

5. Design of Experiments

Introduzione

Il “Design of Experiments” è una metodologia sistematica per pianificare test in cui si modificano le variabili di input di un sistema al fine di osservare e identificare le cause dei cambiamenti nell’output.

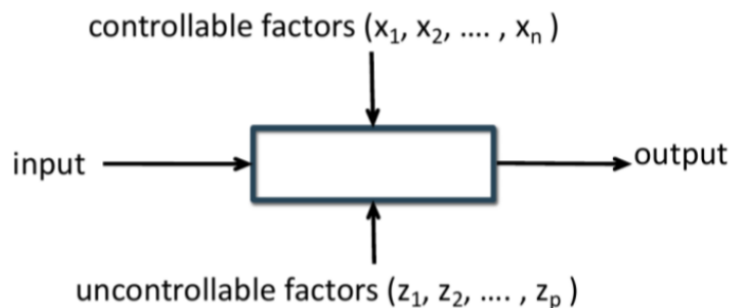


Figura 1

Tra gli obiettivi del DOE ci sono:

- Studiare l’influenza dei fattori controllabili (input) sulla/e uscita/e
- Determinare il/i fattore/i più influente/i
- Studiare le interazioni tra i fattori e la loro influenza sulla/e uscita/e

Questi sono quelli rilevanti per questa sezione del nostro elaborato.

Per una questione di completezza, riportiamo di seguito il formalismo utilizzato durante la trattazione:

- Variabili di risposta: le uscite sulle quali vogliamo controllare l’influenza dei fattori
- Fattori: variabili di input, da scegliere con cura sulla base di ciò che vogliamo valutare
- Livelli: valori che possono assumere i fattori
- Repliche: numero di ripetizione per esperimento

Configurazione

Abbiamo considerato il Web Server Apache HTTP Server 2 come sistema e abbiamo scelto di progettare un **two-factor experiment** con **4 repliche**, dove i fattori sono CTT e Page Type, rispettivamente:

- **CTT**: 25%, 50% e 75% della Usable Capacity (**tre livelli**)
- **Page Type**: dimensione piccola (), media () e grande (), in tutto **tre livelli**

Le **variabili di risposta** scelte sono:

- **Response Time**
- **Latency**
- **Throughput**

Il **numero di esperimenti** totali (osservazioni) ammonta a $3 * 3 * 4 = 36$ per ogni variabile di risposta. Questa sezione dell’elaborato mira ad esporre i risultati dello studio svolto, durante

il quale abbiamo valutato come le prestazioni del sistema (tempo di risposta, latenza e throughput) reagiscono al variare dei parametri di input sopra citati (CTT e dimensione dei file).

Per **generare il carico di lavoro** abbiamo utilizzato **JMeter**, impostando sessioni da 120 secondi con 50 utenti concorrenti e ramp-up period di 25 secondi. A valle dei test, abbiamo trasferito i risultati su JMP, costruendo un piano DOE ad hoc per analizzare le metriche ottenute.

	CTT	Page Type	Response Time ...	Throughput [req/s]	Latency [ms]
1	25% (800)	Light	2,7	13,75	1,93
2	25% (800)	Light	2,54	13,74	1,64
3	25% (800)	Light	2,65	13,74	1,73
4	25% (800)	Light	2,5	13,74	1,61
5	25% (800)	Medium	13,82	13,74	1,61
6	25% (800)	Medium	14,11	13,74	1,68
7	25% (800)	Medium	14,28	13,74	1,68
8	25% (800)	Medium	13,5	13,74	1,79
9	25% (800)	Heavy	28,5	13,74	2,21
10	25% (800)	Heavy	28,88	13,74	2,35
11	25% (800)	Heavy	28,56	13,74	2,21
12	25% (800)	Heavy	27,1	13,75	1,99
13	50% (1600)	Light	2,38	27,08	1,59
14	50% (1600)	Light	2,18	27,08	1,36
15	50% (1600)	Light	2,1	27,08	1,28
16	50% (1600)	Light	2,2	27,08	1,37
17	50% (1600)	Medium	11,73	27,08	1,55
18	50% (1600)	Medium	11,58	27,08	1,5
19	50% (1600)	Medium	11,85	27,08	1,48
20	50% (1600)	Medium	13	27,08	1,63
21	50% (1600)	Heavy	31,2	27,05	2,86
22	50% (1600)	Heavy	30,25	27,03	2,71
23	50% (1600)	Heavy	21,77	27,05	1,94
24	50% (1600)	Heavy	22,85	27,07	1,85
25	75% (2400)	Light	2,63	40,41	1,77
26	75% (2400)	Light	2,69	40,41	1,85
27	75% (2400)	Light	2,38	40,41	1,62
28	75% (2400)	Light	2,41	40,41	1,53
29	75% (2400)	Medium	12,43	40,4	1,55
30	75% (2400)	Medium	12,36	40,4	1,75
31	75% (2400)	Medium	14,42	40,4	2,08
32	75% (2400)	Medium	13,08	40,41	2,29
33	75% (2400)	Heavy	22,08	40,39	1,66
34	75% (2400)	Heavy	20,06	40,4	1,41
35	75% (2400)	Heavy	19,23	40,39	1,34
36	75% (2400)	Heavy	19,61	40,39	1,32

Figura 2

Analisi di Importanza

Prima di riportare i risultati e le tecniche adottate per il calcolo dell'importanza dei fattori, è importante osservare che l'importanza non è un fattore statistico ma un parametro soggettivo. Infatti, siamo noi a scegliere se un fattore è importante o meno a seconda dei risultati ottenuti in relazione alla Sum of Squares Total (SST).

La tecnica adottata per valutare l'importanza dei fattori è la stima del modello sul piano DOE. Abbiamo effettuato una regressione lineare con l'obiettivo di minimizzare l'MSE ed enfasi sul leverage degli effetti, ottenendo i risultati riportati in foto.

Analisi della varianza				
Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	8	3207,6619	400,958	130,3485
Errore	27	83,0532	3,076	Prob > F
C. totale	35	3290,7151		<,0001*

Stime dei parametri				
---------------------	--	--	--	--

Test degli effetti					
Origine	Nparm	DF	Somma dei quadrati	Rapporto F	Prob > F
CTT	2	2	53,4684	8,6911	0,0012*
Page Type	2	2	3058,0227	497,0708	<,0001*
CTT*Page Type	4	4	96,1707	7,8161	0,0003*

Figura 4.a. Analisi degli effetti per Y=Response Time

Analisi della varianza				
Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	8	4264,5367	533,067	13084374
Errore	27	0,0011	4,074e-5	Prob > F
C. totale	35	4264,5378		<,0001*

Stime dei parametri				
---------------------	--	--	--	--

Test degli effetti					
Origine	Nparm	DF	Somma dei quadrati	Rapporto F	Prob > F
CTT	2	2	4264,5336	52337458	<,0001*
Page Type	2	2	0,0017	20,5227	<,0001*
CTT*Page Type	4	4	0,0014	8,3523	0,0002*

Figura 3.b. Analisi degli effetti per Y=Throughput

Analisi della varianza				
Origine	DF	Somma dei quadrati	Media quadratica	Rapporto F
Modello	8	3,3639000	0,420488	7,6530
Errore	27	1,4835000	0,054944	Prob > F
C. totale	35	4,8474000		<,0001*

Stime dei parametri				
---------------------	--	--	--	--

Test degli effetti					
Origine	Nparm	DF	Somma dei quadrati	Rapporto F	Prob > F
CTT	2	2	0,2146167	1,9530	0,1614
Page Type	2	2	0,9230167	8,3995	0,0015*
CTT*Page Type	4	4	2,2262667	10,1296	<,0001*

Figura 5.c. Analisi degli effetti per Y=Latency

L'analisi si è concentrata sui dati derivanti dall'**analisi della varianza** (ANOVA) e dal **test degli effetti**. Questi strumenti sono stati fondamentali per quantificare il peso (*importanza*) dei singoli fattori e degli errori su ciascuna variabile di risposta. Il metodo utilizzato per determinare l'importanza si basa sulla **Sum of Squares** (SS), e in un two-factor experiment la varianza totale viene esplicitata come segue:

$$SST = SSA + SSB + SSAB + SSE$$

Dove:

- **SST**: Varianza totale
- **SSA & SSB**: Varianze dovute ai singoli fattori
- **SSAB**: Varianza dovuta all'interazione tra i due fattori
- **SSE**: Sum of Squares Error (Errore residuo)

L'importanza si calcola normalizzando la singola Sum of Square rispetto a quella totale:

$$Imp(B) = \frac{SSB}{SST}$$

Riportiamo di seguito in una tabella il risultato dell'Analisi di Importanza per ogni variabile di uscita.

Fattore\Var. di risposta	Response Time	Latency	Throughput
CTT	$\frac{SSCTT}{SST} = \frac{53.4684}{3290.7151} = 1.6\%$	$\frac{SSCTT}{SST} = \frac{0.2146}{4.8474} = 4.4\%$	$\frac{SSCTT}{SST} = \frac{4264.5367}{4264.5378} = 99.99\%$
Page Type	$\frac{SSPT}{SST} = \frac{3058.0227}{3290.7151} = 92.9\%$	$\frac{SSPT}{SST} = \frac{0.9230}{4.8474} = 19.1\%$	< 0.01%
CTT * Page Type	$\frac{SSINT}{SST} = \frac{96.1707}{3290.7151} = 3.0\%$	$\frac{SSINT}{SST} = \frac{2.2263}{4.8474} = 45.9\%$	< 0.01%
Errore	$\frac{SSE}{SST} = \frac{83.0532}{3290.7151} = 2.5\%$	$\frac{SSE}{SST} = \frac{1.4835}{4.8474} = 30.6\%$	< 0.01%

Tabella 1. Risultato dell'Analisi di importanza per ogni fattore e per l'errore, per le singole uscite.

P.S: Qui vince un "difetto" nella selezione del tipo di pagine per l'analisi del Throughput (CTT troppo piccolo); procederemo comunque con l'analisi.

I risultati ottenuti sono i seguenti:

1. Per il Response Time, è il Page Type ad avere la maggiore importanza.
2. Per la Latency, invece, l'interazione tra i fattori è sicuramente la più importante, senza però poter trascurare l'impatto dell'errore.
3. Per il Throughput, invece, il CTT è nettamente il più importante tra tutti.

Analisi di Significatività

Per procedere a tale analisi, è stato necessario utilizzare dei test statistici, in quanto la **significatività** è un **fattore statistico**! Essa è il **contributo alla variazione rispetto all'errore**. Tuttavia, è necessario **determinare normalità dei residui delle variabili di risposta** (errori sulla stima del modello dovuti alle repliche degli esperimenti) e **omoschedasticità** di questi **rispetto ai fattori**, che sono le assunzioni fondamentali da verificare per individuare il corretto metodo ANOVA da applicare.

Di seguito sono riportate le colonne dei residui delle variabili di risposta:

Residuo Response Time [elapsed] [ms]	Residuo Throughput [req/s]	Residuo Latency [ms]
0.1025	0.0075	0.2025
-0.0575	-0.0025	-0.0875
0.0525	-0.0025	0.0025
-0.0975	-0.0025	-0.1175
-0.1075	1.24345e-14	-0.08
0.1825	1.24345e-14	-0.01
0.3525	1.24345e-14	-0.01
-0.4275	1.24345e-14	0.1
0.24	-0.0025	0.02
0.62	-0.0025	0.16
0.3	-0.0025	0.02
-1.16	0.0075	-0.2
0.165	-7.10543e-15	0.19
-0.035	-7.10543e-15	-0.04
-0.115	-7.10543e-15	-0.12
-0.015	-7.10543e-15	-0.03
-0.31	-3.55271e-15	0.01
-0.46	-3.55271e-15	-0.04
-0.19	-3.55271e-15	-0.06
0.96	-3.55271e-15	0.09
4.6825	-3.55271e-15	0.52
3.7325	-0.02	0.37
-4.7475	-3.55271e-15	-0.4
-3.6675	0.02	-0.49
0.1025	0	0.0775
0.1625	0	0.1575
-0.1475	0	-0.0725
-0.1175	0	-0.1625
-0.6425	-0.0025	-0.3675
-0.7125	-0.0025	-0.1675
1.3475	-0.0025	0.1625
0.0075	0.0075	0.3725
1.835	-0.0025	0.2275
-0.185	0.0075	-0.0225
-1.015	-0.0025	-0.0925
-0.635	-0.0025	-0.1125

Figura 4. Sezione della tabella sul sw JMP relativa ai residui delle variabili di risposta.

Analisi Normalità

Per procedere all'analisi della normalità, abbiamo visualizzato le distribuzioni dei residui per effettuare un primo test visivo tramite plot quantile-quantile.

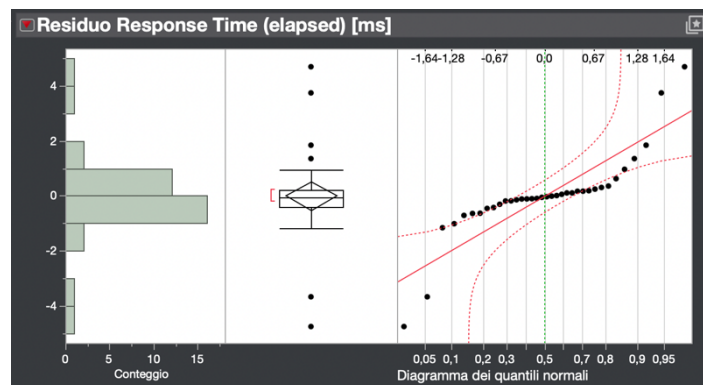


Figura 5.a. Distribuzione del residuo di Response Time

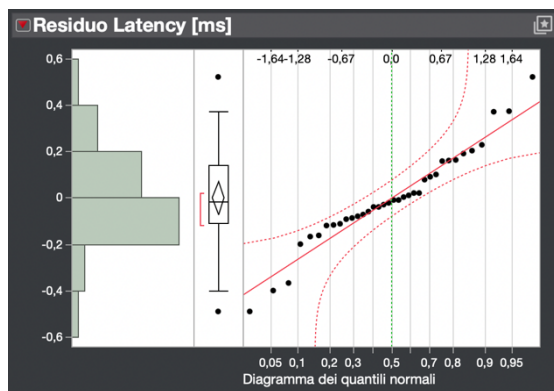


Figura 5.b. Distribuzione del residuo di Latency

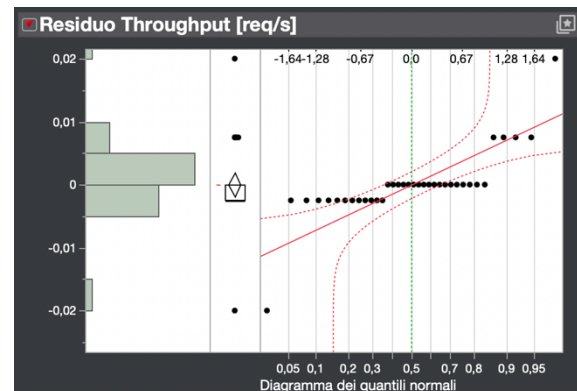


Figura 5.c. Distribuzione del residuo di Throughput

Dall'analisi visuale si evince che:

1. **Response Time:** La normalità è rispettata nella parte centrale del plot quantile, ma si notano forti deviazioni sulle code che escono dalle bande di confidenza
2. **Latency:** È la distribuzione che meglio approssima la normalità, essendo i residui ben allineati sulla retta di riferimento e contenuti quasi interamente nelle bande di confidenza
3. **Throughput:** Deviazioni marcate, specialmente agli estremi, indicano una evidente mancanza di normalità.

Per chiarire il dubbio sul Response Time, è possibile applicare il **Test di Shapiro-Wilk**. Di seguito sono riportati i risultati del Test per Response Time e Latency (inutile per Throughput).

Test della bontà di adattamento				Test della bontà di adattamento			
	W	Prob<W			W	Prob<W	
Shapiro-Wilk	0,789006	<,0001*		Shapiro-Wilk	0,967212	0,3544	
		A ²	P-value simulato			A ²	P-value simulato
Anderson-Darling	3,285416	<,0001*		Anderson-Darling	0,5518546	0,1412	

Figura 6. Test Shapiro-Wilk rispettivamente su residuo Response Time e su residuo Latency

Sia il test visivo sia quello di Shapiro-Wilk ci dicono che per il **residuo di Response Time non ho la normalità**, mentre per la **Latency sì!**

Analisi Omoschedasticità

Prima di procedere all’analisi, è importante osservare che da questa possiamo **escludere le variabili di uscita** sui quali **residui** abbiamo determinato la **non normalità**, ovvero Response Time e Throughput, poiché per queste basta applicare il Test di Wilcoxon/Kruskal-Wallis alle stesse variabili di risposta per ogni fattore.

L’unica oggetto di questa analisi, dunque, è la **Latency**.

Riportiamo di seguito i risultati dei test di verifica dell’omoschedasticità per la Latency rispetto ai fattori.

Test	Rapporto F	Num DF	Den DF	Prob > F	Test	Rapporto F	Num DF	Den DF	Prob > F
O'Brien[.5]	2,3591	2	33	0,1103	O'Brien[.5]	2,9651	2	33	0,0654
Brown-Forsythe	1,7926	2	33	0,1824	Brown-Forsythe	2,5832	2	33	0,0907
Levene	2,2251	2	33	0,1240	Levene	2,5464	2	33	0,0937
Bartlett	3,7926	2	.	0,0225*	Bartlett	3,6701	2	.	0,0255*

Figura 7. Test per verificare l’omoschedasticità di Latency rispetto a CTT e Page Type

Dai test, risulta evidente che abbiamo omoschedasticità per Latency per entrambi i fattori, ergo è necessario applicare l’F-Test per studiare la significatività dei fattori (Figura 3.c).

Per quanto riguarda Response Time e Throughput, di seguito sono riportati i risultati del Test di Wilcoxon/Kruskal-Willis per ogni fattore.

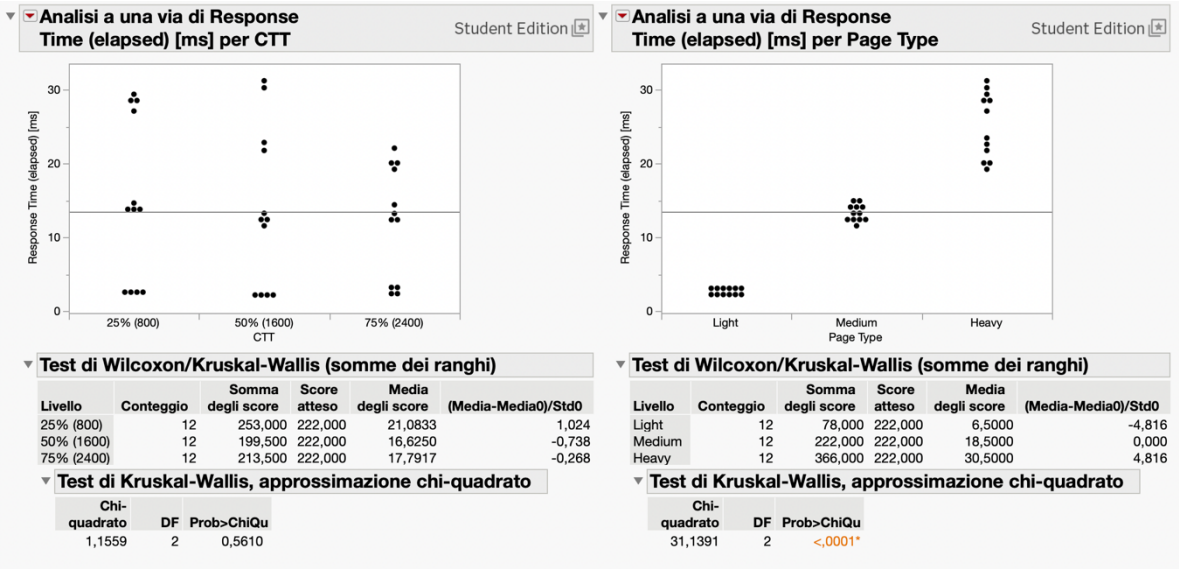


Figura 8.a. Test di Wilcoxon/Kruskal-Willis per Response Time, rispettivamente con CTT e Page Type

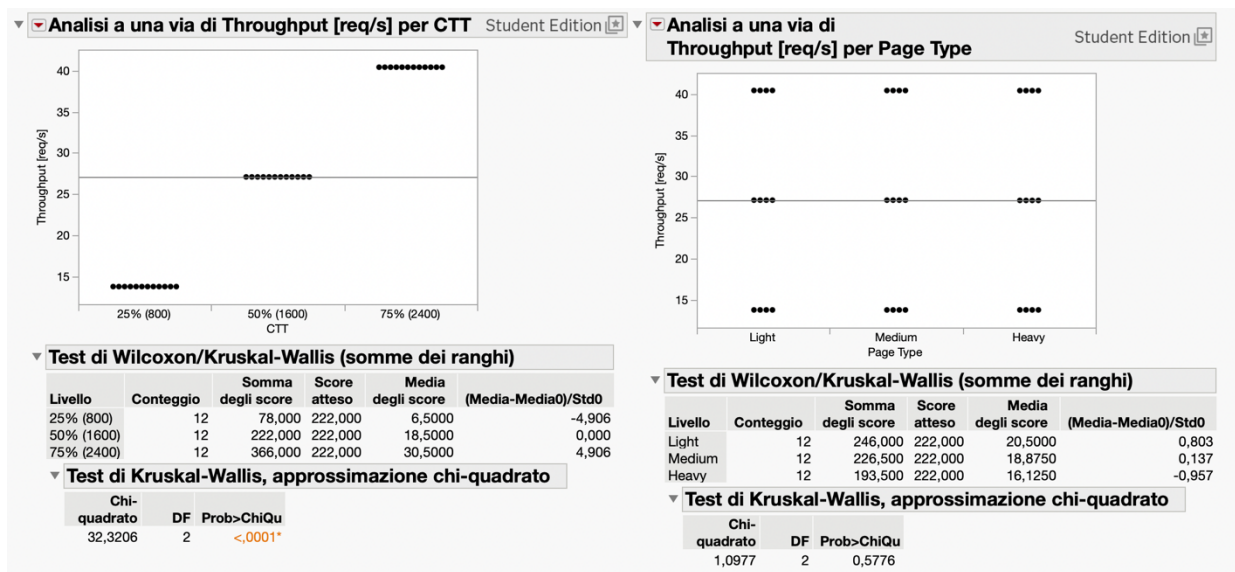


Figura 8.b. Test di Wilcoxon/Kruskal-Wallis per Throughput, rispettivamente con CTT e Page Type

Conclusioni

Sulla base dei risultati ottenuti, possiamo concludere che:

- **Response Time:** il fattore più importante risulta essere Page Type, il quale risulta essere anche significativo, mentre CTT non risulta essere né significativo né importante.
- **Latency:** l'interazione tra i due fattori risulta essere sia importante sia significativa, mentre l'errore risulta essere importante e il Page Type significativo.
- **Throughput:** Il fattore CTT è indiscussamente il più importante e significativo.