

1. Um estudo observacional caso-controle foi conduzido para investigar a associação entre a ocorrência de uma mutação e uma doença. Os dados coletados de duas populações são apresentados a seguir.

	População 1		População 2	
	Mutação	Normal	Mutação	Normal
Casos	10	90	40	20
Controles	90	810	20	10

a) Calcule (à mão) a estatística razão de chances (OR: *odds ratio*) para cada população. Adote um modelo de regressão logística para analisar os dados de cada população. Obtenha o intervalo de confiança para o OR. Interprete.

b) Some as correspondentes caselas de dados das duas populações e calcule (à mão) a estatística OR. Como em a), adote um modelo de regressão logística para analisar os dados. Obtenha o intervalo de confiança para o OR. Interprete.

c) Com base nos resultados obtidos em a) e b), use o Paradoxo de Simpson para explicar um possível efeito de confundimento que pode ocorrer em estudos de associação considerando dados estratificados desse tipo.

d) Para realizar uma meta-análise destes dados, adote um modelo de regressão logística considerando os estratos populacionais como um fator sob estudo.

2. A tabela a seguir apresenta resultados de um estudo de associação de um SNP de interesse com uma doença. Para validação dos resultados do estudo também foram genotipados 20 marcadores considerados neutrais (isto é, que não estão associados com a doença e nem com o SNP sob estudo). As estatísticas Qui-quadrado (com 1 g.l.) da associação destes marcadores com a doença estão também apresentadas.

	CC	CG	GG
Casos	25	20	—
Controles	5	20	45

Estatísticas Qui-quadrado				
0.1101	0.1485	1.3233	0.0191	0.1935
0.0158	1.0742	0.0642	0.0039	0.0101
0.7083	0.0268	2.4173	2.2925	6.4653
0.2750	10.8274	4.7093	0.3573	0.5707

a) Preencha, a seu critério, a casela (casos, GG) com um valor de frequência. Realize um teste de associação do SNP com a doença com base na razão de chances (OR). Considere os seguintes casos: 2 graus de liberdade e 1 grau de liberdade (efeito linear do SNP). Interprete os resultados.

b) Use a informação dos marcadores neutrais para validação dos resultados do estudo. Para isso, considere os seguintes procedimentos:

- Pritchard e Rosenberg (1999) propuseram o uso de marcadores neutrais para construir um teste da inexistência de confundimento devido à estrutura de população (hipótese H₀). Considerando os testes de associação de cada um de L marcadores neutrais obtém-se a estatística

$$\chi_P^2 = \chi_{M1}^2 + \chi_{M2}^2 + \dots + \chi_{ML}^2$$

cuja significância é verificada com base na distribuição Qui-quadrado com L graus de liberdade. Se não houver indicação de estrutura de população por este teste, a associação com o gene candidato deve ser considerada válida.

- Devlin e Roeder (1999) propuseram corrigir a estatística Qui-quadrado do gene candidato (digamos, χ_c^2) dividindo por um fator de inflação, tal que:

$$\chi_{GC}^2 = \chi_c^2 / \max(\hat{\lambda}, 1); \quad \hat{\lambda} = \frac{\text{mediana}(\chi_{M1}^2, \dots, \chi_{ML}^2)}{0.4549}$$

O valor 0.4549 corresponde ao quantil 50% da distribuição Qui-quadrado com 1 g.l. Os L marcadores são denominados neutrais ou de controle genômico (GC). A associação do gene candidato é então testada avaliando a significância da estatística corrigida, χ_{GC}^2 , na distribuição Qui-quadrado com 1 g.l.

3. Considere os dados do GAW16 (Genetic Analysis Workshop 16) disponíveis no edisciplinas. O arquivo "phen.txt" contém informações fenotípicas de 2062 indivíduos classificados de acordo com o diagnóstico de artrite reumatóide, sendo 868 casos (Affection=1) e 1194 controles (Affection=0). Para esses mesmos indivíduos, os arquivos "gene6.6.txt, gene6.7.txt, gene6.8.txt, gene6.9.txt e gene6.10.txt" contém genótipos de SNPs no cromossomo 6, em que 0, 1 e 2 correspondem ao número de cópias de um alelo alvo.

a) Realize uma análise de associação de cada SNP com a doença (casos e controles). Adote um modelo de regressão logística. Apresente os resultados em um gráfico Manhattan e em um gráfico Vulcão. Há SNPs significantes?

b) Considere todos os 5 arquivos com dados de genótipos de SNPs. Padronize esses dados, subtraindo a média e dividindo pelo desvio padrão, e obtenha os dois primeiros Componentes Principais (CP). Usando esses CP, construa um gráfico de dispersão dos 2062 indivíduos. Identifique os grupos caso e controle. Os CP discriminam os indivíduos? Comente.

c) Ajuste modelos de regressão logística usando os dois primeiros CP como covariáveis além do sexo dos indivíduos. Construa os gráficos Manhattan e Vulcão. Há SNPs significantes?

d) Compare os resultados dos modelos de regressão logística ajustados em a) e c).

Boa sorte 😊