

LGN5830 - Biometria de Marcadores Genéticos

Responsáveis: Antonio Augusto Franco Garcia e Roland Vencovsky
Departamento de Genética, ESALQ/USP
Segundo semestre de 2007

Populações de Mapeamento, Teste da Segregação Mendeli- ana, Análise de Ligação.

1 INTRODUÇÃO AO R

- 1.1 Leia o texto “The Friendly Beginners’ R Course”, que está disponível na página do curso. Visite o site, instale o programa, etc. Leia também o tutorial <http://www.leg.ufpr.br/Rtutorial/>.

2 TESTE DA RAZÃO DE VEROSSIMILHANÇA

- 2.1 Mostre qual é a relação numérica entre LRT e LOD . É possível associar p -valores ao LOD ? Justifique.
- 2.2 Em seu clássico experimento, Mendel (1866) cruzou linhagens de ervilhas diferentes quanto à forma (e outras características), tendo obtido os resultados apresentados na tabela abaixo (população F_2). Considere cada uma das plantas separadamente, e também o total de sementes.

1. Obter estimativas de máxima verossimilhança do parâmetro θ (probabilidade de uma dada semente ser lisa).
2. Testar usando LRT se esses valores diferem do esperado sob H_0 .
3. Caso o gene que controla esse caráter não mostrasse dominância, qual distribuição de proba-

bilidades deveria ser usada para analisar esses dados? Justifique.

4. Mendel realizou esse mesmo tipo de experimento para várias outras características (7 ao todo) e plantas. Seriam esperados falsos positivos nos seus resultados? Justifique.

- Nota: observe que essa forma de realizar o teste é diferente da usual (teste de aderência), embora também faça uso da distribuição χ^2 .
- Use o R, tanto quanto possível, para fazer o exercício. Tome como base os códigos usados na aula (ver página do curso).

Planta	Sementes lisas	Sementes rugosas
1	45	12
2	27	8
3	24	7
4	19	10
5	32	11
6	26	6
7	88	24
8	22	10
9	28	6
10	25	7

- 2.3 Considere o arquivo de dados MAIZE.TXT, que refere-se a uma população F_2 analisada com um marcador codominante (dados na página do curso).

1. Complete a Tabela 1, realizando assim testes para a segregação mendeliana de cada marcador.
2. Determine o nível de significância adequado para as estatísticas LRT e LOD , usando a correção de Bonferroni.
3. Interpretar os resultados do ponto de vista genético.

- Obs: use o programa computacional R, adaptando os códigos apresentados na aula (ver página do curso).

	n_1	n_2	n_3	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$L(\hat{\theta})$	$L(\theta)$	LRT	p	LOD
M_1											
M_2											
M_3											
M_4											
M_5											
M_6											
M_7											
M_8											
M_9											
M_{10}											
M_{11}											
M_{12}											

3 ANÁLISE DE LIGAÇÃO

3.1 Para resolver os exercícios abaixo, use os dados do retrocruzamento apresentados na aula (MOUSE.TXT; ver na página do curso) e o software R, tanto quanto possível.

1. Complete a tabela abaixo, realizando assim testes da ligação genética (os termos seguem a notação usada em aula).
 2. Quantos grupos de ligação foram formados? Interprete os resultados.
- Nota: em situações reais, o correto seria testar a ligação entre todas as combinações dois-a-dois dos marcadores. Nesse exercício, isso não foi solicitado apenas para diminuir o número de cálculos necessários.

Marcas	$\hat{r}$	$L(\hat{r})$	$L(r)$	LRT	p	LOD
$M_1 - M_2$						
$M_2 - M_3$						
$M_3 - M_4$						
$M_4 - M_5$						
$M_5 - M_6$						
$M_6 - M_7$						
$M_7 - M_8$						
$M_8 - M_9$						
$M_9 - M_{10}$						
$M_{10} - M_{11}$						
$M_{11} - M_{12}$						
$M_{12} - M_{13}$						
$M_{13} - M_{14}$						

4 BIBLIOGRAFIA

1. Liu, B.H. Statistical Genomics - Linkage, Mapping and QTL Analysis. CRC Press, New York, 1998. Capítulos 5 e 6.
Leitura sobre o assunto como um todo, dando grande enfoque para os procedimentos estatísticos
2. Edwards, A.W.F. Likelihood (expanded edition). The John Hopkins University Press, 1992.
Texto clássico sobre a verossimilhança e seu papel na inferência estatística.
3. Benjamini, Y.; Hochberg, Y. 1995. Controlling the false discovery rate at a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57: 289-300.
4. Benjamini, Y.; Hochberg, Y. 2000. On the adaptive control of the false discovery rate in multiple testing with independent statistics. J. Educ. Behav. Stat. 25: 60-83.
5. Benjamini, Y.; Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29: 1165-1187.
6. Storey, J.D. 2003. The positive false discovery rate: a Bayesian interpretation and the q -value. Ann. Statist. 31: 2013-2035.

As referências 3-6 apresentam o conceito de FDR em detalhes.
7. Oraguzie, N.C.; Rikkerink, E.H.A.; Gardiner, S.E.; de Silva, H.N. (eds.) Association Mapping in Plants. Springer 2007
No capítulo 7, os autores apresentam uma excelente discussão sobre o problema dos múltiplos testes e o conceito de p -valor.