Modelagem da Variância em Experimentos Não-Replicados

Flávio Fogliatto, *Ph.D.*

Panorâmica da abordagem proposta por Ribeiro, Fogliatto & Caten (2001)

- Considere um experimento fatorial com fatores testados a l níveis e sem replicações:
  - ✓ **Situação bastante comum já que replicações são usualmente caras e não realizadas na prática**
- Suponha ser possível determinar um modelo de regressão relacionando observações da resposta j aos níveis dos fatores de controle:
  - ✓ **Usando predições a partir do modelo, determinam-se os resíduos para cada tratamento:**
    - **ou seja, a diferença entre as observações reais e preditas da resposta *j***
    - **Os resíduos dão uma medida da variância no experimento que não pode ser explicada pelo modelo de regressão**

Panorâmica (*Continuação)*

- Deseja-se verificar se os resíduos, dentro de um determinado nível de um fator de controle, diferem significativamente dos resíduos obtidos nos outros níveis do fator:
  - ✓ quando este for o caso, a variância das respostas, estimada pelos resíduos, pode ser considerada como função do fator de controle em questão

O modelo de regressão para a variância da resposta j é determinado em 6 etapas

1. Determine um modelo de regressão relacionando as observações da resposta j aos fatores de controle experimentais
2. Para um dado fator de controle k plote os resíduos nos níveis extremos:
   - objetivo é identificar diferenças nos resíduos nos dois níveis:
     - caso isso ocorra, vá ao passo 3;
     - caso contrário, é pouco provável a obtenção de um modelo singificativo para a variância da resposta j

Passo 3

- Calcule as variâncias amostrais para os resíduos em cada nível extremo de *k:*
  - ✓ Notação: $s^2_{k,(-1)}$ e $s^2_{k,(+1)}$
- Deseja-se testar a hipótese nula H_0: variâncias são iguais nos níveis extremos de k:
  - ✓ Para tanto, monta-se um teste F da seguinte maneira: $F = \dfrac{s^2_{k,(s)}}{s^2_{k,(b)}}$
  - ✓ para um par de níveis extremos s e b de k ($s \neq b$), usando a maior das razões entre variâncias
  - ✓ Se $F \geq F_{\alpha/2}$, as variâncias não são iguais nos dois níveis (use $\alpha \leq$ 0,05), devendo ser modeladas como função do fator k
- Repita os passos 2 e 3 para todos os fatores de controle

Passo 4

- Liste os fatores de controle que afetam a variância da resposta *j* bem como suas variâncias amostrais nos níveis extremos
- Organize os resultados em uma tabela:
  - ✓ Veja exemplo com *M* fatores e níveis extremos codificados como -1 e +1 no próximo slide

Variâncias amostrais

Fatores de controle				
$k=1$	$k=2$	…	$k=M$	*Variâncias amostrais*
-1	0	…	0	$s^2_{1,(-1)}$
+1	0	…	0	$s^2_{1,(+1)}$
0	-1	…	0	$s^2_{2,(-1)}$
0	+1	…	0	$s^2_{2,(+1)}$
⋮	⋮	⋱	⋮	⋮
0	0	…	-1	$s^2_{M,(-1)}$
0	0	…	+1	$s^2_{M,(+1)}$

Passos 5 e 6

5. Obtenha um modelo de regressão relacionando a variância amostral aos níveis extremos dos fatores controláveis:
   - ✓ Esse será o modelo para a variância da resposta *j*
6. Repita os passos 1 → 5 para as demais variáveis de resposta

Alguns comentários

- Procedimento não considera termos de alta ordem (interações e efeitos quadráticos) ao modelar a variância:
  - ✓ É por isso que só se verifica a variância amostral em níveis extremos dos fatores de controle
- Benefícios da inclusão de termos de alta ordem são dúbios:
  - ✓ Eles tendem a não ser significativos na prática
  - ✓ Estimativas de mais alta ordem demandam uma maior quantidade de dados, normalmente não disponíveis em projetos fatoriais
  - ✓ a inclusão de termos de alta ordem demandaria um procedimento mais complexo de ajuste do modelo, o que não é o objetivo aqui

Mais comentários

- Abordagem é similar àquela proposta por Box & Meyer (1986), mas desenvolvida independentemente:
  - ✓ Dispersion effects from fractional designs. *Technometrics*, **28** (1), 19-27, 1986.
- Principais diferenças:
  - ✓ Abordagem de Box & Meyer permite incorporar interações nos modelos para variância
  - ✓ determinação dos termos significativos a serem incluídos nos modelos é feita visualmente, através de papéis de probabilidade
  - ✓ Abordagem é restrita a fatoriais 2^k

Estudo de caso

- Desenvolvimento de composto de borracha usado na manufatura de pneus
- Formulação atual baseada em especificações de clientes
- Experimento incluiu 27 tratamentos onde composto foi obtidos com diferentes formulações e condições de processo:
  - ✓ Ingredientes incluiam enxofre e negro de fumo
  - ✓ Condições de processo relacionados a tempo de mistura
- Tratamentos avaliados com relação a dez variáveis de resposta:
  - ✓ Algumas respostas são densidade, dureza e reologia
- Objetivo: identificar ajuste para fatores de controle que resulte em baixo custo, respostas próximas ao alvo, com variância e sensibilidade mínima

Respostas

Resposta	Import. relativa	Tipo	Valor atual	Alvo	Limite inferior	Limite superior
Y_1	2	Nominal	8.0	8.5	7.93	9.07
Y_2	2	Nominal	85	85	74.2	95.8
Y_3	3	Menor	220	210	-	232.7
Y_4	3	Nominal	27	30	27.02	32.98
Y_5	4	Nominal	62	62	59.49	64.51
Y_6	4	Nominal	1.137	1.13	1.125	1.135
Y_7	5	Menor	80	65	-	78
Y_8	4	Maior	1300	1400	1231.69	-
Y_9	3	Maior	1900	2400	2328,02	-
Y_{10}	4	Maior	500	530	496.42	-

Fatores de controle

Fator controle	Ajuste atual	Intervalo de ajuste Min	Max	Flexibilidade
X_1	4.0	3.0	5.0	10
X_2	6.5	5.5	6.5	10
X_3	19.5	18	22	10
X_4	3.5	2.0	4.0	10
X_5	8	8	12	10

Escolha do design - *diretrizes*

		Numero de fatores de controle	
		Pequeno	*Grande*
Alta Flexibilidade	***Linear***	Fatorial 2^k	Fatorial 2^k fracionado Plackett-Burman
	Não-linear	Fatorial 3^k CCD	Fatorial 3^k fracionado CCD fracionado
Baixa Flexibilidade	***Linear***	Como acima. Use blocagem para fatores de controle pouco importantes.	
	Não-linear	Use estratégia split-plot para fatores de controle importantes.	

Design do estudo

- Dado o grande nº de fatores, sua alta flexibilidade e a possibilidade de efeitos não-lineares, escolheu-se um CCD fracionado em três blocos:
  - ✓ Primeiros dois blocos (18 tratamentos) correspondem a um fatorial fracionado 2^{5-1} com dois pontos de centro adicionados para checar não-leariedade, a qual resultou significativa:
    - De acordo com especialistas, somente fatores X_1 e X_2 poderiam apresentar efeitos não-lineares
  - ✓ Trats. 24 a 27 correspondem a porção estrela de um CCD com 2 fatores ($\alpha = 1$):
    - eles foram rodados em um 3º bloco com 5 pontos de centro (19 a 23)
    - Fatores X_3, X_4, e X_5 mantidos em níveis centrais

Matriz experimental

Trt	Bloco	X1	X2	X3	X4	X5	Trt	Bloco	X1	X2	X3	X4	X5
1	1	-1	-1	-1	1	-1	15	2	1	-1	-1	1	1
2	1	-1	-1	-1	-1	1	16	2	1	1	1	-1	-1
3	1	-1	1	1	1	-1	17	2	1	1	1	1	1
4	1	-1	1	1	-1	1	18	2	0	0	0	0	0
5	1	1	1	-1	1	-1	19	3	-1	0	0	0	0
6	1	1	1	-1	-1	1	20	3	1	0	0	0	0
7	1	1	-1	1	1	-1	21	3	0	-1	0	0	0
8	1	1	-1	1	-1	1	22	3	0	1	0	0	0
9	1	0	0	0	0	0	23	3	0	0	0	0	0
10	2	-1	1	-1	-1	-1	24	3	0	0	0	0	0
11	2	-1	1	-1	1	1	25	3	0	0	0	0	0
12	2	-1	-1	1	-1	-1	26	3	0	0	0	0	0
13	2	-1	-1	1	1	1	27	3	0	0	0	0	0
14	2	1	-1	-1	-1	-1							

9 de 10 respostas apresentaram variância modelável

1	Med.	$\hat{Y}_1 = 7.10 + 1.08X_1 + 0.64X_1^2 + 1.11X_2 + 0.54X_2^2 + 0.42X_4 + 0.256X_1X_2$ (0.91)
	Var.	$\hat{\sigma}_1 = 0.362 + 0.186X_1 + 0.118X_2$
2	Med.	$\hat{Y}_2 = 74.62 - 2.33X_1 - 6.26X_2^2$ (0.39)
	Var.	$\hat{\sigma}_2 = 4.125 - 1.40X_3 + 1.58X_5$
3	Med.	$\hat{Y}_3 = 201.19 - 4.89X_1 - 7.78X_1^2 - 3.89X_2 - 9.8X_5$ (0.79)
	Var.	$\hat{\sigma}_3 = 6.225 + 2.525X_1$
4	Med.	$\hat{Y}_4 = 31.57 + 3.60X_1 + 1.43X_1^2 + 1.98X_2 + 1.58X_2^2 + 1.69X_3 + 1.10X_4 + 2.36X_5$ (0.93)
	Var.	$\hat{\sigma}_4 = 0.623 + 0.253X_2$
5	Med.	$\hat{Y}_5 = 61.73 + 2.06X_1 + 2.46X_1^2 + 2.33X_2 + 0.938X_3 + 0.938X_5$ (0.82)
	Var.	$\hat{\sigma}_5 = 1.633 + 0.892X_1$

Resultados parciais

Recapitulando o procedimento para modelagem da variância

Trt	Bloco	X1	X2	X3	X4	X5	Resp.1 Obs.	Resp. 1 Pred.	Diferença
1	1	-1	-1	-1	1	-1			
2	1	-1	-1	-1	-1	1			
3	1	-1	1	1	1	-1			
4	1	-1	1	1	-1	1			
10	2	-1	1	-1	-1	-1			
11	2	-1	1	-1	1	1			
12	2	-1	-1	1	-1	-1			
13	2	-1	-1	1	1	1			
19	3	-1	0	0	0	0			
5	1	1	1	-1	1	-1			
6	1	1	1	-1	-1	1			
7	1	1	-1	1	1	-1			
8	1	1	-1	1	-1	1			
14	2	1	-1	-1	-1	-1			
15	2	1	-1	-1	1	1			
16	2	1	1	1	-1	-1			
17	2	1	1	1	1	1			
20	3	1	0	0	0	0			

Utilizam-se dados experimentais e modelos para média

Verificando níveis extremos de X_1

Resultados são organizados numa tabela

Fatores de controle				
X_1	X_2	…	X_M	*Variâncias amostrais*
-1	0	…	0	$s^2_{1,(-1)}$
+1	0	…	0	$s^2_{1,(+1)}$
0	-1	…	0	$s^2_{2,(-1)}$
0	+1	…	0	$s^2_{2,(+1)}$
⋮	⋮	⋱	⋮	⋮
0	0	…	-1	$s^2_{M,(-1)}$
0	0	…	+1	$s^2_{M,(+1)}$

***Variância amostral das diferenças de X_1 no nível* -1**

Teste a significância da variância amostral para cada fator controlável

$$F = \frac{s^2_{1,(+1)}}{s^2_{1,(-1)}}$$

Mantenha na tabela somente fatores cujas variâncias resultem significativas!

Testando para níveis extremos de X_1

$$F_{0,025;\,9,9} = 4,03$$

Variâncias obtidas a partir de 9 valores em cada nível

Teste o procedimento nos dados deste experimento fatorial

Trt	*X1*	*X2*	*X3*	*Y1*	*Y2*
1	-1	-1	-1	40,02	37,44
2	1	-1	-1	30,16	36,92
3	-1	1	-1	35,68	42,67
4	1	1	-1	34,35	43,60
5	-1	-1	1	41,03	37,60
6	1	-1	1	28,78	35,43
7	-1	1	1	35,28	45,44
8	1	1	1	33,54	44,41