deduce che la presenza di schermi più piccoli e di tastiera virtuale rallenta l'inserimento dei dati/comandi rispetto a una tastiera fisica su un dispositivo desktop o laptop, ma non in modo significativo.

L'analisi complessiva dei tre grafici presentati conduce ad evidenziare una buona aderenza delle prestazioni dei soggetti sperimentali ai tempi soglia stabiliti, con una coerenza generale nei tempi di completamento del task con una evidente ma fisiologica varianza basata sull'esperienza del soggetto sperimentale e del dispositivo utilizzato dallo stesso.

Tenendo conto di quanto appena affermato e che tutti i 15 soggetti sperimentali hanno completato con successo il task assegnato

nell'esperimento e, dunque, nessuno di loro ha rinunciato a concludere anticipatamente lo stesso senza il raggiungimento dell'obiettivo e tenuto conto che solo 5 su 15 soggetti sperimentali sono andati oltre il TCP (tra l'altro di pochi secondi), si può affermare che tali risultati forniscono una base solida per valutare soddisfacenti le prestazioni della piattaforma I-Mole.

5. SPECIFICAZIONE DEI SERVIZI

Così come definitivo e progettato nell'OR 4, l'obiettivo di ricerca di questo caso di studio è il seguente:

> Analizzare Piattaforma I-Mole al fine di valutarla rispetto alla complessità di specificare il servizio che soddisfa il bisogno da soddisfare tra tutte le alternative che il sistema propone per ogni servizio

L'indagine ha previsto l'esecuzione di 3 RUN sperimentali identificati come RUN30, RUN31 e RUN32.

I soggetti sperimentali interessati a questa indagine sono stati tutti i 45 soggetti sperimentali disponibili; essi sono stati divisi in tre batterie da 15; ogni batteria ha eseguito un RUN.

Durante questa indagine sono state rilevate le seguenti misure:

- TPR Tempo Persona Rilevato durante l'esecuzione del task assegnato;
- CES indice del questionario CES sottoposto ai soggetti sperimentali alla fine dell'esecuzione del task ed i cui risultati sono mostrati nell'OR 5.2.

L'analisi mira a esaminare i risultati emersi dall'esperimento condotto al fine di valutare la piattaforma I-Mole. Partendo dai dati raccolti, si offre un'analisi dettagliata delle principali osservazioni emerse durante il corso della sperimentazione.

Per ogni RUN, in conformità con quanto effettuato nel RUN20, sono stati calcolati i seguenti indici statistici ritenuti particolarmente rilevanti per raggiungere l'obiettivo prefissato.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai soggetti sperimentali per completare il task.

Scarto percentuale: Lo scarto percentuale tra i tempi persona rilevati e il valore soglia fornisce un'indicazione della discrepanza percentuale tra i tempi effettivi di completamento e il tempo massimo accettabile. Un piccolo scarto percentuale indica una buona aderenza al valore soglia. Esso si determina con la seguente formula (TPC - TPR) / TPC).

Deviazione Standard: La deviazione standard è una misura di dispersione che indica quanto i dati sono distribuiti intorno alla media; ovvero, rappresenta la variabilità dei dati rispetto alla media. Una deviazione standard più alta indica una maggiore dispersione dei dati rispetto alla media, mentre una deviazione standard più bassa indica una maggiore coerenza o omogeneità nei dati. Una deviazione standard relativamente alta può avere implicazioni pratiche importanti. Ad esempio, potrebbe indicare che alcuni soggetti completano il task in tempi molto più brevi rispetto ad altri, suggerendo la presenza di potenziali fattori di confondimento, in senso negativo, o differenze individuali nella capacità di eseguire il task.

L'aggiunta dell'analisi della deviazione standard fornisce ulteriori dettagli sulla variabilità dei tempi di completamento del task sulla piattaforma web. Questo parametro è cruciale per comprendere la consistenza delle prestazioni degli utenti e l'affidabilità della piattaforma nel supportare l'esecuzione delle attività.

La deviazione standard è stata calcolata come una percentuale rispetto al Valore Soglia (TPC) e come percentuale dei Tempi Persona Rilevati (TPR). Questo permette di valutare la dispersione dei dati rispetto al tempo medio di completamento del task.

Varianza: La varianza dei tempi persona rilevati fornisce una misura della dispersione dei dati intorno alla media. Una varianza più alta indica una maggiore dispersione dei tempi di completamento; ovvero, indica una maggiore dispersione dei dati rispetto alla media. Una varianza più bassa indica, invece, una maggiore coerenza.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il coefficiente di variazione si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un coefficiente basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare.

5.1 Analisi TPR RUN 30 – RUN 31 – RUN 32

L'obiettivo di questa macroarea dell'indagine è la validazione della complessità di specificare il servizio richiesto dal cliente nella "linea di servizi" che gli propone il sistema.

Durante l'esperimento un tutor indicato da Assioma ha rilevato il valore del **TPR (Tempo Persona Rilevato)** per ogni soggetto sperimentale impegnato nell'esecuzione del task assegnato.

Inoltre, a priori è stato definito il **Tempo Persona di Confronto (TPC)** che rappresenta il valore soglia che gli esperti di dominio (esperti di Assioma che padroneggiano la piattaforma IMOLE nonché il dominio della logistica) hanno stabilito come accettabile per completare il task.

5.1.1 RUN 30

Così come indicato nel paragrafo precedente, di seguito sono indicati i valori degli indici per il RUN 30.

Tempo Persona di Confronto (TPC): Questo valore rappresenta il valore soglia che gli esperti di dominio hanno stabilito come accettabile per completare il task. Per questo esperimento, il TPC è pari a 366 secondi, che corrisponde a 6 minuti e 6 secondi.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 368 secondi, che corrisponde a 6 minuti e 8 secondi. Come si può notare, la media dei Tempi Persona Rilevati (TPR) è superiore al Valore Soglia (TPC) di 2 secondi, indicando un normale andamento nell'esecuzione dei task da parte dei soggetti sperimentali.

Scarto percentuale: Lo scarto percentuale tra i tempi persona rilevati e il valore soglia (TPC - TPR) / TPC), è del -0,55%. Questo indica che la media dei tempi persona rilevati è leggermente superiore al valore soglia, il che suggerisce che in media i soggetti hanno completato il task in un tempo leggermente più alto rispetto a quanto previsto dagli esperti.

Deviazione standard: Nel RUN30 la deviazione standard è risultata pari a 33,54 secondi; tale valore, non certamente alto, indica, dunque, che i tempi di completamento del task tra i soggetti sperimentali variano di circa 33,54 secondi rispetto alla media di 368 secondi (6 minuti e 8 secondi). Tale indice è ritenuto positivo dagli sperimentatori visto che in letteratura è indicato il 20% come valore di accettazione.

Varianza: La varianza è il quadrato della deviazione standard e rappresenta la media dei quadrati delle differenze tra ogni valore e la media. Essa è risultata pari a 1206 secondi al quadrato.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il CV si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un CV basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare. Nel RUN 30 tale valore è risultato pari a 9,12.

Nell'esperimento RUN 30, il risultato ottenuto suggerisce che i dati sono relativamente consistenti e ciò soddisfa appieno il gruppo di lavoro impegnato nella sperimentazione.

5.1.2 RUN 31

Si passa, dunque, ad illustrare i valori quantitativi raccolti e determinati per il RUN 31:

Tempo Persona di Confronto (TPC): Questo valore rappresenta il valore di soglia che gli esperti di dominio hanno stabilito come accettabile per completare il task. Anche per questo esperimento, il TPC è pari a 366 secondi, che corrisponde a 6 minuti e 6 secondi.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 356 secondi, che corrisponde a 5 minuti e 56 secondi. Come si può notare, la media dei Tempi Persona Rilevati (TPR) è inferiore al Valore Soglia (TPC) di 10 secondi, indicando un normale andamento nell'esecuzione dei task da parte dei soggetti sperimentali.

Scarto percentuale: Lo scarto percentuale tra i tempi persona rilevati e il valore soglia (TPC - TPR) / TPC), è del 2,73%. Questo indica che la media dei tempi persona rilevati è leggermente inferiore al valore soglia, il che suggerisce che in media i soggetti hanno completato il task in un tempo leggermente più breve rispetto a quanto previsto dagli esperti.

Deviazione standard: Nel RUN31 la deviazione standard è risultata pari a 30,94 secondi (l'8,45% del TPC); tale valore, non certamente alto, indica, dunque, che i tempi di completamento del task tra i soggetti sperimentali variano di circa 30,94 secondi rispetto alla media di 356 secondi (5 minuti e 56 secondi).

Tale indice è ritenuto positivo dagli sperimentatori visto che in letteratura è indicato il 20% come valore di accettazione.

Varianza: La varianza è il quadrato della deviazione standard e rappresenta la media dei quadrati delle differenze tra ogni valore e la media. Essa è risultata pari a 1026 secondi al quadrato.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il CV si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un CV basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare. Nel RUN 31 tale valore è pari a 8,69.

Nell'esperimento RUN 31, il risultato ottenuto suggerisce che i dati sono relativamente consistenti e ciò soddisfa appieno il gruppo di lavoro impegnato nella sperimentazione.

5.1.3 RUN 32

Si passa, infine, ad illustrare i valori quantitativi raccolti e determinati per il RUN 32:

Tempo Persona di Confronto (TPC): Questo valore rappresenta il valore di soglia che gli esperti di dominio hanno stabilito come accettabile per completare il task. Anche per questo esperimento, il TPC è pari a 366 secondi, che corrisponde a 6 minuti e 6 secondi.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 372 secondi, che corrisponde a 6 minuti e 12 secondi. Come si può notare, la media dei Tempi Persona Rilevati (TPR) è superiore al Valore Soglia (TPC) di 6 secondi, indicando un normale andamento nell'esecuzione dei task da parte dei soggetti sperimentali.

Scarto percentuale: Lo scarto percentuale tra i tempi persona rilevati e il valore soglia (TPC - TPR) / TPC), è del -1,64%. Questo indica che la media dei tempi persona rilevati è leggermente superiore al valore soglia, il che suggerisce che in media i soggetti hanno completato il task in un tempo leggermente maggiore rispetto a quanto previsto dagli esperti.

Deviazione standard: Nel RUN32 la deviazione standard è risultata pari a 36,24 secondi (il 9,90% del TPC); tale valore, non certamente alto, indica, dunque,

che i tempi di completamento del task tra i soggetti sperimentali variano di circa 36,24 secondi rispetto alla media di 372 secondi (6 minuti e 12 secondi). Tale indice è ritenuto positivo dagli sperimentatori visto che in letteratura è indicato il 20% come valore di accettazione.

Varianza: La varianza è il quadrato della deviazione standard e rappresenta la media dei quadrati delle differenze tra ogni valore e la media. Essa è risultata pari a 1407 secondi al quadrato.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il CV si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un CV basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare. Nel RUN 32 tale valore è pari a 9,49.

Nell'esperimento RUN 32, il risultato ottenuto suggerisce che i dati sono relativamente consistenti e ciò soddisfa appieno il gruppo di lavoro impegnato nella sperimentazione.

5.1.4 Conclusioni analisi RUN 30 – RUN 31 – RUN 32

I risultati dei tre RUN sperimentali sono coerenti con gli obiettivi del gruppo di lavoro. A rafforzare tale convinzione contribuiscono anche i 6 grafici che seguono fortemente consigliati dalla letteratura per la validazione dei risultati degli esperimenti di questo tipo; ovvero, tre grafici a dispersione dei tempi rilevati durante l'esecuzione dei task assegnati ai 15 soggetti sperimentali e tre istogramma che illustrano la media dei tempi per classe di soggetti sperimentali.

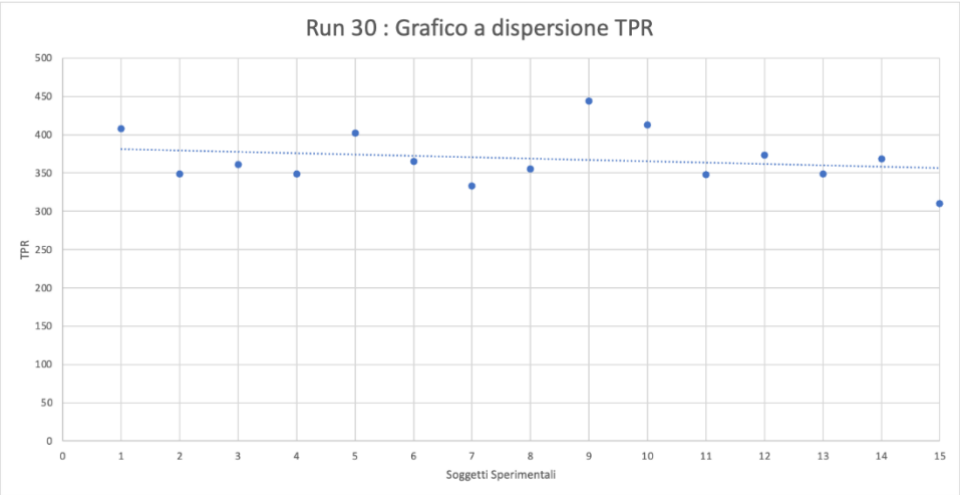


Figura 5.1 - Grafico a dispersione TPR - RUN 30

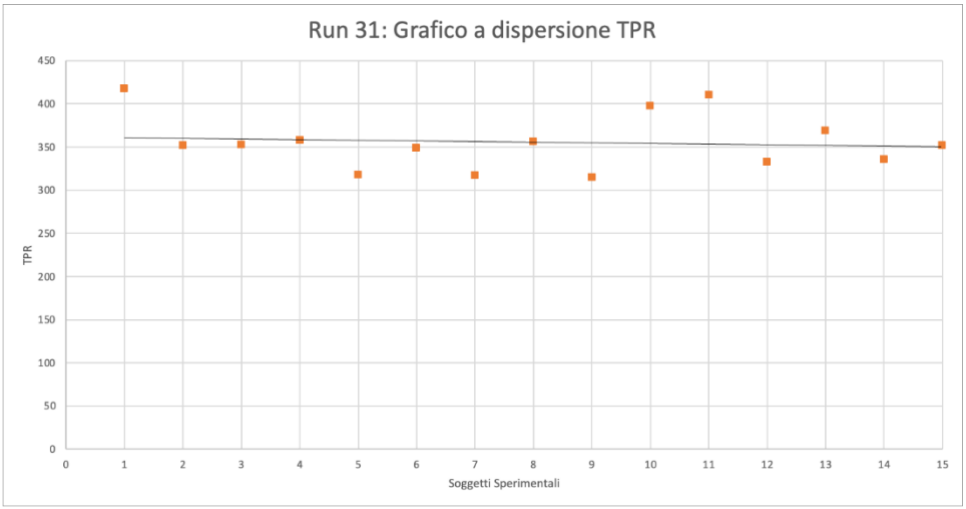


Figura 5.2- Grafico a dispersione TPR - RUN 31

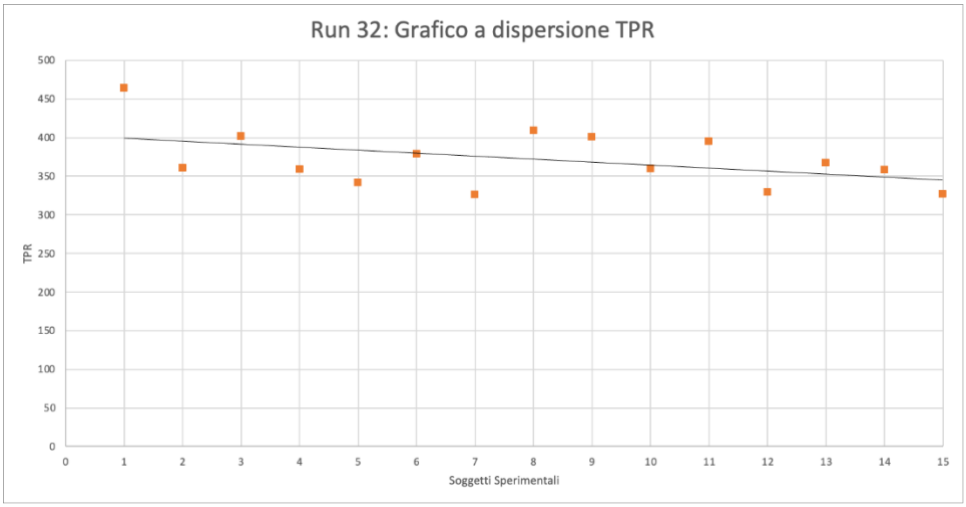


Figura 5.3 – Grafico a dispersione TPR - RUN 32

Dall'analisi del tre grafici a dispersione dei TPR è possibile dedurre diverse osservazioni; ovvero:

- Conformità al valore medio atteso: Il fatto che la linea di tendenza del grafico a dispersione si sovrapponga quasi alla linea retta del valore medio atteso suggerisce che la media dei dati osservati è molto vicina al valore medio atteso o atteso. Questo indica che, in generale, i dati osservati tendono a seguire il valore medio atteso.
- Coerenza dei risultati: La distribuzione dei dati intorno alla linea di tendenza suggerisce una certa coerenza nei risultati. Ciò può essere

interpretato come una stabilità o una consistenza nei tempi di completamento del task sulla piattaforma web tra i diversi partecipanti.

- Possibile aderenza agli standard di performance: Visto che il valore medio atteso (TPC) rappresenta un obiettivo di performance e, dunque, un valore di riferimento, la conformità dei dati osservati a questo valore indica che la piattaforma I-Mole raggiunge gli standard di performance attesi.

Dunque, le osservazioni suggeriscono una certa coerenza e conformità dei dati rispetto al valore TPC, indicando una stabilità e una possibile aderenza agli standard di performance.

Nonostante la conformità generale ai tempi soglia, è importante esaminare la dispersione dei dati intorno alla linea di tendenza. Una maggiore dispersione potrebbe indicare una variabilità nei tempi di completamento tra i partecipanti, che potrebbe essere dovuta a differenze individuali nelle capacità cognitive o nelle strategie utilizzate per completare il task. A tal fine è sicuramente utile esaminare le tendenze o pattern nel grafico a dispersione; ovvero, far emergere classi di soggetti sperimentali con prestazioni consistentemente migliori o peggiori rispetto alla media, il che potrebbe suggerire la presenza di fattori o variabili aggiuntive che influenzano i tempi di completamento. Per tal motivo, e come preannunciato, vengono visualizzati i tre istogramma che illustrano la media dei tempi per classe di soggetti sperimentali.

Figura 5.3 – Media dei tempi RUN30 per classe

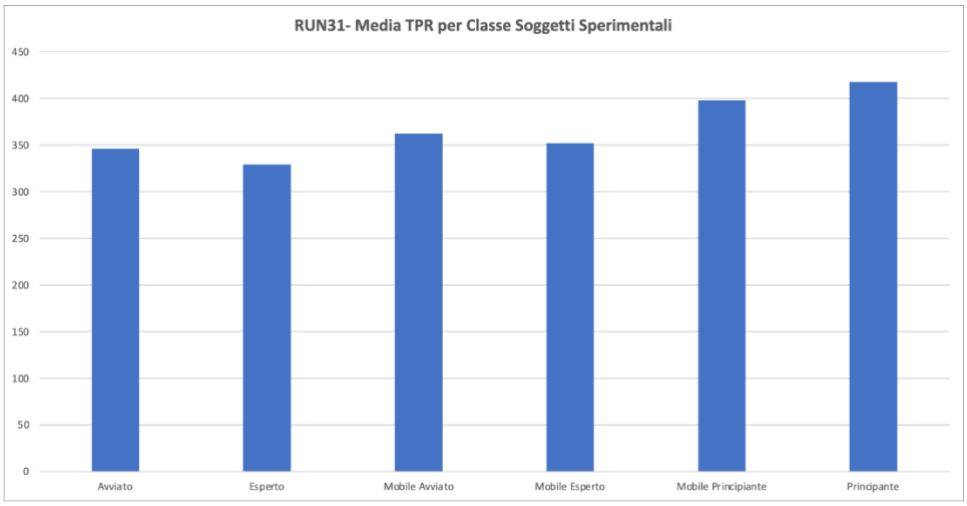


Figura 5.4 - Media dei tempi RUN31 per classe

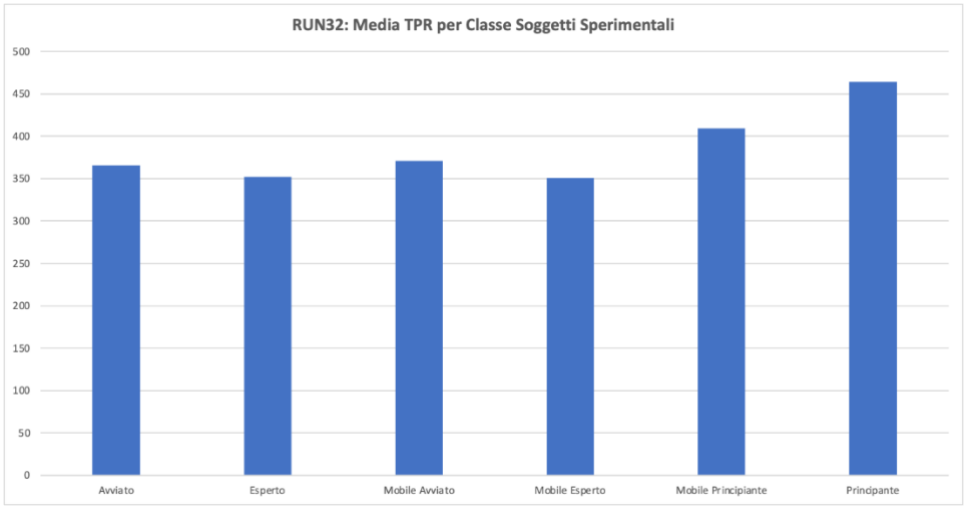


Figura 5.5 - Media dei tempi RUN32 per classe

L'analisi degli istogrammi delle medie dei TPR per classe dei soggetti sperimentali evidenzia differenze nei tempi di completamento tra le classi di partecipanti; è di tutta evidenza che i "principianti" hanno impiegato mediamente più tempo nell'esecuzione del task assegnato rispetto agli altri gruppi. Questo suggerisce che ci sono differenze, anche se non sostanziali, nelle prestazioni tra i partecipanti in base al loro livello di esperienza o competenza nel contesto del task, vista la loro poca familiari con una piattaforma web come quella di I-Mole e, dunque, richiedono più tempo per comprendere le istruzioni o completare le attività richieste.

Le differenze nei tempi di completamento tra i gruppi possono avere importanti implicazioni pratiche per la progettazione della piattaforma web e

per la pianificazione di future sperimentazioni. Ad esempio, potrebbe essere necessario fornire istruzioni più dettagliate o un supporto aggiuntivo per i "principianti" al fine di ridurre i tempi di completamento e migliorare l'esperienza complessiva dell'utente.

La letteratura consiglia, in tali casi, di esaminare eventuali tendenze o pattern all'interno delle varie classi di soggetti sperimentali per far emergere la presenza di ulteriori fattori o variabili che influenzano i tempi di completamento. Per tal motivo, si è deciso di realizzare un'ulteriore tipologia di istogrammi, che segue, dove i soggetti sono stati raggruppati in base al tipo di dispositivo utilizzato durante l'esperimento.

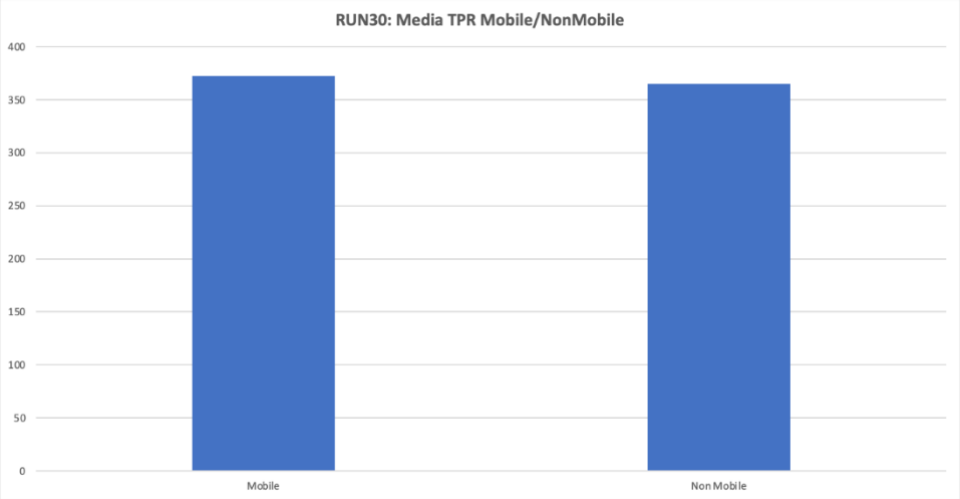


Figura 5.6 - Media TPR Mobile/ Non mobile

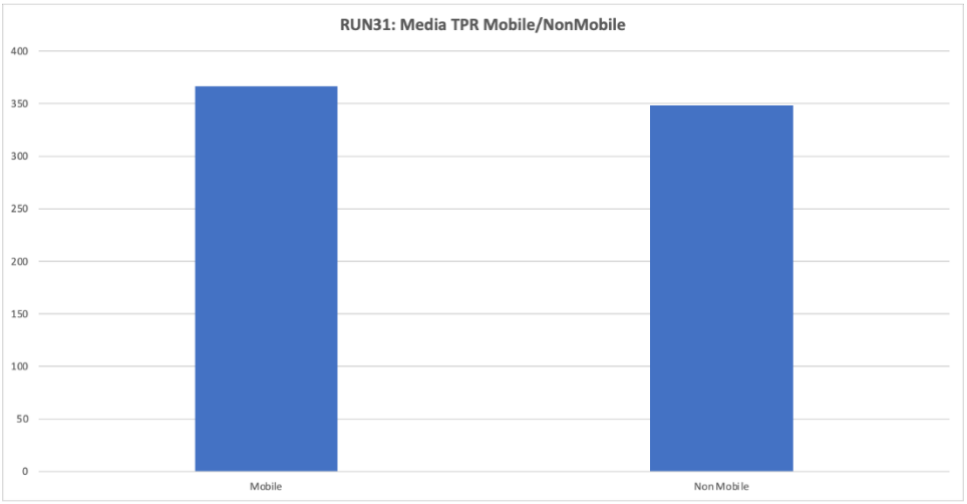


Figura 5.7 - Media TPR Mobile/ Non mobile

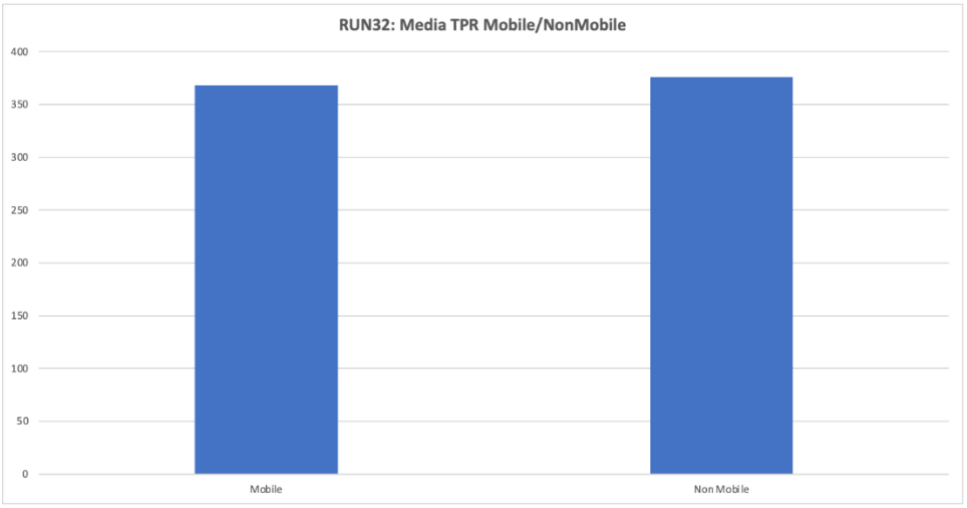


Figura 5.8 - Media TPR Mobile/ Non mobile

L'analisi di questi ultimi istogrammi induce all'osservazione che il tipo di dispositivo utilizzato non incide sulla tempistica di esecuzione di un task. Questo conferma ciò che si è affermato per la sperimentazione precedente laddove si evidenziava la natura "responsive" della piattaforma I-Mole.

L'analisi complessiva dei nove grafici presentati conduce ad evidenziare una buona aderenza delle prestazioni dei soggetti sperimentali ai tempi soglia stabiliti, con una coerenza generale nei tempi di completamento del task con una evidente ma fisiologica varianza basata sull'esperienza del soggetto sperimentale e del dispositivo utilizzato dallo stesso.

Tenendo conto di quanto appena affermato e che tutti i 45 soggetti sperimentali hanno completato con successo il task assegnato nell'esperimento e, dunque, nessuno di loro ha rinunciato a concludere anticipatamente lo stesso senza il raggiungimento dell'obiettivo, si può affermare che tali risultati forniscono una base solida per valutare soddisfacenti le prestazioni della piattaforma I-Mole.

6. RICERCA DI UN SERVIZIO

Così come definitivo e progettato nell'OR 4, l'obiettivo di ricerca di questo caso di studio è il seguente:

> Analizzare Piattaforma I-Mole al fine di valutarla rispetto alla complessità di ricercare un servizio offerto dalla piattaforma

La valutazione dell'indagine è stata effettuata rilevando i dati di 3 esecuzioni (RUN40, RUN41 e RUN42).

I soggetti sperimentali interessati a questa indagine sono tutti i 45 utenti disponibili; essi sono stati suddivisi in tre batterie da 15; ogni batteria ha eseguito un RUN.

Durante questa indagine sono state rilevate le seguenti misure:

- TPR Tempo Persona Rilevato durante l'esecuzione del task assegnato;
- CES indice del questionario CES sottoposto ai soggetti sperimentali alla fine dell'esecuzione del task ed i cui risultati sono mostrati nell'OR 5.2.

L'analisi mira a esaminare i risultati emersi dall'esperimento condotto al fine di valutare la piattaforma I-Mole. Partendo dai dati raccolti, si offre un'analisi dettagliata delle principali osservazioni emerse durante il corso della sperimentazione.

Per ogni RUN, in conformità con quanto effettuato nelle due precedenti sperimentazioni, sono stati calcolati i seguenti indici statistici ritenuti particolarmente rilevanti per raggiungere l'obiettivo prefissato.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai soggetti sperimentali per completare il task.

Scarto percentuale: Lo scarto percentuale tra i tempi persona rilevati e il valore soglia fornisce un'indicazione della discrepanza percentuale tra i tempi effettivi di completamento e il tempo massimo accettabile. Un piccolo scarto percentuale indica una buona aderenza al valore soglia. Esso si determina con la seguente formula (TPC - TPR) / TPC).

Deviazione Standard: La deviazione standard è una misura di dispersione che indica quanto i dati sono distribuiti intorno alla media; ovvero, rappresenta la

variabilità dei dati rispetto alla media. Una deviazione standard più alta indica una maggiore dispersione dei dati rispetto alla media, mentre una deviazione standard più bassa indica una maggiore coerenza o omogeneità nei dati. Una deviazione standard relativamente alta può avere implicazioni pratiche importanti. Ad esempio, potrebbe indicare che alcuni soggetti completano il task in tempi molto più brevi rispetto ad altri, suggerendo la presenza di potenziali fattori di confondimento, in senso negativo, o differenze individuali nella capacità di eseguire il task.

L'aggiunta dell'analisi della deviazione standard fornisce ulteriori dettagli sulla variabilità dei tempi di completamento del task sulla piattaforma web. Questo parametro è cruciale per comprendere la consistenza delle prestazioni degli utenti e l'affidabilità della piattaforma nel supportare l'esecuzione delle attività.

La deviazione standard è stata calcolata come una percentuale rispetto al Valore Soglia (TPC) e come percentuale dei Tempi Persona Rilevati (TPR). Questo permette di valutare la dispersione dei dati rispetto al tempo medio di completamento del task.

Varianza: La varianza dei tempi persona rilevati fornisce una misura della dispersione dei dati intorno alla media. Una varianza più alta indica una maggiore dispersione dei tempi di completamento; ovvero, indica una maggiore dispersione dei dati rispetto alla media. Una varianza più bassa indica, invece, una maggiore coerenza.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il coefficiente di variazione si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un coefficiente basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare.

6.1 Analisi TPR – RUN 40 – RUN 41 – RUN 42

L'obiettivo di questa macroarea dell'indagine è la validazione della complessità di ricercare il servizio richiesto dal cliente nella "linea di servizi" che il sistema propone.

Durante l'esperimento un tutor indicato da Assioma ha rilevato il valore del **TPR (Tempo Persona Rilevato)** per ogni soggetto sperimentale impegnato nell'esecuzione del task assegnato.

Inoltre, a priori è stato definito il **Tempo Persona di Confronto (TPC)** che rappresenta il valore soglia che gli esperti di dominio (esperti di Assioma che padroneggiano la piattaforma IMOLE nonché il dominio della logistica) hanno stabilito come accettabile per completare il task.

6.1.1 RUN 40

Così come indicato nel paragrafo precedente, di seguito sono indicati i valori degli indici per il RUN 40.

Tempo Persona di Confronto (TPC): Questo valore rappresenta il valore soglia che gli esperti di dominio hanno stabilito come accettabile per completare il task. Per questo esperimento, il TPC è pari a 300 secondi, che corrisponde a 5 minuti.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 290 secondi, che corrisponde a 4 minuti e 50 secondi. Questo indica che, in media, i partecipanti hanno completato il task entro un periodo di tempo inferiore rispetto al TPC, il che è un risultato positivo.

Scarto percentuale: Lo scarto percentuale tra i Tempi Persona Rilevati e il valore soglia (TPPC - TPPR) / TPPC), è del 3,33%. Questo valore fornisce un'indicazione della differenza percentuale tra il tempo medio impiegato dai partecipanti e il tempo soglia. Un valore basso, come nel nostro caso, è un'altra indicazione positiva poichè indica una migliore aderenza al TPC.

Deviazione standard: La deviazione standard è una misura di dispersione dei dati rispetto alla media. In questa sperimentazione, la deviazione standard è stata di 42,50 secondi; ovvero, i tempi di completamento del task variano in media di circa 42,50 secondi rispetto alla media di 290 secondi. Ancora una volta, questa è una indicazione positiva poichè suggerisce una certa coerenza nei tempi di completamento tra i partecipanti.

Varianza: La varianza è il quadrato della deviazione standard e fornisce una misura della dispersione dei dati. Essa è risultata pari a 1934,84 secondi al quadrato.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il CV si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un CV basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare. Nel RUN 40 tale valore è pari a 14,65%.

L'analisi della piattaforma IMOLE evidenziata dall'esecuzione del RUN 40 evidenzia una serie di risultati positivi, compresa l'aderenza agli obiettivi di tempo di completamento, l'accuratezza dei tempi rilevati e la stabilità delle prestazioni il che soddisfa appieno il gruppo di lavoro impegnato nella sperimentazione.

6.1.2 RUN 41

Si passa, dunque, ad illustrare i valori quantitativi raccolti e determinati per il RUN 41:

Tempo persona di Confronto (TPC): Questo valore rappresenta il valore soglia che gli esperti di dominio hanno stabilito come accettabile per completare il task. Anche per questo esperimento, il TPC è pari a 300 secondi, che corrisponde a 5 minuti.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 281 secondi, che corrisponde a 4 minuti e 41 secondi. Come si può notare, la media dei Tempi Persona Rilevati (TPR) è inferiore al Valore Soglia (TPC) di 19 secondi, indicando un normale andamento nell'esecuzione dei task da parte dei soggetti sperimentali.

Scarto percentuale: Lo scarto percentuale tra i Tempi Persona Rilevati e il valore soglia (TPC - TPR) / TPC), è del 6,33%. Questo indica che la media dei tempi persona rilevati è leggermente inferiore al valore soglia, il che suggerisce che in media i soggetti hanno completato il task in un tempo più breve rispetto a quanto previsto dagli esperti.

Deviazione standard: Nel RUN41 la deviazione standard è risultata pari a 29,99 secondi; tale eccellente valore indica, dunque, che i tempi di completamento del task tra i soggetti sperimentali variano di circa 30 secondi rispetto alla media di 281 secondi (4 minuti e 41 secondi). Tale indice è ritenuto positivo dagli sperimentatori visto che in letteratura è indicato il 20% come valore di accettazione.

Varianza: La varianza è il quadrato della deviazione standard e rappresenta la media dei quadrati delle differenze tra ogni valore e la media. Essa è risultata pari a 963,50 secondi al quadrato.

Coefficiente di variazione (CV): consente di vedere quanto le varianze variano rispetto alla media. Il CV si calcola come il rapporto tra la deviazione standard e la media dei TPR. Un CV basso indica che le varianze sono coerenti; se, invece, esso risultasse essere alto potrebbe indicare che ci sono discrepanze significative da analizzare. Nel RUN 41 tale valore è pari a 10,67%.

L'analisi dei dati raccolti fornisce una panoramica completa delle prestazioni della piattaforma I-MOLE rispetto alla funzionalità di ricerca di un servizio progettata per il RUN 41. Infatti, mentre la media dei tempi persona rilevati è leggermente inferiore al TPC, altri indicatori come lo scarto percentuale, la deviazione standard e il coefficiente di variazione indicano una certa coerenza nei tempi rilevati, che soddisfa appieno il gruppo di lavoro impegnato nella sperimentazione.

6.1.3 RUN 42

Si passa, infine, ad illustrare i valori quantitativi raccolti e determinati per il RUN 42:

Valore soglia (TPC): Questo valore rappresenta il tempo massimo che gli esperti di dominio hanno stabilito come accettabile per completare il task. Anche per questo esperimento, il TPC è pari a 300 secondi, che corrisponde a 5 minuti.

Media dei Tempi Persona Rilevati (TPR): Questo valore rappresenta la media dei tempi impiegati dai 15 soggetti sperimentali per completare il task. Esso è risultato pari a 288 secondi, che corrisponde a 4 minuti e 48 secondi. Come si può notare, la media dei Tempi Persona Rilevati (TPR) è inferiore al Valore Soglia (TPC) di 12 secondi, indicando un normale andamento nell'esecuzione dei task da parte dei soggetti sperimentali.

Scarto percentuale: Lo scarto percentuale tra i Tempi Persona Rilevati e il valore soglia (TPC - TPR) / TPC), è del 4,00%. Questo indica che la media dei tempi persona rilevati è leggermente inferiore al valore soglia, il che suggerisce che in media i soggetti hanno completato il task in un tempo leggermente maggiore rispetto a quanto previsto dagli esperti.

Deviazione standard: Nel RUN42 la deviazione standard è risultata pari a 31,58 secondi; tale valore, non certamente alto, indica, dunque, che i tempi di completamento del task tra i soggetti sperimentali variano di 31,51 secondi rispetto alla media di 288 secondi (4 minuti e 48 secondi). Tale indice è ritenuto positivo dagli sperimentatori visto che in letteratura è indicato il 20% come valore di accettazione.

Varianza: La varianza è il quadrato della deviazione standard e rappresenta la media dei quadrati delle differenze tra ogni valore e la media. Essa è risultata pari a 1069 secondi al quadrato.

Nell'esperimento RUN 42, il risultato ottenuto suggerisce che i dati sono relativamente consistenti e ciò soddisfa appieno il gruppo di lavoro impegnato nella sperimentazione.

6.1.4 Conclusioni analisi RUN 40 – RUN 41 – RUN 42

I risultati dei tre RUN sperimentali (RUN 40, RUN 41, RUN 42) sono coerenti con gli obiettivi del gruppo di lavoro. A rafforzare tale convinzione contribuiscono anche i 6 grafici che seguono fortemente consigliati dalla letteratura per la validazione dei risultati degli esperimenti di questo tipo; ovvero, tre grafici a dispersione dei tempi rilevati durante l'esecuzione dei task assegnati ai 15 soggetti sperimentali e tre istogramma, che seguono, che illustrano la media dei tempi per classe di soggetti sperimentali.

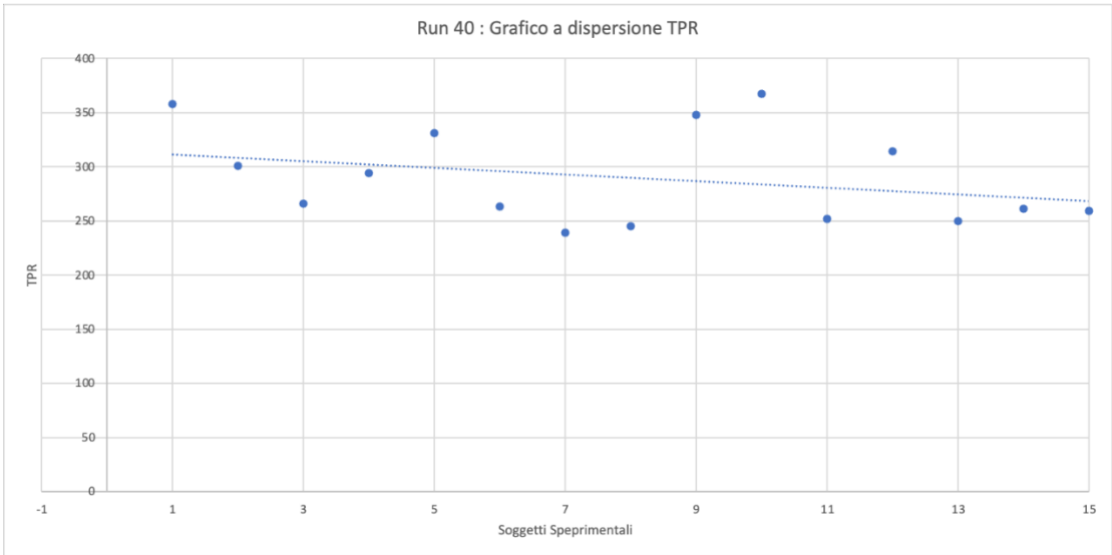


Figura 6.1- Grafico a dispersione TPR - RUN 40

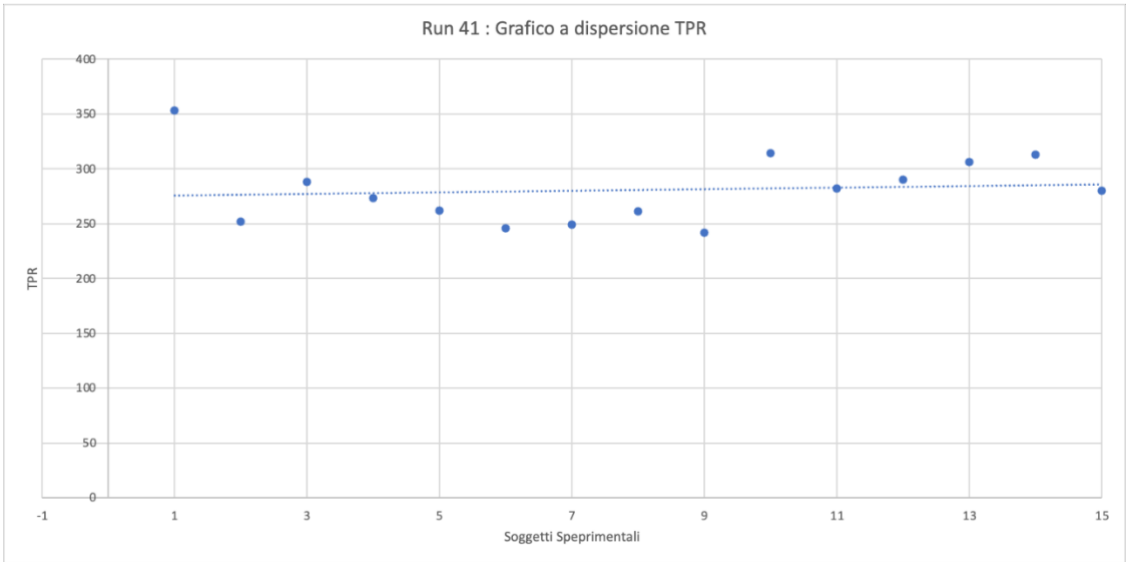


Figura 6.2 - Grafico a dispersione TPR - RUN 41

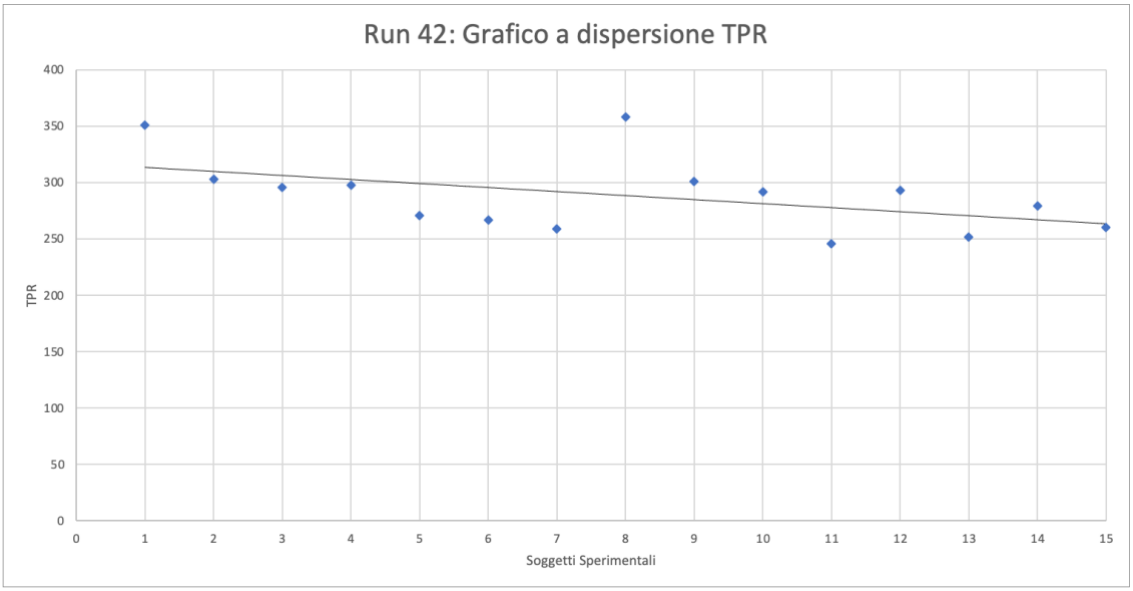


Figura 6.3 - Grafico a dispersione TPR - RUN 42

Dall'analisi del tre grafici a dispersione dei TPR è possibile dedurre diverse osservazioni; ovvero:

- Conformità al valore medio atteso: Il fatto che la linea di tendenza del grafico a dispersione si sovrapponga quasi alla linea retta del valore medio atteso suggerisce che la media dei dati osservati è molto vicina al valore soglia stabilito dagli esperti di Assioma. Questo indica che, in generale, i dati osservati tendono a seguire il valore medio atteso.

- Coerenza dei risultati: La distribuzione dei dati intorno alla linea di tendenza suggerisce una certa coerenza nei risultati. Ciò può essere interpretato come una stabilità o una consistenza nei tempi di completamento del task sulla piattaforma web tra i diversi partecipanti.
- Possibile aderenza agli standard di performance: Visto che il valore medio atteso (TPC) rappresenta un obiettivo di performance e, dunque, un valore di riferimento, la conformità dei dati osservati a questo valore indica che la piattaforma I-Mole raggiunge gli standard di performance attesi.

Dunque, le osservazioni suggeriscono una certa coerenza e conformità dei dati rispetto al valore TPC, indicando una stabilità e una possibile aderenza agli standard di performance.

Nonostante la conformità generale ai tempi soglia, è importante esaminare la dispersione dei dati intorno alla linea di tendenza. Una maggiore dispersione potrebbe indicare una variabilità nei tempi di completamento tra i partecipanti, che potrebbe essere dovuta a differenze individuali nelle capacità cognitive o nelle strategie utilizzate per completare il task. A tal fine è sicuramente utile esaminare le tendenze o pattern nel grafico a dispersione; ovvero, far emergere classi di soggetti sperimentali con prestazioni consistentemente migliori o peggiori rispetto alla media, il che potrebbe suggerire la presenza di fattori o variabili aggiuntive che influenzano i tempi di completamento. Per tal motivo, e come preannunciato, vengono visualizzati i tre istogramma che illustrano la media dei tempi per classe di soggetti sperimentali.

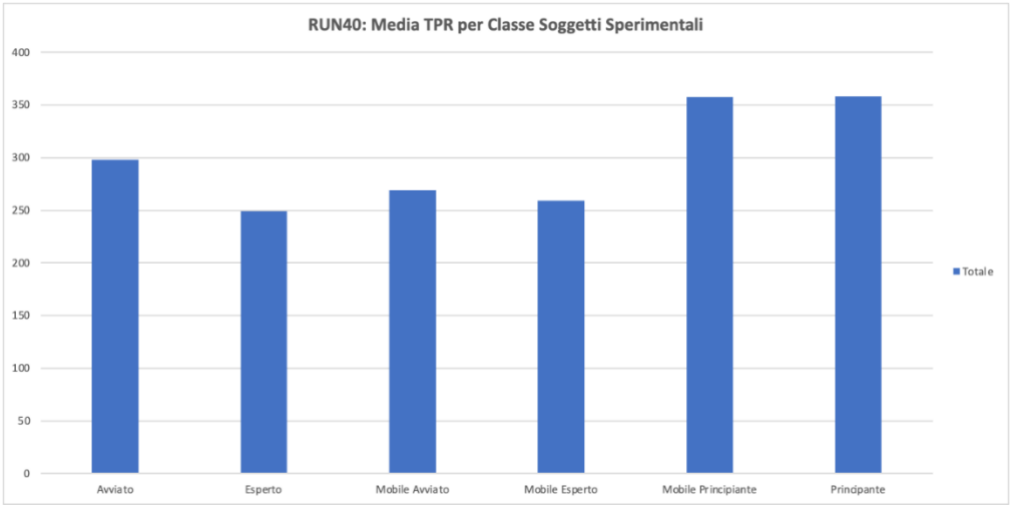


Figura 6.4 – Media dei tempi RUN 40 per classe

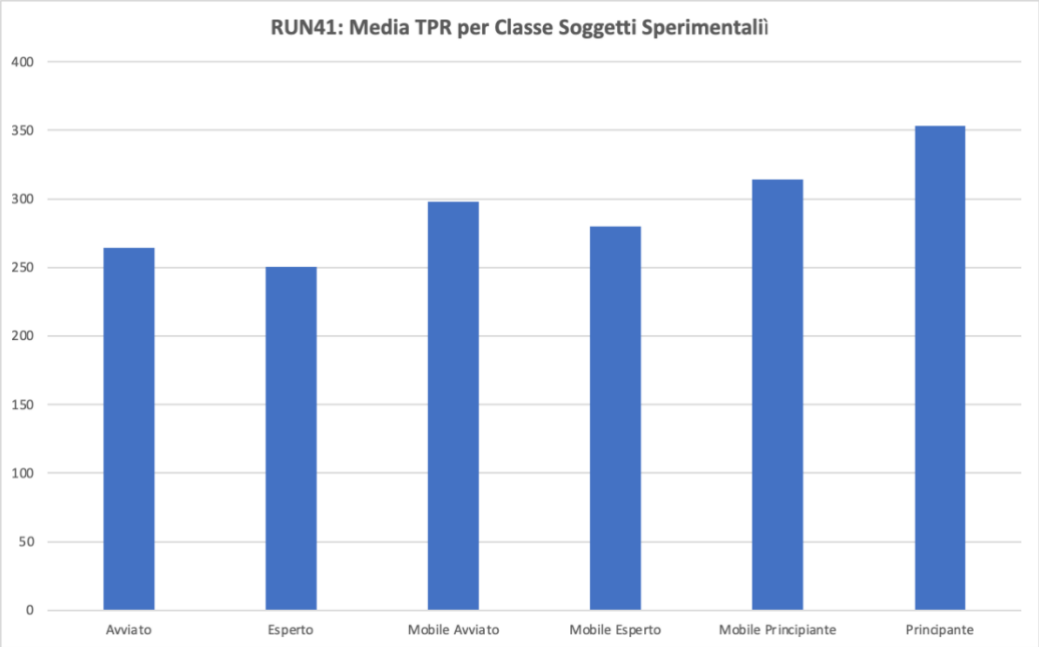


Figura 6.5 Media dei tempi RUN 41 per classe

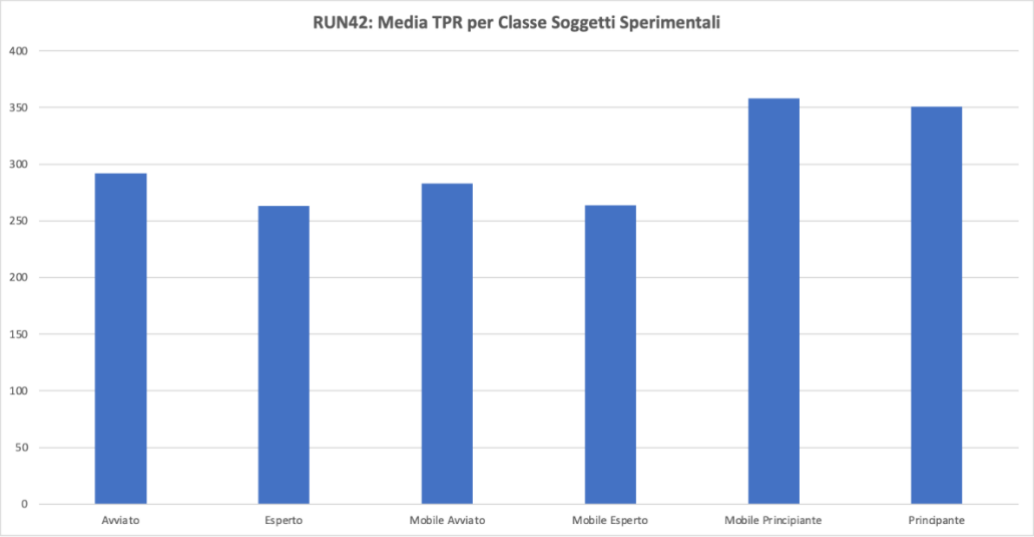


Figura 6.6 Media dei tempi RUN 42 per classe

Anche per queste tre esecuzioni dall'analisi di questi ultimi istogrammi si deduce che il tipo di dispositivo utilizzato non incide granché sulla tempistica di esecuzione di un task. Questo conferma ciò che si è affermato per la sperimentazione eseguita con il RUN 20, laddove si evidenziava la natura "responsive" della piattaforma I-Mole.

L’analisi complessiva dei nove grafici presentati conduce ad evidenziare una buona aderenza delle prestazioni dei soggetti sperimentali ai tempi soglia stabiliti, con una coerenza generale nei tempi di completamento del task con una evidente ma fisiologica varianza basata sull’esperienza del soggetto sperimentale e del dispositivo utilizzato dallo stesso.

Tenendo conto di quanto appena affermato e che tutti i 45 soggetti sperimentali hanno completato con successo il task assegnato nell’esperimento e, dunque, nessuno di loro ha rinunciato a concludere anticipatamente lo stesso senza il raggiungimento dell’obiettivo, si può affermare che tali risultati forniscono una base solida per valutare soddisfacenti le prestazioni della piattaforma I-Mole.

Considerazioni finali

In riunioni fatte in seguito all’esperimento ed effettuate con esperti di Assioma, per l’analisi di tali risultati ed al fine di ottimizzare ulteriormente le prestazioni della piattaforma web e ridurre la variabilità nei tempi di completamento si è deciso di verificare se ci sono aree della piattaforma web che potrebbero essere ottimizzate per migliorare la velocità di caricamento delle pagine e delle interazioni utente. Ciò potrebbe includere ottimizzazioni del codice, riduzione del numero di richieste di rete o miglioramenti nell'architettura del server.

Le differenze nei tempi di completamento tra i gruppi possono avere importanti implicazioni pratiche per la progettazione della piattaforma web e per la pianificazione di future sperimentazioni. Ad esempio, potrebbe essere necessario fornire istruzioni più dettagliate o un supporto aggiuntivo per i "principianti" al fine di ridurre i tempi di completamento e migliorare l'esperienza complessiva dell'utente.

7. ANALISI CES

Così come indicato nell'OR 4.1 di progettazione dell'esperimento, ogni attività effettuata dai soggetti sperimentali è stata monitorata da un tutor di Assioma; a tal fine oltre alla misura quantitativa TPR, analizzata nei paragrafi precedenti, è stata rilevata una qualitativa derivante dal questionario CES (Customer Effort Score) che ogni soggetto sperimentale ha compilato dopo che aver eseguito il task assegnatogli.

Seguendo le linee guida indicate nell'OR 4.3 di individuazione ed allestimento dei cantieri sperimentali, il questionario CES sottoposto ai soggetti sperimentali è costituito da una sola domanda; ovvero:

"Come valuti l'impegno richiesto per comporre il servizio desiderato tra le diverse opzioni proposte dal sistema?

Il soggetto sperimentale ha risposto rispettando la seguente scala Llikert:

Facile Neutro Difficile

Nelle sette esecuzioni, precedentemente illustrate ed analizzate sono state ottenute 105 risposte, indicate nell'OR 5.2 di rilevazione dati sperimentali che evidenziano, per ogni esecuzione, un indice sempre maggiore di 68% che in letteratura è consolidato come valore soglia accettabile per una valutazione positiva della piattaforma analizzata.

Mentre nella prima esecuzione il valore del CES è stato pari a 73,33%, nelle restanti sei il valore del CES si è sempre attestato all'86,67% ottenendo sempre 14 risposte "Facile" su 15 questionari compilati.

L'indice CES totale ottenuto è dell'84,76% che indica, dunque, che la maggioranza dei soggetti sperimentali ha valutato l'esperienza complessiva dell'uso della piattaforma web come relativamente basso in termini di sforzo richiesto. Questo suggerisce che la maggior parte dei partecipanti ha percepito l'interazione con la piattaforma come relativamente agevole e priva di ostacoli significativi.

Nel progetto I-Mole l'indice CES è stato utilizzato come misura della soddisfazione dell'utente relativa all'esperienza complessiva di utilizzo della piattaforma Web. Un indice elevato, come nel caso del di questa sperimentazione, indica un alto grado di soddisfazione dei soggetti sperimentali e che, quindi, la piattaforma web ha raggiunto un buon livello di

soddisfazione dell'utente in termini di facilità d'uso e di esperienza complessiva.

Tenendo conto di quanto appena affermato e che ben 97 risposte ottenute (su 105) hanno valutato l'esperienza come 'facile' a fronte di 2 valutazioni “difficile” e 6 “neutro”, si può affermare che tali risultati forniscono una base solida per valutare soddisfacente il grado di accettazione della piattaforma I-Mole.

8. USABILITÀ E UX DEL PORTALE

8.1 Analisi per l'indagine di Usabilità del Portale

Così come definitivo e progettato nell'OR 4, l'obiettivo di ricerca di questo caso di studio è il seguente:

> Analizzare Piattaaforma I-Mole al fine di valutarla rispetto alla complessità di usare la piattaforma dal punto di vista dell'usabilità nel contesto della usability evaluation.

La valutazione dell'indagine è stata effettuata facendo compilare ai soggetti sperimentali i questionari SUS dopo ogni RUN delle sperimentazioni illustrate nei precedenti paragrafi.

L'usabilità di una piattaforma web è un aspetto cruciale per garantire un'esperienza soddisfacente agli utenti.

Il System Usability Scale (SUS) è uno strumento ampiamente utilizzato per valutare soggettivamente l'usabilità di sistemi interattivi, come le piattaforme web; esso è un questionario composto da dieci item che coprono una vasta gamma di aspetti legati all'usabilità, come la facilità di apprendimento, l'efficienza nell'uso e la soddisfazione complessiva.

8.1.1 Usabilità della Composizione dei Servizi

Questa indagine ha raccolto i sondaggi SUS dei quindici soggetti sperimentali impegnati nel RUN 20 ed i cui risultati sono mostrati nel T.R. 5.2.

Il punteggio SUS complessivo ottenuto è pari a 77,66 su una scala da 0 a 100. Questo punteggio riflette un giudizio complessivamente positivo sull'usabilità della piattaforma I-MOLE. Tuttavia, è importante notare che ci potrebbero essere aree specifiche che potrebbero beneficiare di miglioramenti per garantire un'esperienza utente ottimale.

Quindi, mentre i risultati complessivi del questionario SUS indicano un livello di usabilità ragionevolmente alto della piattaforma web, è importante analizzare i singoli item del questionario per identificare gli aspetti specifici che potrebbero richiedere attenzione. Ad esempio, potremmo esaminare se ci sono domande specifiche del SUS su cui gli utenti hanno dato punteggi inferiori alla media, indicando aree di possibile miglioramento.

Per tal motivo, è stato realizzato il grafico che segue.

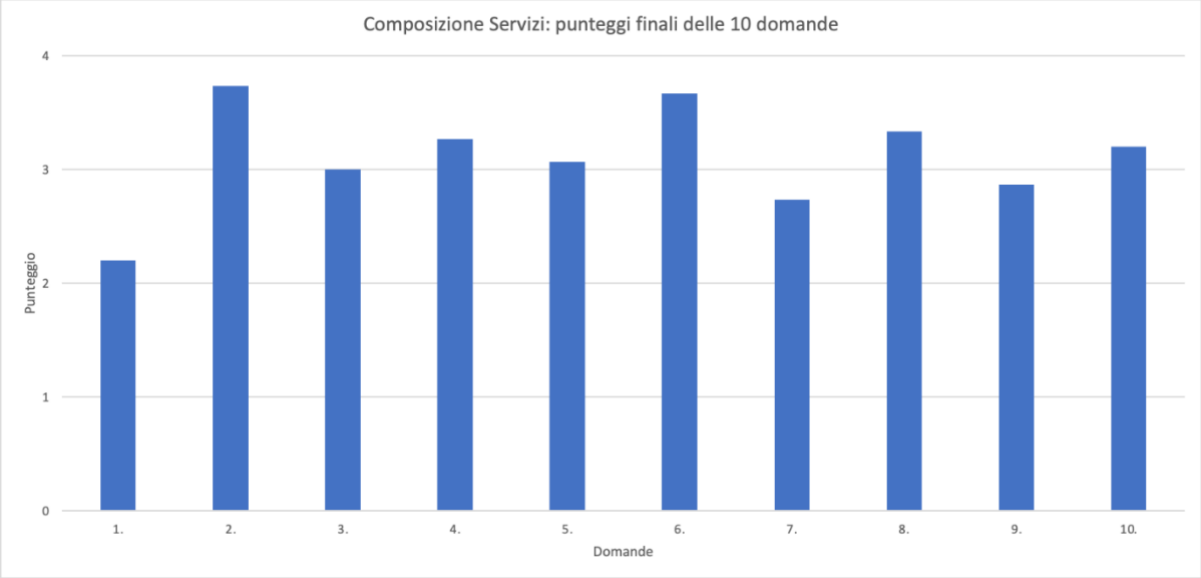


Figura 8.1 - Punteggio domande - Composizione Servizi

Analizzando il grafico si può notare che:

Domanda 1 – Penso che mi piacerebbe utilizzare questo sito frequentemente (2,20): Il punteggio medio indica che la maggior parte degli utenti è neutrale o leggermente incline verso il desiderio di utilizzare il sito frequentemente. Questa è la domanda che ha ricevuto il punteggio inferiore dell'intero questionario SUS. La motivazione è da ricercarsi, probabilmente nell'interpretazione che i soggetti sperimentali hanno dato alla stessa: quindi, non tanto il fatto che potrebbero esserci elementi della piattaforma che non sono particolarmente attraenti per gli utenti quanto, semmai, la personale idea che tali temi non siano centrali rispetto al proprio vissuto quotidiano.

Domanda 2 – Ho trovato il sito inutilmente complesso (3,73): Un punteggio così alto indica che la maggior parte degli utenti ha trovato il sito facile da usare. La seconda domanda del questionario SUS ha totalizzato il punteggio più alto e ciò è un segnale importante che indica la bontà della progettazione della piattaforma I-MOLE e conferma, così come affermato dal team Assioma, l'adozione di un design centrato sull'utente che mette al centro le esigenze e le prospettive degli stessi durante tutto il processo di progettazione.

Domanda 3 – Ho trovato il sito molto semplice da usare (3,00): Nonostante il punteggio medio positivo, questo punteggio è inferiore rispetto ad altre domande che valutano la facilità d'uso del sito. Potrebbe essere necessario

indagare ulteriormente su quali aspetti del sito sono percepiti come meno semplici ed a tale fine si intende valutare come questo indice evolva nelle prossime sperimentazioni.

Domanda 4 – Penso che avrei bisogno del supporto di una persona già in grado di utilizzare il sito (3,27): Un punteggio relativamente alto indica che la maggior parte degli utenti si sente abbastanza sicura nell'utilizzare il sito in autonomia.

Domanda 5 – Ho trovato le varie funzionalità del sito bene integrate (3,07): Un punteggio medio positivo suggerisce che la maggior parte degli utenti ha trovato che le funzionalità del sito fossero integrate in modo soddisfacente.

Domanda 6 – Ho trovato incoerenze tra le varie funzionalità del sito (3,67): Un punteggio alto indica che molti utenti non hanno percepito delle incoerenze tra le diverse funzionalità del sito. Tale risultato conforta il gruppo di lavoro che, infatti, ha notato che durante la sperimentazione sono si sono verificati pochi momenti di confusione e frustrazione da parte dei soggetti sperimentali.

Domanda 7 – Penso che la maggior parte delle persone possano imparare ad utilizzare il sito facilmente (2,73): Un punteggio inferiore alla media suggerisce che gli utenti potrebbero non essere così sicuri che la maggior parte delle persone possa imparare facilmente ad utilizzare il sito. A tal fine è stato segnalato al team di Assioma che potrebbe essere necessario migliorare la chiarezza delle istruzioni e/o dei tutorial messi a disposizione degli utenti.

Domanda 8 – Ho trovato il sito molto difficile da utilizzare (3,33): Ottimo il risultato conseguito da questa domanda: molti utenti hanno trovato il sito abbastanza semplice da utilizzare.

Domanda 9 – Mi sono sentito a mio agio nell'utilizzare il sito (2,87): Un punteggio inferiore alla media, ma comunque abbastanza alto, suggerisce che alcuni utenti potrebbero non sentirsi completamente a proprio agio nell'utilizzare il sito I-MOLE. Visto anche i risultati dei quesiti precedenti non ci sono appunti particolari forniti al team Assioma.

Domanda 10 – Ho avuto bisogno di imparare molti processi prima di riuscire ad utilizzare al meglio il sito (3,20): Un punteggio relativamente alto suggerisce che molti utenti non hanno avvertito la necessità di apprendere molti processi per utilizzare appieno il sito. Ciò è un indicatore di una curva di apprendimento ripida e di una documentazione utile allo scopo.

A conforto dei buoni risultati ottenuti da questo RUN della sperimentazione vengono riprodotti i seguenti due grafici.

Nel primo si nota che tutti i risultati si dispongono lungo la linea di tendenza del grafico a dispersione, il che denota una consistenza nell'esperienza utente tra i diversi soggetti sperimentali; ovvero, che l'usabilità del sito web è stata percepita in modo simile da tutti i soggetti sperimentali. L'usabilità è una qualità soggettiva di un'applicazione; ma in questo caso si può affermare con assoluta certezza che il design del sito web di I-Mole è stato sviluppato in modo coerente e uniforme, senza significative variazioni nell'usabilità tra le diverse parti del sito o tra le diverse funzionalità.

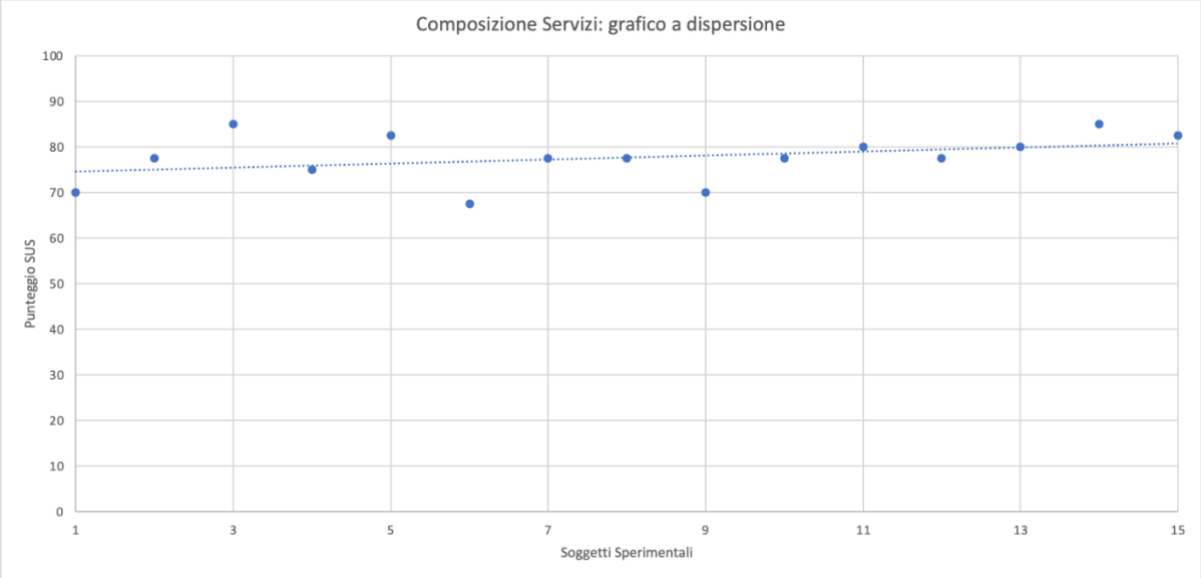


Figura 8.2 – Grafico a dispersione – Composizione servizi

Nel secondo grafico si può notare che l'esperienza pregressa dei vari soggetti sperimentali è quasi ininfluente per la semplicità di utilizzo della piattaforma I-Mole. Addirittura, per questo RUN, i soggetti sperimentali classificati come "Avviato" hanno fornito un punteggio SUS maggiore di quello risultante dal sondaggio degli "Esperti".

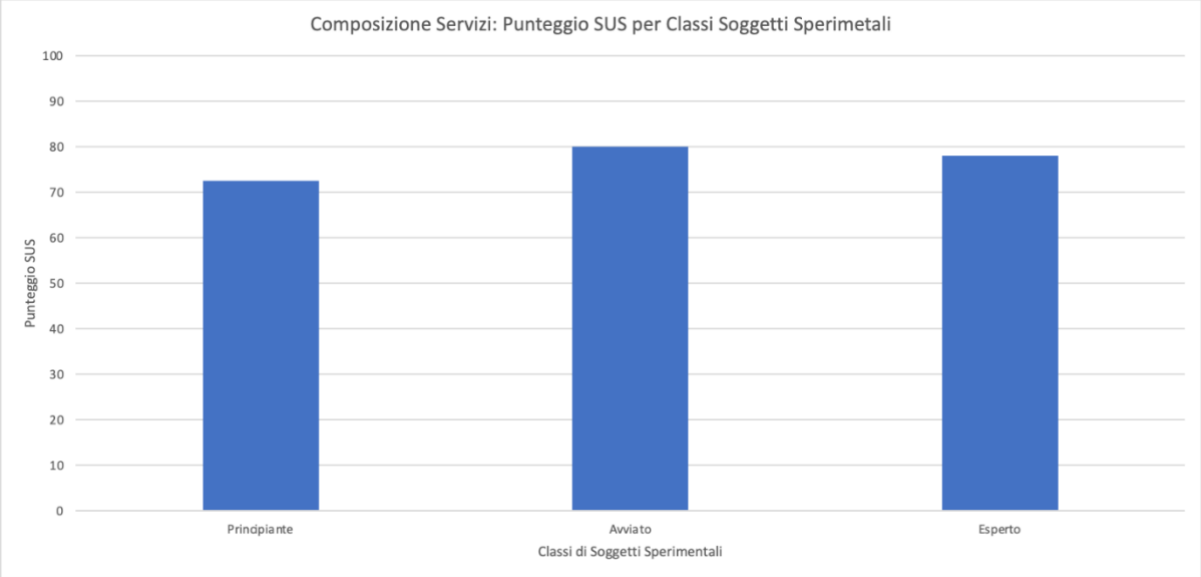


Figura 8.3 - Punteggio SUS per classi - Composizione servizi

8.1.2 Usabilità della Specificazione dei Servizi

Questa indagine ha raccolto i sondaggi SUS dei quarantacinque soggetti sperimentali impegnati nel RUN 30, 31 e 32 ed i cui risultati sono mostrati nel T.R. 5.2.

Il punteggio SUS complessivo ottenuto è pari a 88,89 su una scala da 0 a 100. Questo punteggio riflette un giudizio molto positivo sull'usabilità della piattaforma I-MOLE. È parere degli scriventi che alcuni suggerimenti informalmente forniti dai soggetti sperimentali alla fine del RUN20 hanno influito sul miglioramento della performance SUS della piattaforma I-MOLE.

Anche in questo caso si ritiene importante analizzare i singoli item del questionario per identificare gli aspetti specifici che potrebbero richiedere attenzione. Ad esempio, potremmo esaminare se ci sono domande specifiche del SUS su cui gli utenti hanno dato punteggi inferiori alla media, indicando aree di possibile miglioramento.

Per tal motivo, è stato realizzato il grafico che segue.

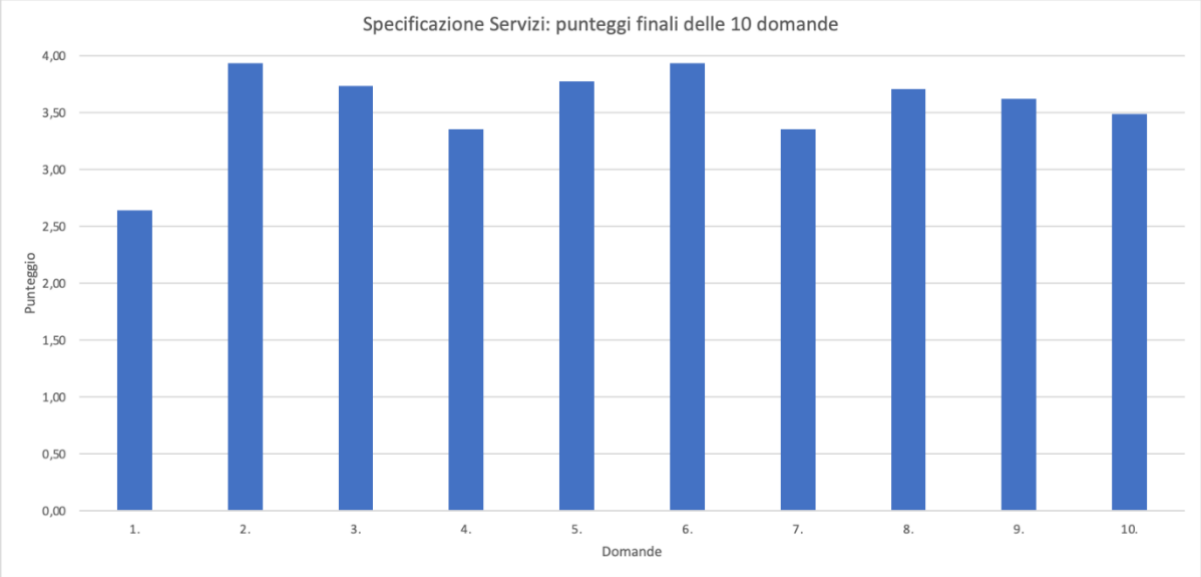


Figura 8.4 - Punteggio domande - Specificazione Servizi

Analizzando il grafico si può notare che:

Domanda 1 – Penso che mi piacerebbe utilizzare questo sito frequentemente (2,64): Il punteggio medio indica che la maggior parte degli utenti è leggermente incline verso il desiderio di utilizzare il sito frequentemente. Anche in questa sperimentazione, questa è la domanda che ha ricevuto il punteggio inferiore dell'intero questionario SUS. È convinzione sempre maggiore del gruppo di lavoro che tale risultato, comunque positivo, è dovuto all'interpretazione che i soggetti sperimentali hanno dato alla domanda: quindi; non tanto il fatto che potrebbero esserci elementi della piattaforma che non sono particolarmente attraenti per gli utenti quanto, semmai, la personale idea che tali temi non siano centrali rispetto al proprio vissuto quotidiano.

Domanda 2 – Ho trovato il sito inutilmente complesso (3,93): Un punteggio ancor più alto del RUN precedente indica che la maggior parte degli utenti ha trovato il sito facile da usare. Anche in questa sperimentazione la seconda domanda del questionario SUS ha totalizzato il punteggio più alto (insiemne alla domanda 6) e ciò è un segnale importante che indica la bontà della progettazione della piattaforma I-MOLE e conferma, così come affermato dal team Assioma, l'adozione di un design centrato sull'Utente che mette al centro le esigenze e le prospettive degli utenti durante tutto il processo di progettazione.

Domanda 3 – Ho trovato il sito molto semplice da usare (3,73): L'alto punteggio conforta il gruppo di lavoro circa la bontà della progettazione della piattaforma web I-MOLE.

Domanda 4 – Penso che avrei bisogno del supporto di una persona già in grado di utilizzare il sito (3,36): Un punteggio relativamente alto indica che la maggior parte degli utenti si sente abbastanza sicura nell'utilizzare il sito in autonomia.

Domanda 5 – Ho trovato le varie funzionalità del sito bene integrate (3,78): Nulla da commentare ulteriormente, praticamente tutti i soggetti sperimentali hanno trovato che le funzionalità del sito fossero integrate ottimamente.

Domanda 6 – Ho trovato incoerenze tra le varie funzionalità del sito (3,93): Questa domanda, ad ex aequo con la domanda 2, è quella con il punteggio più alto di questa sperimentazione. Un punteggio così alto indica che molti utenti non hanno percepito delle incoerenze tra le diverse funzionalità del sito.

Domanda 7 – Penso che la maggior parte delle persone possano imparare ad utilizzare il sito facilmente (3,36): Sarà per le indicazioni emerse dalla riunione per la valutazione del RUN20, ma la rivisitazione delle istruzioni e/o dei tutorial messi a disposizione degli utenti ha sortito un ottimo punteggio circa la percezione dei soggetti sperimentali che la maggior parte delle persone possa imparare facilmente ad utilizzare il sito.

Domanda 8 – Ho trovato il sito molto difficile da utilizzare (3,71): Ottimo il risultato conseguito da questa domanda: molti utenti hanno trovato il sito abbastanza semplice da utilizzare.

Domanda 9 – Mi sono sentito a mio agio nell'utilizzare il sito (3,62): Anche in questo caso, il punteggio risulta altissimo. Pertanto, i soggetti sperimentali si sono sentiti mediamente a proprio agio nell'utilizzare il sito I-MOLE.

Domanda 10 – Ho avuto bisogno di imparare molti processi prima di riuscire ad utilizzare al meglio il sito (3,49): Un punteggio molto alto che suggerisce che molti utenti non hanno avvertito la necessità di apprendere molti processi per utilizzare appieno il sito. Ciò è la conferma di una curva di apprendimento ripida e di una documentazione utile allo scopo.

A conforto dei buoni risultati ottenuti dai tre RUN di questa sperimentazione vengono riprodotti i seguenti due grafici.

Nel primo si nota che tutti i risultati si dispongono lungo la linea di tendenza del grafico a dispersione, il che denota una consistenza nell'esperienza utente tra i diversi soggetti sperimentali; ovvero, che l'usabilità del sito web è stata percepita in modo simile da tutti i soggetti sperimentali. L'usabilità è una qualità soggettiva di un'applicazione; ma in questo caso si può affermare con assoluta certezza che il design del sito web di I-Mole è stato sviluppato in modo coerente e uniforme, senza significative variazioni nell'usabilità tra le diverse parti del sito o tra le diverse funzionalità.

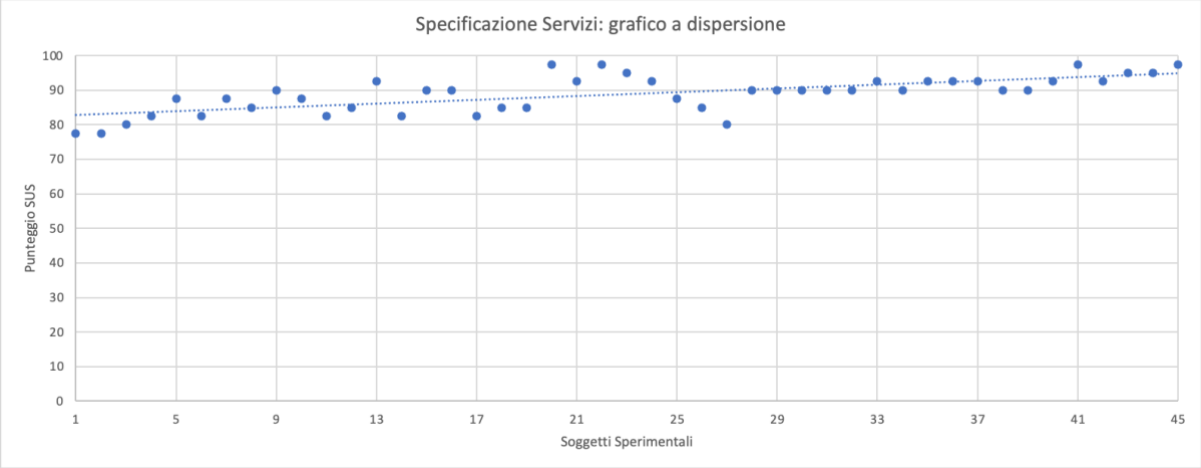


Figura 8.5 – Grafico a dispersione – Specificazione servizi

Nel secondo grafico si può notare che l'esperienza pregressa dei vari soggetti sperimentali è quasi ininfluente per la semplicità di utilizzo della piattaforma I-Mole. Infatti, se è vero che per i principianti l'usabilità è stata percepita in modo inferiore rispetto alle altre classi di soggetti sperimentali è anche vero che il valore raggiunto è superiore all''80%, ben superiore, dunque, alla soglia del 68% indicata in letteratura come minima percentuale per una valutazione positiva dell'usabilità.

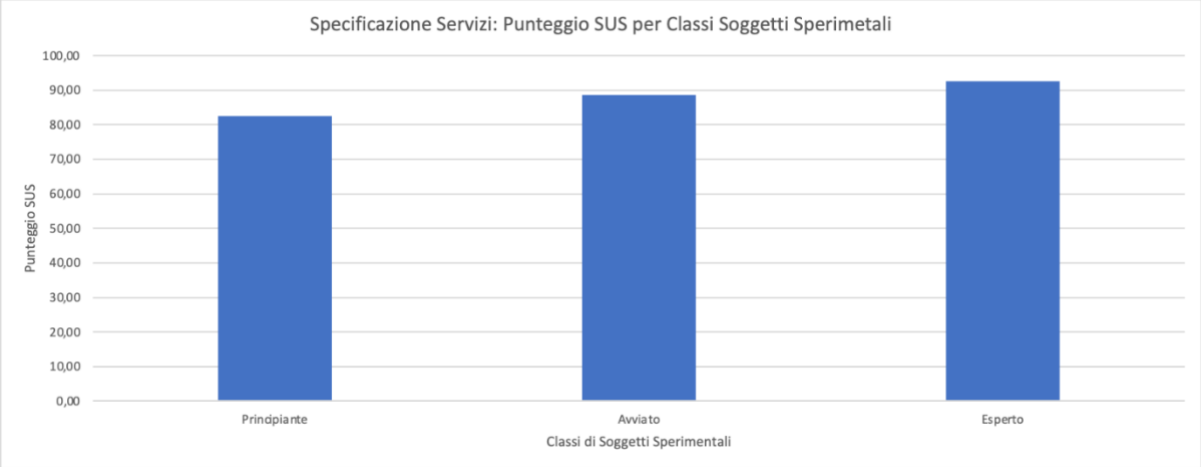


Figura 8.6 - Punteggio SUS per classi - Specificazione servizi

8.1.3 Usabilità della Ricerca dei Servizi

Questa indagine ha raccolto i sondaggi SUS dei quarantacinque soggetti sperimentali impegnati nel RUN 40, 41 e 42 ed i cui risultati sono mostrati nel T.R. 5.2.

Il punteggio SUS complessivo ottenuto è pari a 94,94 su una scala da 0 a 100. Questo eccellente punteggio riflette un giudizio molto positivo sull'usabilità della piattaforma I-MOLE. È parere degli scriventi che la familiarità acquisita dai soggetti sperimentali con le precedenti sperimentazioni hanno influito in modo significativo sul miglioramento della performance SUS della piattaforma I-MOLE.

Come si evidenzierà in seguito, il trend dei punteggi di tutte le domande del questionario SUS è crescente: tenendo conto che per quindici soggetti sperimentali era la terza volta che valutavano la piattaforma I_MOLE e per trenta di loro era solo la seconda volta, se ne deduce che la curva di apprendimento nonché la familiarità di utilizzo crescono rapidamente ed in modo direttamente proporzionale alle "visite" effettuate sulla piattaforma I-MOLE a conferma della bontà della progettazione user centered adottata da Assioma.

Anche in questo caso si ritiene importante analizzare i singoli item del questionario per identificare gli aspetti specifici che potrebbero richiedere attenzione. Ad esempio, potremmo esaminare se ci sono domande specifiche del SUS su cui gli utenti hanno dato punteggi inferiori alla media, indicando aree di possibile miglioramento.

Per tal motivo, è stato realizzato il grafico che segue.

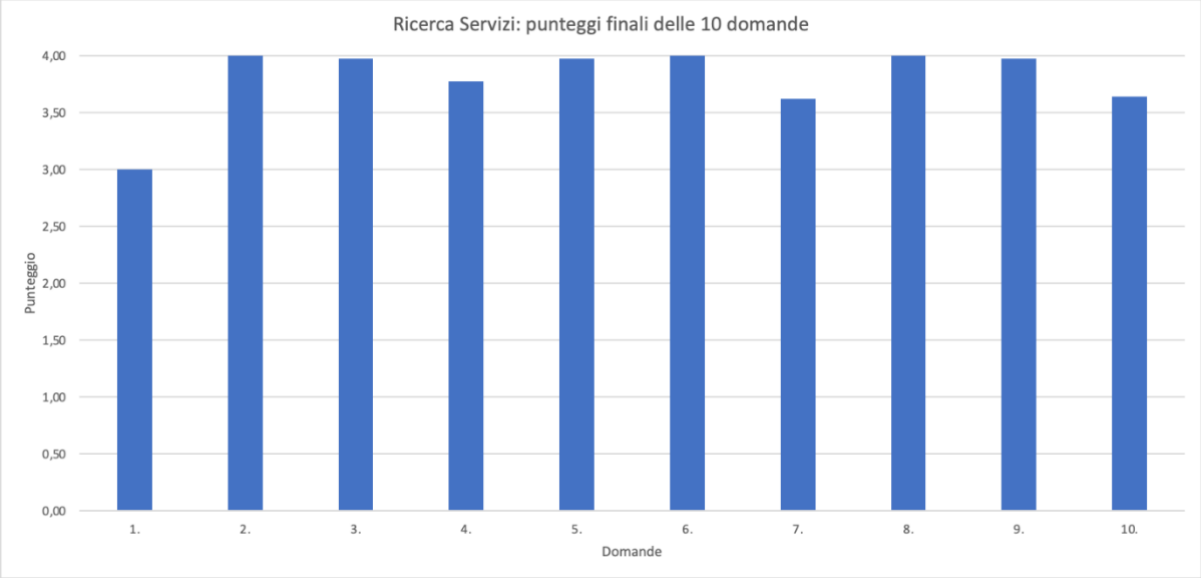


Figura 8.7 - Punteggio domande - Ricerca Servizi

Analizzando il grafico si può notare che:

Domanda 1 – Penso che mi piacerebbe utilizzare questo sito frequentemente (3,00): Il trend del punteggio della Domanda 1 è crescente; la maggior parte degli utenti è, dunque, incline verso il desiderio di utilizzare il sito frequentemente.

Domanda 2 – Ho trovato il sito inutilmente complesso (4,00): Il trend del punteggio della Domanda 2 è crescente: si noti che in questa sperimentazione tutti gli utenti hanno valutato semplice l'utilizzo della piattaforma I-MOLE attribuendo il massimo punteggio, appunto 4.

Domanda 3 – Ho trovato il sito molto semplice da usare (3,98): Il trend del punteggio della Domanda 3 è crescente: si noti che in questa sperimentazione solo un utente non ha attribuito il massimo punteggio.

Domanda 4 – Penso che avrei bisogno del supporto di una persona già in grado di utilizzare il sito (3,78): Il trend del punteggio della Domanda 4 è crescente: il punteggio così alto indica che tutti i soggetti sperimentali si sono sentiti abbastanza o completamente sicuri nell'utilizzare il sito in autonomia.

Domanda 5 – Ho trovato le varie funzionalità del sito bene integrate (3,98): Il trend del punteggio della Domanda 5 è crescente: si noti che in questa sperimentazione solo un utente non ha attribuito il massimo punteggio.

Domanda 6 – Ho trovato incoerenze tra le varie funzionalità del sito (4,00): Il trend del punteggio della Domanda 6 è crescente: si noti che in questa sperimentazione tutti gli utenti ritengono che non ci siano incoerenze tra le varie funzionalità della piattaforma I-MOLE attribuendo il massimo punteggio, appunto 4.

Domanda 7 – Penso che la maggior parte delle persone possano imparare ad utilizzare il sito facilmente (3,62): Il trend del punteggio della Domanda 7 è crescente: il punteggio così alto indica che tutti i soggetti sperimentali pensa che la maggior parte delle persone possa imparare facilmente ad utilizzare il sito.

Domanda 8 – Ho trovato il sito molto difficile da utilizzare (4,00): Il trend del punteggio della Domanda 8 è crescente: si noti che in questa sperimentazione tutti gli utenti hanno valutato semplice l'utilizzo della piattaforma I-MOLE attribuendo il massimo punteggio, appunto 4.

Domanda 9 – Mi sono sentito a mio agio nell'utilizzare il sito (3,98): Il trend del punteggio della Domanda 9 è crescente: si noti che in questa sperimentazione solo un utente non ha attribuito il massimo punteggio. Pertanto, i soggetti sperimentali si sono sentiti sicuramente a proprio agio nell'utilizzare il sito I-MOLE.

Domanda 10 – Ho avuto bisogno di imparare molti processi prima di riuscire ad utilizzare al meglio il sito (3,64): Il trend del punteggio della Domanda 10 è crescente: il punteggio così alto indica che tutti i soggetti sperimentali non hanno avvertito la necessità di apprendere molti processi per utilizzare appieno il sito. Ciò è la conferma di una curva di apprendimento ripida e di una documentazione utile allo scopo.

A conforto dei buoni risultati ottenuti dai tre RUN di questa sperimentazione vengono riprodotti i seguenti due grafici.

Nel primo si nota che tutti i risultati si dispongono lungo la linea di tendenza del grafico a dispersione, il che denota una consistenza nell'esperienza utente tra i diversi soggetti sperimentali; ovvero, che l'usabilità del sito web è stata percepita in modo simile da tutti i soggetti sperimentali. L'usabilità è una qualità soggettiva di un'applicazione; ma in questo caso si può affermare con assoluta certezza che il design del sito web di I-Mole è stato sviluppato in

modo coerente e uniforme, senza significative variazioni nell'usabilità tra le diverse parti del sito o tra le diverse funzionalità.

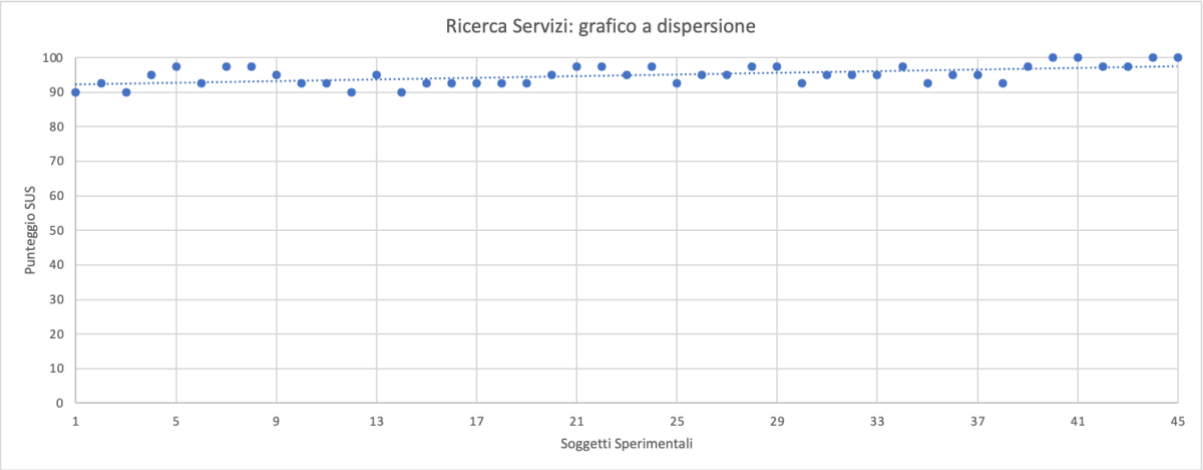


Figura 8.8 – Grafico a dispersione – Ricerca servizi

Nel secondo grafico si può notare che l'esperienza pregressa dei vari soggetti sperimentali è quasi ininfluente per la semplicità di utilizzo della piattaforma I-Mole.

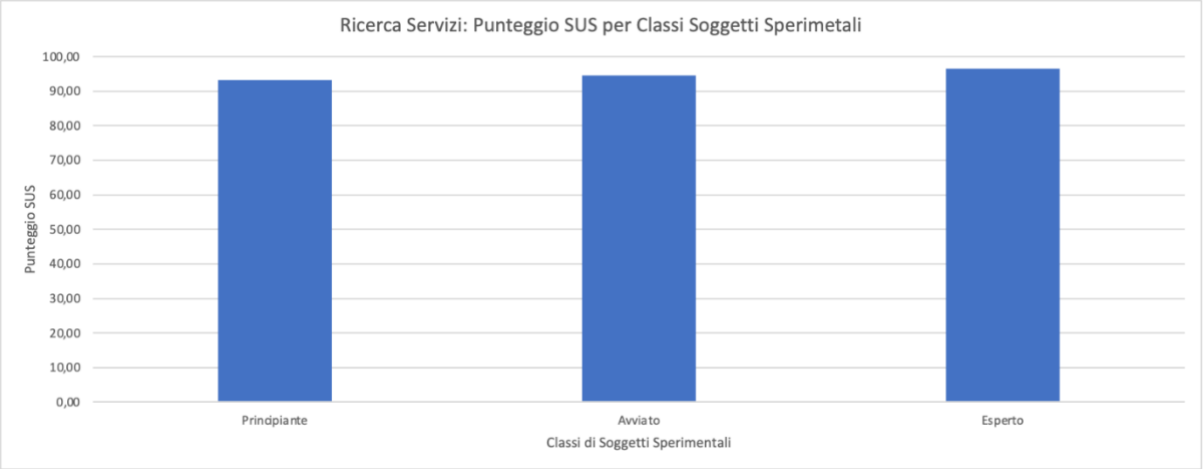


Figura 8.9 - Punteggio SUS per classi - Ricerca servizi

8.1.4 Analisi dell'indice SUS nelle tre sperimentazioni

Infine, è stato analizzato l'andamento degli indici SUS attraverso le tre sperimentazioni. È stato così ottenuto il grafico che segue.

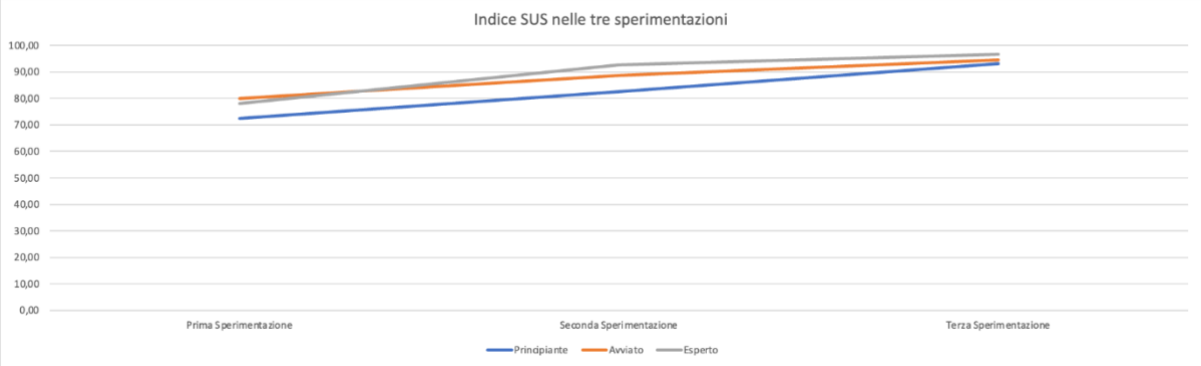


Figura 8.10 - Andamento indici SUS

L'usabilità per i soggetti sperimentali "Principianti" mostra un chiaro miglioramento nel corso delle tre sperimentazioni. Partendo da un punteggio iniziale di 72,50, si è registrato un aumento significativo fino a 93,21 nella terza sperimentazione. Questo suggerisce che i principianti hanno migliorato l'uso del sito web di I-MOLE man mano che si sono familiarizzati con esso e ciò lascia dedurre che l'usabilità del sito supporta l'esperienza all'uso.

Anche per i soggetti sperimentali "Avviati", si osserva un miglioramento dell'usabilità nel corso delle tre sperimentazioni. Sebbene il punteggio iniziale fosse già relativamente alto (80,00), si è registrato un ulteriore aumento fino a 94,60 nella terza sperimentazione. Ciò indica che anche gli utenti avviati hanno beneficiato dell'interazione continua con la piattaforma web.

Anche per gli esperti, si registra un aumento significativo dell'usabilità nel corso delle sperimentazioni. Sebbene il punteggio iniziale fosse già abbastanza alto (78,00), si è verificato un notevole miglioramento fino a 96,54 nella terza sperimentazione. Questo suggerisce che anche gli utenti esperti hanno trovato nuovi modi per ottimizzare la loro esperienza con la piattaforma nel tempo.

Dunque, l'andamento generale mostra un miglioramento costante dell'usabilità della piattaforma I-MOLE attraverso le sperimentazioni successive per tutte le classi di soggetti sperimentali. Il fatto che anche gli esperti abbiano registrato un miglioramento significativo suggerisce che la piattaforma I-MOLE sia stata in grado di soddisfare le esigenze anche dei soggetti sperimentali con capacità più avanzate

È di tutta evidenza, dunque, che nelle tre sperimentazioni successive ci sia stato un costante miglioramento dell'usabilità della piattaforma web per tutte

le classi di utenti, dimostrando un impegno continuo verso l'ottimizzazione dell'esperienza utente.

8.2 Analisi per l'indagine di UX del Portale

Così come definitivo e progettato nell'OR 4, l'obiettivo di ricerca di questo caso di studio è il seguente:

> Analizzare Piattaforma I-Mole al fine di valutarla rispetto alla User Interface dal punto di vista complessità dell'interfaccia nel contesto dell'interazione uomo macchina.

I 45 soggetti sperimentali individuati per le precedenti indagini, dopo aver eseguito le azioni previste dai task a loro assegnati ed aver risposto al quesito per il calcolo dell'indice CES hanno risposto al questionario UEQ.

L'UEQ non produce un punteggio complessivo per l'esperienza dell'utente. A causa della costruzione del questionario, non ha senso costruire un punteggio complessivo (ad esempio calcolando la media di tutte le scale), poiché questo valore non può essere interpretato correttamente. I valori dei singoli item sono elencati per consentire di individuare gli outlier nelle valutazioni. Se un item mostra grandi scostamenti rispetto alle valutazioni degli altri item della stessa scala, questo può essere un indizio del fatto che l'item è interpretato in modo errato (ad esempio, a causa di un contesto particolare nella valutazione) da un numero maggiore di partecipanti.

Come evidenziato nei precedenti deliverable, l'intervallo delle scale è compreso tra -3 (horribly bad) e +3 (extremely good).

In letteratura è consolidata la seguente scala di valutazione:

- i valori $>$ 0,8 rappresentano una valutazione positiva della scala corrispondente
- I valori compresi tra -0,8 e 0,8 rappresentano una valutazione più o meno neutra della scala corrispondente
- i valori $<$ -0,8 rappresentano una valutazione negativa della scala corrispondente

Nella tabella che segue sono evidenziate la media, la varianza e la deviazione standard per ogni item del questionario.

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Scale
1	2,7	0,4	0,6	15	annoying	enjoyable	Attractiveness
2	2,7	0,4	0,6	15	not understandable	understandable	Perspicuity
3	1,1	0,8	0,9	15	creative	dull	Novelty
4	2,5	0,3	0,5	15	easy to learn	difficult to learn	Perspicuity
5	1,8	0,3	0,6	15	valuable	inferior	Stimulation
6	1,5	0,4	0,6	15	boring	exciting	Stimulation
7	1,9	0,6	0,8	15	not interesting	interesting	Stimulation
8	2,3	0,8	0,9	15	unpredictable	predictable	Dependability
9	2,0	1,1	1,1	15	fast	slow	Efficiency
10	2,3	1,1	1,0	15	inventive	conventional	Novelty
11	2,1	0,8	0,9	15	obstructive	supportive	Dependability
12	2,5	0,6	0,7	15	good	bad	Attractiveness
13	2,7	0,2	0,5	15	complicated	easy	Perspicuity
14	2,0	2,0	1,4	15	unlikable	pleasing	Attractiveness
15	2,5	0,7	0,8	15	usual	leading edge	Novelty
16	2,4	0,4	0,6	15	unpleasant	pleasant	Attractiveness
17	2,3	0,8	0,9	15	secure	not secure	Dependability
18	1,9	0,6	0,8	15	motivating	demotivating	Stimulation
19	2,3	0,9	1,0	15	meets expectations	does not meet expectations	Dependability
20	2,5	0,4	0,6	15	inefficient	efficient	Efficiency
21	2,3	0,6	0,8	15	clear	confusing	Perspicuity
22	2,2	0,6	0,8	15	impractical	practical	Efficiency
23	2,1	0,4	0,6	15	organized	cluttered	Efficiency
24	1,9	0,9	1,0	15	attractive	unattractive	Attractiveness
25	2,3	0,5	0,7	15	friendly	unfriendly	Attractiveness
26	2,5	0,6	0,7	15	conservative	innovative	Novelty

Il grafico che segue evidenzia l'alto valore della media ottenuto da quasi tutti gli item.

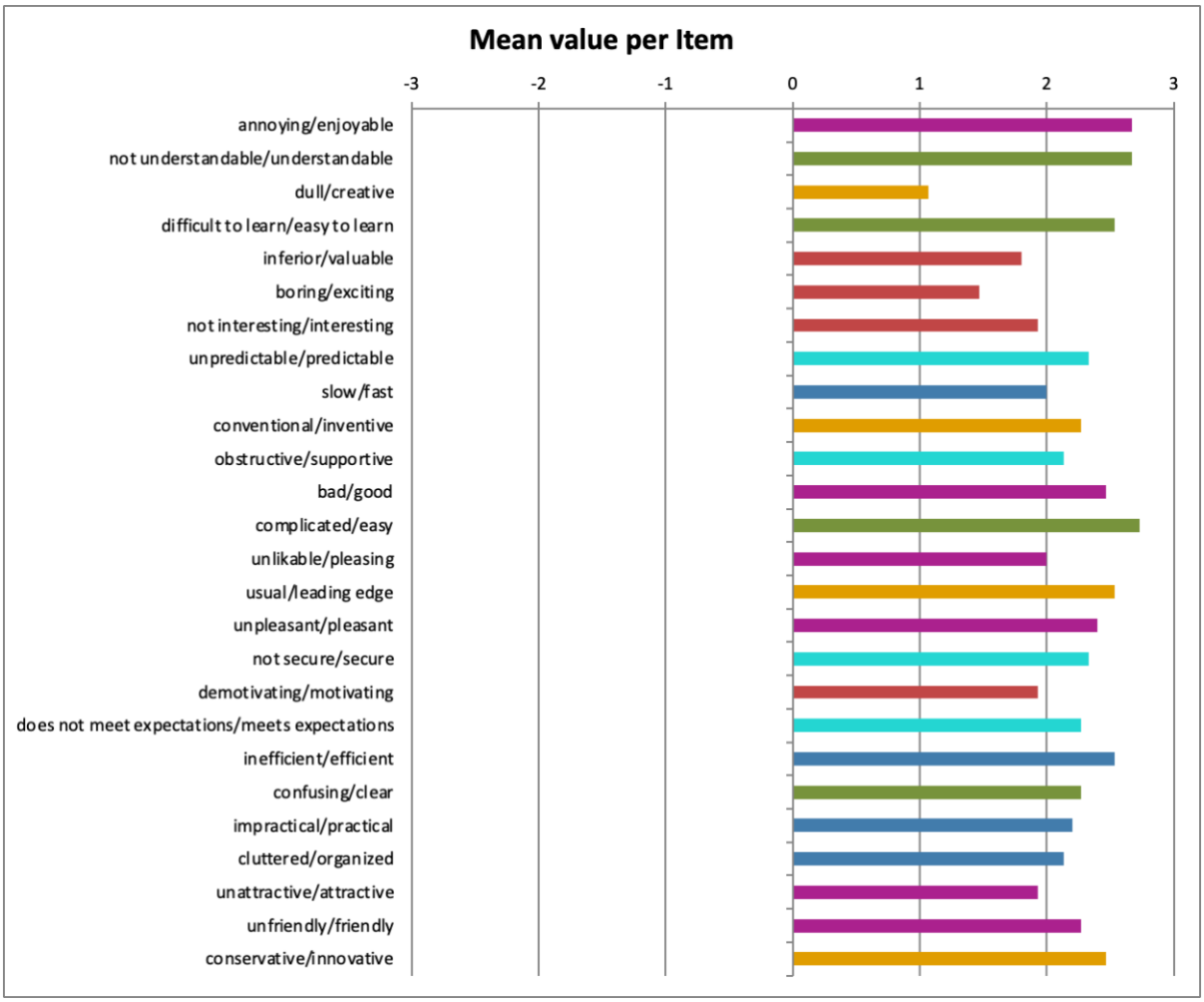


Figura 8.11 - Media valori per Item

Quindi, di seguito la tabella con la media e la varianza per sei classi di domande dello UEQ.

UEQ Scales (Mean and Variance)		
Attractiveness	2,289	0,43
Perspicuity	2,550	0,15
Efficiency	2,217	0,27
Dependability	2,267	0,42
Stimulation	1,783	0,13
Novelty	2,083	0,21

Nel grafico che segue, è mostrato l'istogramma delle medie delle sei classi di domande dello UEQ; qui è possibile notare come tutti i valori si posizionano nell'area verde, il che certifica la buona percezione della UX da parte dei soggetti sperimentali.

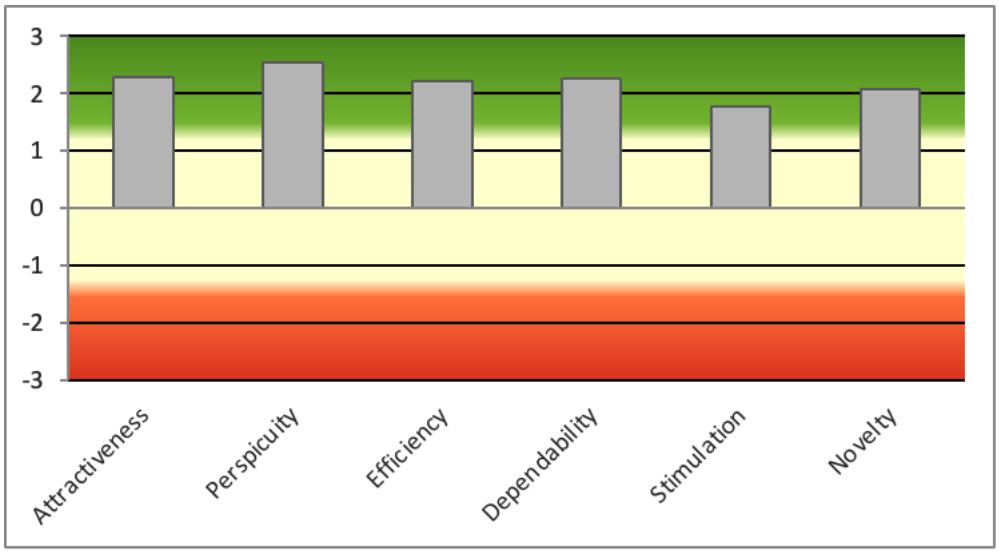


Figura 8.12 - Media classi di domande UEQ

Le scale dell'UEQ possono essere raggruppate in qualità pragmatica (Perspicuità, Efficienza, Affidabilità) e qualità edonica (Stimolazione, Originalità). La qualità pragmatica descrive gli aspetti qualitativi legati al compito, quella edonica gli aspetti qualitativi non legati al compito. Di seguito viene calcolata la media dei tre aspetti della qualità pragmatica ed edonica.

Pragmatic and Hedonic Quality	
Attractiveness	2,29
Pragmatic Quality	2,34
Hedonic Quality	1,93

Pragmatic Quality (Qualità Pragmatica): Il punteggio medio di 2,34 indica che gli utenti hanno valutato positivamente l'aspetto pratico e funzionale della piattaforma web. Questo suggerisce che il sistema soddisfa in modo efficace le esigenze degli utenti e consente loro di raggiungere i loro obiettivi in modo rapido ed efficiente. La valutazione positiva della qualità pragmatica è un indicatore importante dell'usabilità e dell'efficacia del sistema nel supportare le attività degli utenti.

Hedonic Quality (Qualità Edonica): Il punteggio medio di 1,93 indica che gli utenti hanno valutato in modo meno positivo l'aspetto emotivo e piacevole della piattaforma web. Questo potrebbe indicare che gli utenti non hanno trovato il sistema particolarmente coinvolgente o gratificante dal punto di vista emotivo. Potrebbero esserci elementi della piattaforma che non soddisfano pienamente le esigenze emotive degli utenti o che non sono considerati particolarmente attraenti o divertenti.
Attractiveness (Attrattività): Il punteggio medio di 2,29 per l'attrattività indica una valutazione moderatamente positiva da parte degli utenti riguardo alla piacevolezza e all'appeal visivo della piattaforma web. Anche se non altissimo, questo punteggio suggerisce che la piattaforma ha un certo grado di attrattività estetica e che gli utenti la trovano almeno in parte piacevole da utilizzare.
Il grafico che segue mostra l'istogramma delle medie tre classi di domande dello UEQ appena analizzate; qui è possibile notare come tutti i valori si posizionano nell'area verde, il che certifica ulteriormente la buona percezione della UX da parte dei soggetti sperimentali.

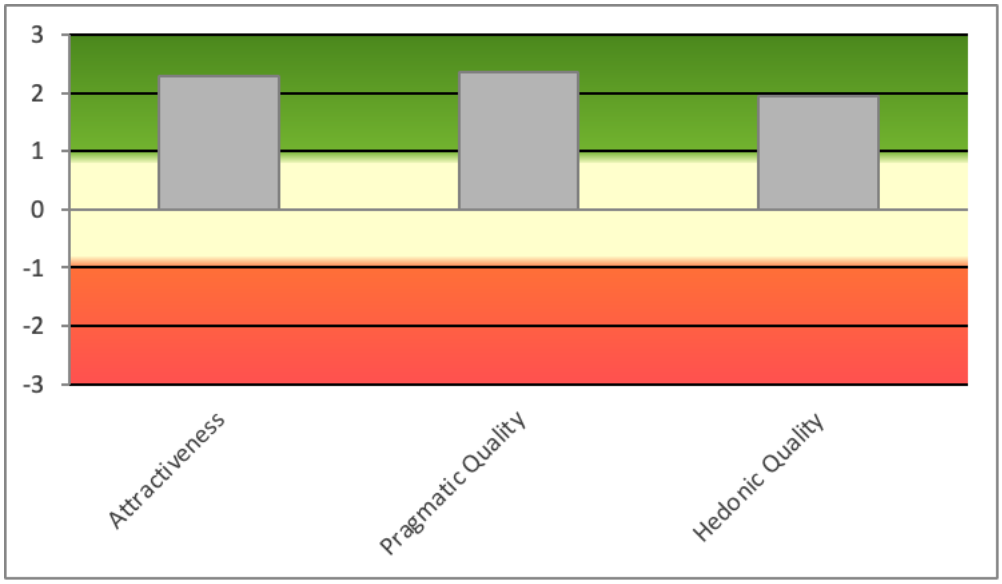


Figura 8.13 - Media tre classi domande UEQ

In sintesi, l'analisi delle dimensioni Pragmatica e Edonica della qualità suggerisce che la piattaforma web ha ricevuto valutazioni positive per quanto riguarda l'aspetto pratico e funzionale, mentre c'è spazio per miglioramenti per quanto riguarda l'aspetto emotivo e piacevole dell'esperienza utente. Queste informazioni possono essere utili per identificare aree di focus per futuri miglioramenti della piattaforma al fine di massimizzare sia l'usabilità che la soddisfazione degli utenti.

9. SECURITY TESTING

In questa sezione vengono descritte le analisi effettuate mediante l'utilizzo del tool di analisi illustrato nel TR5.1. L'obiettivo è generare una serie di diagrammi a partire dai report ottenuti dal vulnerability scanner Nessus al fine di supportare le analisi successive.

9.1 Distribuzione delle vulnerabilità per Aging

Nel corso di questa prima analisi, ci proponiamo di fornire una descrizione delle vulnerabilità basata sulla loro "età". Nel contesto della sicurezza informatica, il concetto di invecchiamento di una vulnerabilità si riferisce fondamentalmente all'anno in cui quella specifica vulnerabilità è stata individuata e catalogata per la prima volta. Appena una vulnerabilità viene scoperta, viene registrata nell'NVD (National Vulnerability Database) insieme all'anno in cui è stata individuata. Comprendere la distribuzione delle vulnerabilità in base agli anni è di vitale importanza, poiché maggiore è l'anzianità di una vulnerabilità, maggiore è il rischio ad essa associato. La prima verifica che possiamo eseguire sui dati a disposizione consiste nel valutare come le vulnerabilità si distribuiscono anno per anno. Per questa analisi, utilizzeremo la seguente metrica del MITRE:

- **Attack Vector (AV)**: fa riferimento al contesto in cui la vulnerabilità potrebbe essere sfruttata.

Tipo	Descrizione
Network (N)	Il componente vulnerabile è connesso allo stack di rete e la vulnerabilità può essere sfruttata attraverso l'intera rete.
Adjacent (A)	Il componente vulnerabile è connesso allo stack di rete, in questo caso, la vulnerabilità può essere sfruttata solamente a livello di protocollo o da una rete adiacente.
Local (L)	Il componente vulnerabile non è connesso allo stack di rete e la vulnerabilità può essere sfruttata accedendo al sistema localmente. In

	questo caso, l'attaccante fa affidamento sulla User Interaction di una seconda persona per lanciare l'attacco.
Physical (P)	Il componente vulnerabile non è connesso allo stack di rete e la vulnerabilità può essere sfruttata accedendo al sistema fisicamente.

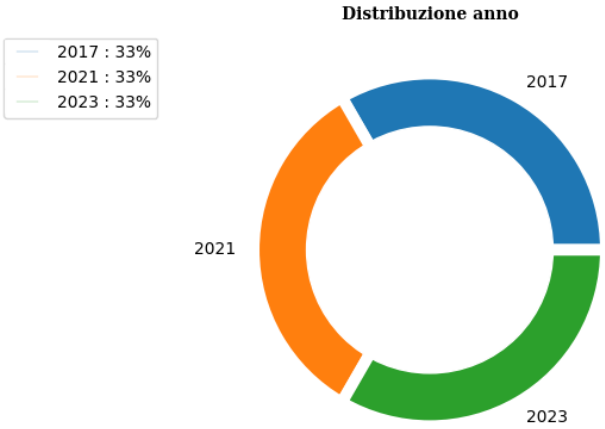


Figura 9.1 - Distribuzione percentuali vulnerabilità/anno

Dalla Figura è possibile osservare due aspetti significativi. La prima riguarda il numero di vulnerabilità identificate: esse sono equamente distribuite sui tre anni. La seconda invece sottolinea come data la loro distribuzione, la maggior parte di esse fa riferimento ad anni più recenti; questo è dovuto al fatto che col passare del tempo vengono identificate e patchate nuove vulnerabilità che affliggono i sistemi informatici, le quali potrebbero essere patchate in un ordine arbitrario e non in base al livello di rischio. Al fine di valutare questo ultimo aspetto, è stato realizzato il grafico seguente.

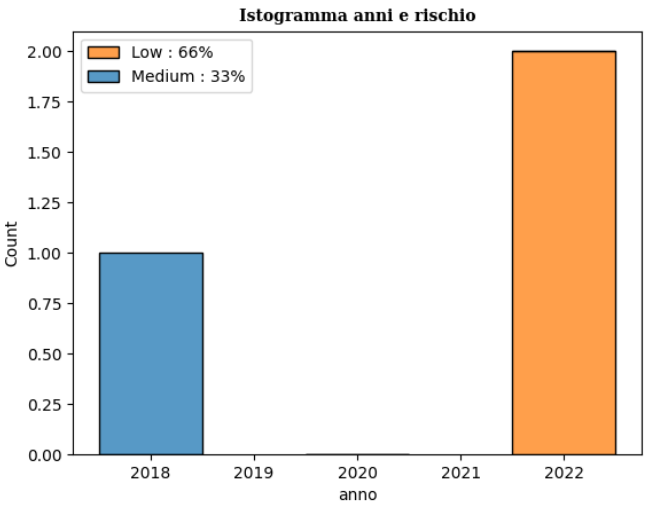


Figura 9.2 - Istogramma anni/rischio

In Figura **Error! Reference source not found.** è possibile osservare la distribuzione delle vulnerabilità per anno sulla base del loro livello di rischio. Negli anni più recenti si concentrano un numero maggiore di vulnerabilità, ma il loro livello di rischio è LOW invece, le vulnerabilità con livello di rischio MEDIUM si concentrano sugli anni meno recenti. Questa analisi suggerisce che le componenti utilizzate sono sempre più soggette a vulnerabilità, soprattutto di livello LOW. Si può ipotizzare che le patch debbano essere applicate in base al livello di rischio, anziché partendo da quelle più recenti. Dal grafico in Figura **Error! Reference source not found.**invece si evince che il vettore di attacco più utilizzato per sfruttare queste vulnerabilità fa riferimento al tipo Network. Il mezzo principale grazie al quale potrebbe essere possibile exploitare queste vulnerabilità identificate sul sistema target si riferiscono ad accessi tramite la rete.

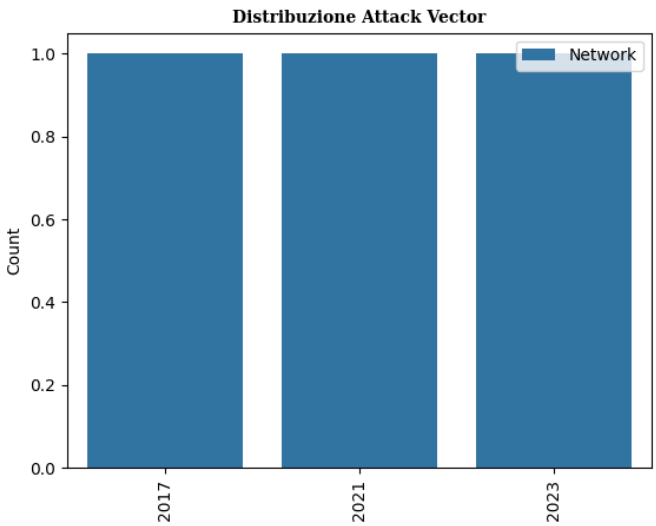


Figura 9.3 - Distribuzione Attack Vector/anno

9.2 Distribuzione delle vulnerabilità catalogate per classe

Riguardo questa seconda analisi, le vulnerabilità verranno classificate sulla base di metriche che il Mitre ha selezionato e considerato più rilevanti. Le metriche che verranno utilizzate prevalentemente in questa fase sono i seguenti:

- Attack Complexity (AC): fa riferimento alla difficoltà della progettazione di un attacco sfruttando una determinata vulnerabilità.

Tipo	Descrizione
Low (L)	Non esistono speciali restrizioni di accesso o circostanze particolari per sfruttare una vulnerabilità.
High (H)	Sfruttare con successo una vulnerabilità non può essere soltanto una casualità, ma può dipendere da fattori che richiedono grande effort da parte dell'attaccante.

Privileges Required (PR): fa riferimento ai privilegi necessari ad un attaccante per poter sfruttare con successo una vulnerabilità.

Tipo	Descrizione
None (N)	L'attaccante non necessita di alcun privilegio sul sistema target per sfruttare una determinata vulnerabilità.
Low (L)	L'attaccante richiede privilegi minimi che forniscono capacità di base all'utente che normalmente potrebbero influenzare solo le impostazioni e i file di proprietà di un utente.
High (H)	L'attaccante richiede privilegi elevati che forniscono un controllo significativo sul componente vulnerabile, consentendo l'accesso alle impostazioni e ai file dell'intero componente.

- **User Interaction (UI):** fa riferimento alla necessità dell'attaccante di essere supportato da altre persone al fine di sfruttare una vulnerabilità.

Tipo	Descrizione
None (N)	Un attaccante può sfruttare la vulnerabilità senza che una persona interagisca direttamente col sistema target.
Required (R)	Un attaccante ha necessità di una persona che effettui una serie di azioni sul sistema target prima che la vulnerabilità possa essere sfruttata.

La prima analisi effettuata riguarda la relazione tra le metriche "Attack Vector" e "Attack Complexity".

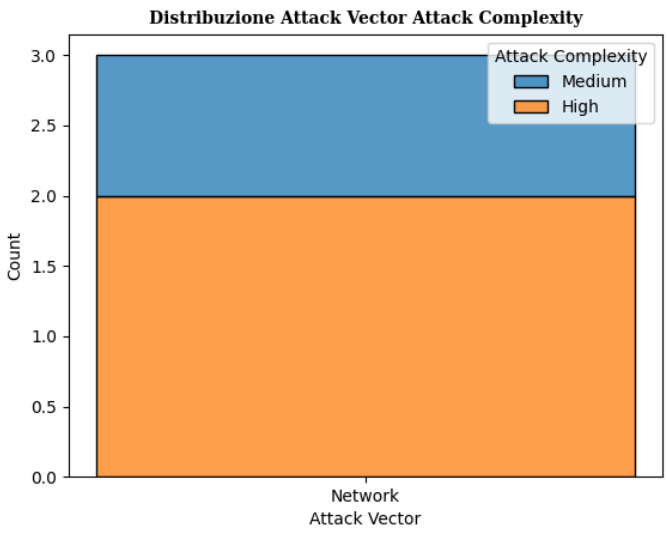


Figura 9.4 - Distribuzione Attack Vector/Attack Complexity

Dal grafico precedente (Figura) possiamo notare come nel caso di Attack Vector Network, l'Attack Complexity può assumere sia un valore MEDIUM che HIGH. Ciò fa riferimento al fatto che per sfruttare queste vulnerabilità il vettore di attacco è sicuramente la rete ma, per un attaccante, sfruttarle richiede un determinato livello di complessità.

Successivamente è stato possibile mettere in relazione la "User Interaction" e l'"Attack Vector":

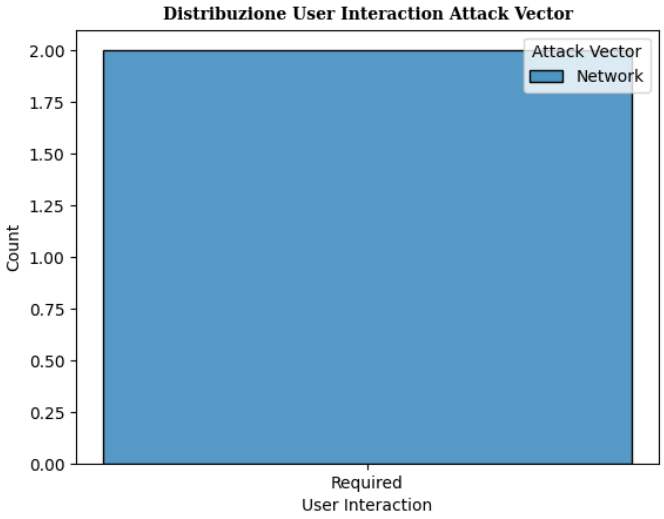


Figura 9.5 - Distribuzione User Interaction/Attack Vector

In base al grafico in Figura , possiamo notare come nel caso di User Interaction Required, l'Attack Vector assume il valore di Network. Questo aspetto implica che l'eventuale attacco viene veicolato attraverso la rete e che è necessario il supporto di uno o più utenti terzi al fine di sfruttare tali vulnerabilità.

9.3 Distribuzione delle vulnerabilità per layer

Una qualsiasi applicazione può essere realizzata mediante diverse architetture. Solitamente una delle architetture più utilizzata è quella che prevede una suddivisione in tre livelli, o layer: presentazione, persistenza e business logic. L'obiettivo di questo paragrafo è vedere come le vulnerabilità si distribuiscono sui vari layer del dataset di analisi.

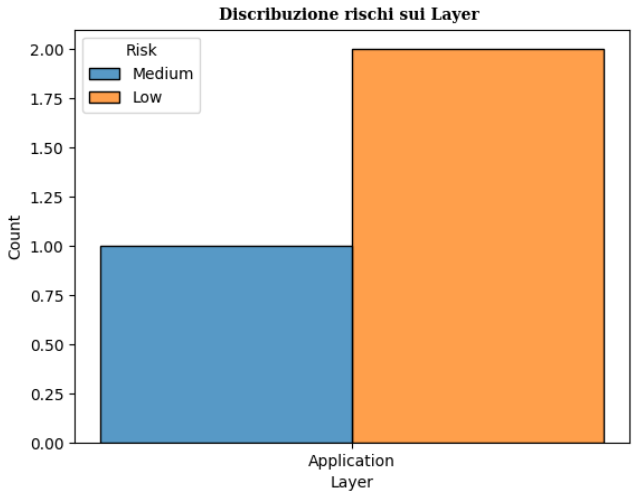


Figura 9.6 - Distribuzione Rischi/Layer

Analizzando il grafico precedente (Figura) appare evidente come le vulnerabilità più frequenti per il layer Application presentano livelli di rischio MEDIUM e LOW. È notevole osservare che tutte le vulnerabilità sono concentrate nel layer dell'applicazione, indicando la sua elevata fragilità. La vulnerabilità di un layer dipende da diversi fattori, tra cui la presenza di exploit che possano sfruttare le vulnerabilità. Se il layer è afflitto da numerose vulnerabilità, soprattutto se sfruttabili tramite exploit, la situazione diventa critica dal punto di vista della sicurezza.

9.4 Iniziative di miglioramento

L'analisi condotta ha rivelato la presenza di vulnerabilità di livello MEDIUM e LOW legate a componenti software. È importante sottolineare che il Vulnerability Assessment ha identificato potenziali vulnerabilità, ma non implica necessariamente che tutte siano attive nella piattaforma. La valutazione del rischio, la difficoltà di sfruttamento e l'aging delle vulnerabilità sono aspetti chiave da considerare. Nota positiva è il fatto che le vulnerabilità rilevate sono recenti, non obsolete, nonostante l'uso di componenti vulnerabili. Altri punti salienti includono la predominanza di vulnerabilità nel layer Application e la necessità di un supporto da terze parti per sfruttare gli exploit. Si consiglia quindi una valutazione iniziale delle componenti software per prevenire l'adozione di elementi vulnerabili nello sviluppo della piattaforma. Un approccio proattivo alla sicurezza delle componenti può contribuire a mitigare i rischi associati alle vulnerabilità.

10. SICUREZZA CON LA BLOCKCHAIN

10.1 Risultati

In questa sezione vengono mostrati i grafici dei tempi misurati con e senza la blockchain e variando il numero di nodi della rete blockchain creata.

Nel primo grafico, sono confrontati i tempi misurati con e senza l'utilizzo della blockchain. Nel caso della scrittura in blockchain i tempi risultano maggiori.

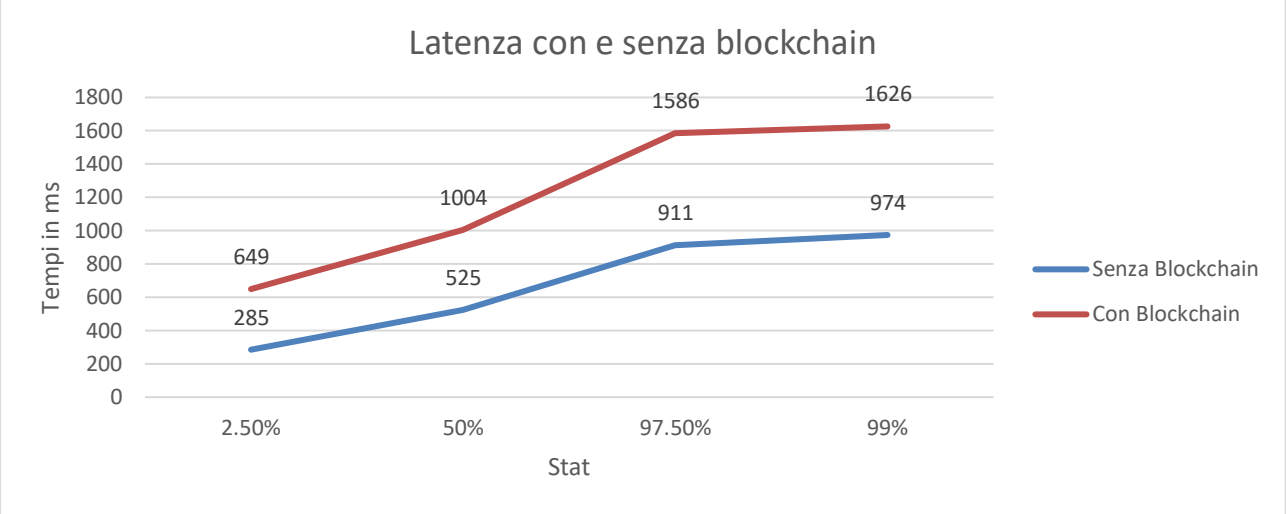


Figura 10.1 - Confronto Tempi con e senza Blockchain

Nel secondo grafico, sono presenti i tempi di latenza misurati al variare del numero di nodi. Dal grafico si nota che i tempi sono maggiori nel caso di due nodi e iniziano a diminuire aumentando il numero di nodi fino a 6. Invece, quando si hanno 8 nodi tali tempi risultano maggiori rispetto a 6 nodi, ma comunque inferiori rispetto al caso di 4 nodi.

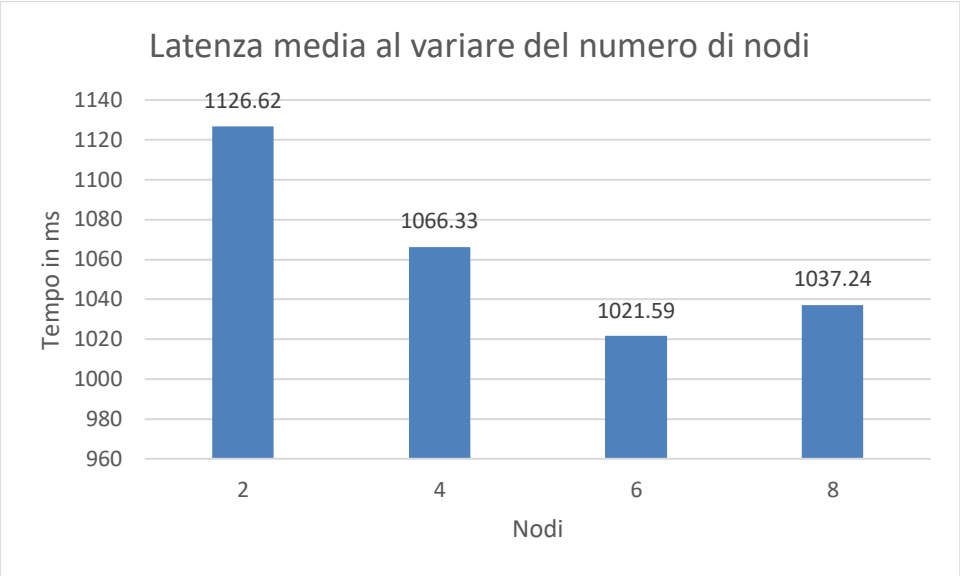


Figura 10.2 – Latenza al variare del numero dei nodi

10.2 Analisi dei risultati

Quando viene misurata la latenza, è importante tenere in considerazione alcuni aspetti che potrebbero influenzare tali valori, come:

1. Latenza di rete: i tempi di risposta possono essere influenzati dalla distribuzione geografica dei nodi e la latenza di rete tra essi.
2. Algoritmo di consenso: L'algoritmo di consenso utilizzato nella rete può influire sulla velocità effettiva delle transazioni e sui tempi di risposta.
3. Prestazioni del nodo: La potenza di calcolo e le risorse (CPU, memoria, velocità del disco) dei singoli nodi possono influenzare la velocità con cui possono elaborare e rispondere alle richieste di lettura.
4. Numero di transazioni: Il carico di lavoro complessivo sulla rete, compreso il numero di transazioni elaborate, può influire sui tempi di risposta.
   All'aumentare del numero di transazioni, potrebbe aumentare anche il carico sulla rete e sui nodi.

Altre considerazioni da fare riguardano il fatto che:

- al crescere delle esecuzioni, crescono anche la quantità di transazioni che vengono salvate nella blockchain.
- al crescere del numero dei nodi, più nodi della rete fanno parte del protocollo di consenso e dovranno interagire tra di loro per decidere chi dovrà validare il blocco contenente le transazioni. Questa procedura può richiedere più tempo.