

1. Os dados a seguir correspondem a médias de uma variável quantitativa de acordo com o genótipo conjunto de dois locos de marcadores moleculares (do tipo *Single Nucleotide Polimorphisms*, SNP_A e SNP_B)), para 7 populações sob estudo. As ações gênicas nessas populações são denominadas como: P1: Ação gênica intermediária; P2: Dominância completa; P3: Ação gênica complementar; P4: Epistasia complexa; P5: Dominância parcial; P6: Superdominância; P7: Ação gênica duplicada.

P1				P2				P3				P4			
	bb	Bb	BB		bb	Bb	BB		bb	Bb	BB		bb	Bb	BB
aa	20	18	16		20	20	16		20	20	10		38.7	6.6	3.2
Aa	17	15	13		20	20	16		20	20	10		4.0	2.0	1.3
AA	14	12	10		14	14	10		10	10	10		2.7	3.4	1.6

P5				P6				P7			
	bb	Bb	BB		bb	Bb	BB		bb	Bb	BB
aa	20	19	16		20	21	16		20	20	20
Aa	19	18	15		22	23	18		20	20	20
AA	14	13	10		14	15	10		20	20	10

- Para cada população (P1 a P7) construa o gráfico de perfis de médias correspondente. Com base nesse gráfico, há indicação de efeito de interação entre os dois fatores genéticos (SNP_A e SNP_B)? Comente.
- Vamos agora gerar dados de Delineamentos Completamente Aleatorizados (DCA) com esquema Fatorial 3x3 (Fator SNP_A e fator SNP_B, ambos, em 3 níveis genotípicos). Considere $\sigma^2 = 1$ e, com base nas 9 médias apresentadas em P1 a P7, gere $r=25$ valores para cada um dos 9 grupos, sob o modelo de distribuição Normal. Em cada caso, obtenha a tabela de ANOVA e interprete os resultados.

2. Considere os dados do arquivo "map.txt" que traz informações de 13 variáveis fenotípicas ("SEX", "Peso", "Altura", "CIRCABD", "GLICOSE", "CTOTAL", "TRIG", "Idade", "mediaPAS", "mediaPAD", "MedPres", "MedCol", "MedDb"), além do genótipo de 102 SNPs ("rs12628452_A", "rs7289830_T", etc.) para uma amostra de 1675 indivíduos.

- Seguindo o princípio da Aleatorização Mendeliana (Burgess and Thompson, 2015), suponha que o nível genotípico ($aa=0$, $Aa=1$, $AA=2$) de cada SNP foi aleatoriamente atribuído a cada indivíduo. Assim, escolha uma das variáveis quantitativas do banco de dados, e avalie o efeito de cada SNP sobre essa variável por meio de um modelo de ANOVA. Adote um esquema DCA com 1 fator em 3 níveis. Caso seja necessário, proponha uma transformação da variável para atender às premissas clássicas.
- Apresente os valores-p da estatística F do teste global da ANOVA em um "gráfico Manhattan" (gráfico de dispersão: $-\log_{10}(\text{valor-p})$ x índice do SNP). Interprete os resultados. Discuta sobre possíveis correções para os múltiplos (102) testes realizados simultaneamente.
- Para cada SNP, teste seu efeito aditivo, $H_0: \mu_{AA} - \mu_{aa} = 0$. Apresente os resultados desses testes por meio de um "gráfico vulcão" (gráfico de dispersão: $-\log_{10}(\text{valor-p})$ x estimativa do contraste ($\mu_{AA} - \mu_{aa}$)). Interprete os resultados.
- Agora, para cada SNP, teste seu efeito de dominância, $H_0: \mu_{Aa} - (\mu_{AA} + \mu_{aa})/2 = 0$. Apresente os resultados desses testes por meio de um "gráfico vulcão" (gráfico de dispersão: $-\log_{10}(\text{valor-p})$ x estimativa do contraste $\mu_{Aa} - (\mu_{AA} + \mu_{aa})/2$). Interprete os resultados.
- Hipoteticamente, assumo que há interesse, e é apropriado, realizar uma meta-análise dos resultados de significância do efeito aditivo desses 102 marcadores. Use as fórmulas apresentadas a seguir para combinar os resultados e obter um valor-p global.

Table 1. Formulae for meta-analysis

	Analytical strategy	
	Sample size based	Inverse variance based
Inputs	N_i - sample size for study i P_i - P -value for study i Δ_i - direction of effect for study i	β_i - effect size estimate for study i se_i - standard error for study i
Intermediate Statistics	$Z_i = \Phi^{-1}(P_i/2) * \text{sign}(\Delta_i)$ $w_i = \sqrt{N_i}$	$w_i = 1/SE_i^2$ $se = \sqrt{1/\sum_i w_i}$ $\beta = \sum_i \beta_i w_i / \sum_i w_i$
Overall Z-Score	$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$	$Z = \beta/SE$
Overall P -value	$P = 2\Phi(- Z)$	

Fonte: Cristen, et al. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26(17): 2190–2191, doi:10.1093/bioinformatics/btq340.

- f) Ainda, considerando o banco de dados “map.txt”, escolha um dentre os 102 SNPs e avalie o possível efeito de interação do SNP e do sexo do indivíduo sobre a variável que está sendo analisada. Para tanto, ajuste um modelo ANOVA para um esquema DCA Fatorial 3x2.

3. Os dados a seguir se referem à contagem diferencial de um tipo de pulgão que ocorre em plantas, antes e depois de um tratamento por 10 dias. Três tratamentos, com água (T1), esporos (T2) e com óleo natural (T3) foram usados em 5 plantas para avaliar o controle da infestação pelo pulgão. No experimento, três folhas contaminadas foram selecionadas de cada uma das plantas, e os tratamentos foram então aleatoriamente atribuídos às folhas. A contagem de pulgões em quadrantes definidos nas folhas foi feita antes e depois do tratamento e a diferença registrada. Os dados estão indicados a seguir.

Trat	Planta 1	Planta 2	Planta 3	Planta 4	Planta 5
T1: Água	-7	11	10	?	4
T2: Esporos	2	22	2	2	5
T3: Óleo	7	32	15	16	10

- Complete a casela faltante (indicada como ?) e realize uma análise de variância destes dados. Que suposições foram adotadas na análise? Essas estão satisfeitas?
- Compare as fontes de variação presentes no experimento.
- Suspeitando que a hipótese de normalidade não é apropriada a estes dados, construa um Teste de Aleatorização para o efeito global dos tratamentos.

Boa sorte ☺