

Computação Experimental - Lista 3

Antônio Henrique de Moura Rodrigues, Yan Victor dos Santos, Samuel Andrade do Couto, Victor Neves Martorelli¹

Professor: Genaína Nunes Rodrigues²

26 de junho de 2018

Listagem de Exercícios dos Slides

Questão 1

Seu melhor amigo acabou de desenvolver um algoritmo revolucionário, chamado Xulambis, para o reconhecimento de ações humanas. Visando uma posterior comparação com o algoritmo estado-da-arte Zambis, ele fez um estudo para estimar a sensibilidade do algoritmo Xulambis, em termos de taxa de reconhecimento (%), aos seus dois parâmetros principais, α e β . Com base nos resultados de um projeto fatorial (com uma única replicação), mostrado nas 3 primeiras colunas da tabela abaixo, ele concluiu que o algoritmo Xulambis é muito mais sensível ao parâmetro α que ao parâmetro β , obtendo um desempenho melhor (maior taxa de reconhecimento) para valores mais altos de α .

Fator α	Fator β	Projeto Fatorial (uma replicação)	Replicações extras	
			2 ^a replicação	3 ^a replicação
-1	-1	70%	74%	64%
1	-1	80%	90%	84%
-1	1	45%	34%	43%
1	1	65%	55%	64%

- a. Com base apenas na única replicação inicialmente feita (3^a coluna da tabela), você concorda com a conclusão do seu amigo quanto à sensibilidade de Xulambis aos dois parâmetros considerados?

Resposta: Não, é notável que o algoritmo Xulambis é mais sensível ao parâmetro β (com 61%) que ao parâmetro α (com 34%).

¹ ahmoura@live.com, yanvictor_ds@hotmail.com, samuelcoouto@hotmail.com, victormartorelli@outlook.com

² genaina@cic.unb.br

qA	0.075
qB	-0.1
qAB	0.025
q0	0.65
SSA	0.0225
SSB	0.04
SSAB	0.0025
SST	0.065
SSA/SST	0.3461538462
SSB/SST	0.6153846154
SSAB/SST	0.03846153846

b. Com base apenas na única replicação inicialmente feita (3^a coluna da tabela), você concorda com a conclusão do seu amigo de que o desempenho de Xulambis é melhor para valores mais altos de α ?

Resposta: Discordamos. De acordo com os resultados da questão anterior, como o Projeto Fatorial apresenta 61% de desempenho para o fator β , comparados aos 34,6% de desempenho para o fator α , não é possível afirmar que ele terá desempenho maior para α .

c. Conhecendo bem a variabilidade inerente ao problema de reconhecimento de ações humanas, você pede ao seu amigo que execute pelo menos mais 2 replicações para cada cenário. Os resultados são mostrados na 4^a e na 5^a colunas da tabela. Suas conclusões em (a) e (b) se mantêm?

Resposta: Se baseando na imagem abaixo, as colunas em amarelo representam os cálculos para repetições. Analisando os novos impactos $SSA/SST = 29,8\%$ e $SSB/SST = 62,3\%$, as afirmações em 1.1 e 1.2 são mantidas.

d. Qual a porcentagem total da variação nos dados, observada no projeto com 3 replicações, você consegue explicar com a variação dos fatores?

Resposta: Basta analisar a imagem da tabela em 1.3.

e. Ainda considerando o projeto com 3 replicações, o efeito devido à interação entre os fatores α e β é significativo com 95% de confiança? Qual a maior confiança que você pode atribuir à afirmativa de que a interação entre os dois fatores é significativa?

Resposta: De acordo com a imagem abaixo, pode-se observar que o intervalo de confiança de AB, com 95%, contém o ZERO neste intervalo. Portanto, o efeito devido à interação entre os fatores α e β não são significativos. A maior confiança encontrada foi de 50%.

qAr1	qAr2	qAr3	qAr	qA	SSE	t-stud(95%)	t-stud(50%)
0.3	0.37	0.41	0.36	0.09	0.023066666667	1.8595	0.706
qBr1	qBr2	qBr3	qBr	qB	t*Sab (95%)		t*Sab (50%)
-0.4	-0.75	-0.41	-0.52	-0.13	2.882391617		1.09436326
qABr1	qABr2	qABr3	qABr	qAB	0.05369667898		
0.1	0.05	0.01	0.053333333333	0.013333333333	I.C.ab (95%)		I.C.ab (50%)
SSA	0.0972		SSA/SST	0.2988929889	Sab	Min= -1.549058284	Min= 0.2389700735
SSB	0.2028		SSB/SST	0.6236162362	Max= 4.21572495	Max= 2.427696593	
SSAB	0.002133333333		SSAB/SST	0.006560065601	DF		
Y =SST	0.3252		SSE/SST	0.07093070931	8		

Questão 2

Um sistema foi avaliado quanto ao impacto de 4 parâmetros – A, B, C e D. Para tanto, um projeto fatorial 2^{4-1} foi realizado e os resultados são mostrados abaixo. O subscrito representa o nível de cada fator (1 para inferior, e 2 para superior).

	C ₁ D ₁	C ₁ D ₂	C ₂ D ₁	C ₂ D ₂
A ₁ B ₁	-	40	15	-
A ₁ B ₂	-	20	10	-
A ₂ B ₁	100	-	-	30
A ₂ B ₂	120	-	-	50

a. Qual o polinômio gerador deste projeto? Liste todos os confoundings , caso existam, e discuta o que eles representam em termos do impacto na precisão do modelo.

Resposta: O polinômio gerador é I=ACD. A única forma de tornar a tabela correta foi fazendo D = AC. Isto impediu de saber o verdadeiro impacto em AC. Os Confoundings podem ser vistos na imagem abaixo:

Confounding
I = ACD
A = CD
B = ABCD
C = AD
AB = BCD
D = AC
BC = ABD
ABC = BD

- b. Quantifique todos os efeitos e as porcentagens de variação explicada por cada um deles.

Resposta:

A	B	C	D	E	F	G	H	I
A	B	C	AB	AC = D	BC	ABC	I	Y
-1	-1	-1	1	1	1	-1	1	40
-1	-1	1	1	-1	-1	1	1	15
-1	1	-1	-1	1	-1	1	1	20
-1	1	1	-1	-1	1	-1	1	10
1	-1	-1	-1	-1	1	1	1	100
1	-1	1	-1	1	-1	-1	1	30
1	1	-1	1	-1	-1	-1	1	120
1	1	1	1	1	1	1	1	50

qA	qB	qC	qAB	qD	qBC	qABC	q0
26.875	1.875	-21.875	8.125	-13.125	1.875	-1.875	48.125
SSA	SSB	SSC	SSAB	SSD	SSBC	SSABC	SS0
5778.125	28.125	3828.125	528.125	1378.125	28.125	28.125	18528.125
49.82%	0.24%	33.01%	4.55%	11.88%	0.24%	0.24%	100.00%

- c. Você sugeriria um projeto melhor? Qual e por quê?

Resposta: Como podemos observar na imagem da tabela anterior, os impactos calculados mostraram um valor alto para os fatores A e C. Portanto, ao escolhermos D=AC, estamos provavelmente perderemos informações que podem ser importantes. Portanto, seria correto sugerir um outro projeto, escolhendo D=BC ou D=ABC que possuem impactos baixos.

Questão 3

O tempo de execução de um algoritmo foi avaliado em função de 3 fatores, a CPU (fator A), o tamanho da memória (fator B) e o número de discos (fator C). Foram executadas três

replicações para cada configuração considerada, conforme mostrado na tabela abaixo. A tabela também mostra alguns resultados agregados da análise das observações. Responda:

Experimento	I	A	B	C	Tempo de execução(ms)			
					t_1	t_2	t_3	\bar{t}
1	1	-1	-1	1	98	100	102	100
2	1	1	-1	-1	245	249	256	250
3	1	-1	1	-1	45	54	52	50
4	1	1	1	1	300	301	299	300
Total	700	400	0	100				
Total/4	175	100	0	25				

$$SSE = 117$$

- a. Qual o polinômio gerador deste projeto? Liste todos os confoundings, caso existam, e discuta o que eles representam em termos do impacto na precisão do modelo?

Resposta: Fazendo C=AB, temos que o polinômio gerador é I=ABC.

Os confoundings são: A=BC, B=AC, C=Ab.

A	B	C	AB	AC	BC	ABC	I	Y1	Y2	Y3	Y
-1	-1	1	1	-1	-1	1	1	98	100	102	100
1	-1	-1	-1	-1	1	1	1	245	249	256	250
-1	1	-1	-1	1	-1	1	1	45	54	52	50
1	1	1	1	1	1	1	1	300	301	299	300
qA	qB	qC	qAB	qAC	qBC	qABC	q0	EY1	EY2	EY3	Confounding
99.91666667	0.08333333333	24.91666667	24.91666667	0.08333333333	99.91666667	175.08333333	175.08333333	-2	0	2	I = ABC
SSA	SSB	SSC	SSAB	SSAC	SSBC	SSABC	SS0	-5	-1	6	A = BC
119800.0833	0.08333333333	7450.083333	7450.083333	0.08333333333	119800.0833	367850.0833	367850.0833	-5.333333333	3.666666667	1.666666667	B = AC
					SSE	116.666667	SST	127366.9167	MSE	14.58333333	
qA	qB	qC	qAB	qAC	qBC	qABC	q0	Se	3.818813079		
-	+	-	+	-	+	-	+	DF	8		
Intervalos de Confiança								t-student (90%)	1.86		
97.8662094	101.9671239	-1.967123933	2.133790599	22.8662094	26.96712393	173.0328761	177.1337906	Sq	1.10239638		
Estimativa								t-student * Sq	2.050457266		
95.37727659	104.4560567	-4.456056748	4.622723415	20.37727659	20.37727659	170.5439433	179.6227234	Sy	2.213842813		
								IC 90%	4.539390082		

A tabela acima mostra que o fator B não tem impacto. Fazer C=AB retira o real impacto do fator C, e ainda usamos B como fator na análise sendo que este não tem nenhum impacto.

- b. Quais são as estimativas e desvios padrões para cada efeito? Apresente o seu modelo para estimativa do tempo de execução do algoritmo em questão.

Resposta: As estimativas e desvios padrões para cada efeito é apresentado na tabela da questão 3.1.

- c. Qual a porcentagem da variação explicada por cada efeito e pelo modelo? Você está satisfeito com seu modelo? Justifique. Qual o seu próximo passo na avaliação do

impacto dos fatores CPU, memória e disco no desempenho do algoritmo em questão?

Resposta: Os cálculos podem ser vistos na tabela da questão 3.1. Uma maneira de melhorar o modelo, seria não analisar B, uma vez que seu impacto é nulo.

d. Com uma confiança de 90%, você poderia dizer que o efeito A é significativamente diferente do efeito C? Justifique.

Resposta: Sim, uma vez que seus intervalos não incluem zero, portanto se tornam diferentes.

e. Você pode propor um melhor projeto de experimentos, mantendo o mesmo custo, que o proposto acima? Justifique

Resposta: Um projeto melhor seria escolher o fator C, e ignorar o fator B na hora de retirar fatores, porque o fator B não tem nenhum impacto, e C seria calculado sem perder informação.

Exercícios do Livro do Jain, capítulo 20: 20.1

Questão 20.1

For a single-factor design, suppose we want to write an expression for \pm_j in terms of y_{ij} 's:

$$\pm_j = a_{11j}y_{11} + a_{12j}y_{12} + \dots + a_{raj}y_{ra}$$

What are the values of the $a_{..j}$'s? From the preceding expression, the error in \pm_j is seen to be

$$e\pm_j = a_{11j}e_{11} + a_{12j}e_{12} + \dots + a_{raj}e_{ra}$$

Assuming errors e_{ij} are normally distributed with zero mean and variance \tilde{A}_e^2 , write an expression for the variance of $e\pm_j$. Verify that your answer matches that in Table 20.5.

Resposta: