4 - Teste de Friedman - Várias amostras relacionadas

Extensão do teste de amostra pareada para mais do que dois tratamentos

Problema: Detectar diferenças entre $K \geq 2$ tratamentos

Bloco: grupo consistindo de K unidades experimentais similares.
As K unidades experimentais dentro de cada bloco devem ser associadas aleatoriamente aos K tratamentos. (Assim, cada tratamento é aplicado uma única vez dentro de cada bloco).

b - número de blocos

Dentro de cada bloco associa-se postos de 1 a K aos resultados dos K tratamentos.

Este experimento é denominado "Planejamento em Blocos Completamente Casualizados".

Dados:

b v.a. K dimensionais independentes,

$(X_{i1}, X_{i2} \dots X_{iK})$, $i = 1, 2 \dots b$, onde

X_{ij} - medida do j-ésimo tratamento no i-ésimo bloco

Tratamento

Bloco		1	2	3	...	K
	1	X_{11}	X_{12}	X_{13}		X_{1K}
	2	X_{21}	X_{22}	X_{23}		X_{2K}
	3					
	⋮					
	b	X_{b1}	X_{b2}	X_{b3}		X_{bK}

Suposições

1- Os blocos são independentes entre si.

2- Dentro de cada bloco, as observações podem ser ordenadas segundo algum critério. (Um número moderado de empates é tolerável).

Feita essa ordenação sejam

$R(X_{ij})$ = posto de X_{ij}

$$R_j = \sum_{i=1}^{b} R(X_{ij}) \qquad j = 1, 2, \dots, k$$

↳ soma dos postos associados ao j-ésimo tratamento.

H_0: Não existem diferenças entre os tratamentos (Todas as distribuições de postos são igualmente prováveis)

H_a: Pelo menos um dos tratamentos tende a produzir observações maiores.

Estatística de Teste:

soma média de postos

Friedman:

$$T = \frac{12}{bk(k+1)} \sum_{j=1}^{k} \left[R_j - \frac{b(k+1)}{2} \right]^2$$

$b \dfrac{(1+k)k}{2}$ = soma de todos os postos

Na presença de empates utilizar

$$T_1 = \frac{(k-1)\sum_{j=1}^{k}\left[R_j - \frac{b(k+1)}{2}\right]^2}{A_1 - C_1} = \frac{(k-1)\left[\sum_{j=1}^{k} R_j^2 - bC_1\right]}{A_1 - C_1}$$

$$A_1 = \sum_{i=1}^{b}\sum_{j=1}^{k}\left[R(X_{ij})\right]^2 \qquad C_1 = \frac{bk(k+1)^2}{4}$$

Pesquisas mais recentes sugerem utilizar

$$T_2 = \frac{(b-1)T_1}{b(k-1) - T_1}$$

que é a estatística da tabela de análise de variâncias com dois fatores calculada para os postos $R(X_{ij})$.

A distribuição exata de T, T_1 e T_2 é difícil de ser obtida e por isso utiliza-se aproximações. A forma tradicional do teste de Friedman utiliza a estatística T que sob H_0 tem dist. aproximadamente χ^2 com $k-1$ g l.

Conover (2ª e 3ª edições) sugere adotar a estatística T_2, que sob H_0 tem dist. aproximadamente F com $k-1$ gl no numerador e $(b-1)(k-1)$ gl no denominador.

Hipóteses:

H_0: Todas as distribuições de postos são igualmente prováveis ($\Leftrightarrow$ os tratamentos têm efeitos idênticos)

H_a: Pelo menos um dos tratamentos tende a produzir observações maiores.

Rejeita-se H_0 ao nível de significância α se $T_2 > w_{1-\alpha}$, em que $w_{1-\alpha}$ é o quantil de ordem $1-\alpha$ da distribuição F com $k-1$ gl no numerador e $(b-1)(k-1)$ gl no denominador.

A aproximação melhora com o aumento de b.

Obs:

- Se não existirem empates,

$$A_1 = \frac{b\,k(k+1)(2k+1)}{6}$$

e $$A_1 - C_1 = \frac{b\,k(k+1)(k-1)}{12}$$

- Se H_0 é rejeitada, podem ser construídos contrastes. Os tratamentos i e j são considerados distintos se

$$|R_j - R_i| > t_{1-\alpha/2}\left[\frac{2\left(bA_1 - \sum R_j^2\right)}{(b-1)(k-1)}\right]^{1/2},$$

sendo $t_{1-\alpha/2}$ o quantil de ordem $1-\alpha/2$ da distribuição t-student com $(b-1)(k-1)$ g.l.

- Na 3ª edição do livro Conover, pag 373, encontra-se um teste alternativo de igualdade de médias de tratamentos para o planejamento em blocos completamente casualizados.

Distribuição exata das estatísticas do teste sob H_0:

Sob H_0, cada distribuição dos k postos dentro de um bloco tem probabilidade $\frac{1}{k!}$. Pela independência dos blocos, para a tabela toda, a probabilidade de uma particular distribuição de postos é

$$\frac{1}{(k!)^b}.$$

Portanto $P(T=x) = \dfrac{\#(b,x)}{(k!)^b}$,

$\#(b,x) =$ nº de arranjos com $T=x$.

Exemplo

Solicitou-se a 12 pessoas que selecionassem quatro áreas de terra idênticas e plantassem quatro tipos diferentes de grama. Após um certo período de tempo, cada pessoa deve ordenar os tipos de grama em ordem de preferência, 1 pior, 4 melhor. Testar a hipótese de igualdade de preferências contra a alternativa de que um tipo de grama tende a ser preferido.

	Tratamento (tipo de grama)			
Blocos	A	B	C	D
1	4	3	2	1
2	4	2	3	1
3	3	1,5	1,5	4
4	3	1	2	4
5	4	2	1	3
6	2	2	2	4
7	1	3	2	4
8	2	4	1	3
9	3,5	1	2	3,5
10	4	1	3	2
11	4	2	3	1
12	3,5	1	2	3,5
R_i	38	23,5	24,5	34

$$A_1 = \sum_{i=1}^{12} \sum_{j=1}^{4} \left[R(X_{ij}) \right]^2 = 356{,}5$$

empates

$$C_1 = \frac{b\,K\,(K+1)^2}{4} = \frac{12.4.25}{4} = 300$$

$$T_1 = \frac{3\left[38^2 + 23{,}5^2 + 24{,}5^2 + 34^2 - 12.300 \right]}{356{,}5 - 300} = 8{,}097$$

$$T_2 = \frac{11.\,8{,}097}{12.3 - 8{,}097} = 3{,}19$$

Para $\alpha = 0{,}05$, o quantil de ordem $0{,}95$ da distribuição F com 3 g l no numerador e 33 g l no denominador é $F_c = 2{,}9$.

Rejeita-se H_0, os dados sugerem que existe uma tendência de pelo menos um tipo de grama ser preferido com relação aos demais.

p-valor (nível descritivo) $= P(F_{3,33} \geq 3{,}19) \simeq 0{,}04$.

Rejeitaria-se H_0 para $\alpha \geq 0{,}04$.

Comparações Múltiplas

$$t_{0,975}\left[\frac{2(bA_1-\sum R_j^2)}{(b-1)(k-1)}\right]^{1/2} = 11,49$$

Conclui-se pela desigualdade de médias para

$$|R_i - R_j| > 11,49$$

⇒ Conclui-se que a grama 1 pode ser considerada superior às gramas 2 e 3.

Nenhuma outra diferença se mostrou significante.

Ver Teste de Quade - Conover, pag 373