

Análise nos genótipos do cromossomo 7 dos dados sobre densidade óssea

Adriana Eva Fernandes

Universidade Federal de São Carlos
Docente: Daiane Aparecida Zuanetti

14 de novembro de 2023

Descrição do conjunto de dados

O conjunto de dados corresponde a um conjunto de informações sobre 651 ratos do sexo feminino F2, obtidos por meio do cruzamento dos pais F1 NZB/B1NJ e RF/J. O objetivo desse cruzamento foi identificar os locais genéticos que controlavam as características do fêmur, como suas propriedades geométricas, minerais ósseos e densidade (BMD).

Neste estudo, realizamos uma análise da BMD (densidade óssea total volumétrica em miligramas por centímetro cúbico) em uma escala logarítmica em 7 marcadores genéticos do cromossomo 7 observados em cada rato, que eram do tipo microssatélites.

Conjunto de dados

Tabela: Dados genéticos do cromossomo 7

| Indivíduo | MK- 1-1 | MK- 1-2 | MK- 1-3 | MK- 1-4 | MK- 1-5 | MK- 1-6 | MK- 1-7 | BMD (log) |
|-----------|------------|------------|------------|------------|------------|------------|------------|--------------|
| Ind-1 | -1 | -1 | 0 | 1 | 1 | 1 | 1 | 6.630683 |
| Ind-2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 6.651572 |
| Ind-3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 6.570883 |
| Ind-4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 6.637258 |
| Ind-5 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 6.614726 |
| Ind-6 | -1 | -1 | -1 | 0 | 0 | 0 | 0 | 6.733402 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |

Descrição do conjunto de dados

No cromossomo 7 de cada indivíduo, foram observados os respectivos marcadores genéticos, denotados por:

- MK-1-1: marcador que representa o locus 0 do cromossomo 7;
- MK-1-2: marcador que representa o locus 23.5 do cromossomo 7.
- MK-1-3: marcador que representa o locus 26.8 do cromossomo 7;
- MK-1-4: marcador que representa o locus 36.0 do cromossomo 7;
- MK-1-5: marcador que representa o locus 51.8 do cromossomo 7;
- MK-1-6: marcador que representa o locus 59 do cromossomo 7;
- MK-1-7: marcador que representa o locus 71 do cromossomo 7;

Descrição do conjunto de dados

No marcador de cada indivíduo, foram observados os respectivos genótipos, denotados por:

- 1: genótipo que representa os ratos homozigotos dominantes, o que significa que ambos os alelos presentes no locus genético são do tipo dominantes;
- 0: genótipo que representa os ratos heterozigotos, o que significa que eles possuem um alelo dominante e um alelo recessivo no locus genético;
- -1: genótipo que representa os ratos homozigotos recessivos, o que significa que ambos os alelos presentes no locus genético são do tipo a;

e a densidade óssea, denotada por:

- BMD: densidade óssea total volumétrica em miligramas por centímetro cúbico) em uma escala logarítmica.

Objetivo das análises

Os objetivos dessas análises são:

- Comparar as frequências genotípicas em cada um dos marcadores do cromossomo 7;
- Comparar as frequências alélicas em cada um dos marcadores do cromossomo 7;
- Verificar a existência de possíveis relações entre as frequências alélicas de marcadores diferentes do cromossomo 7;
- Verificar qual é o alelo mais raro do cromossomo 7;
- Verificar qual marcador genético do cromossomo 7 parece apresentar mais desbalanceamento entre os alelos;
- Verificar se os marcadores do cromossomo 7 estão em EHW;
- Verificar se os marcadores do cromossomo 7 estão, aos pares, em equilíbrio ou desequilíbrio de associação;
- Verificar quais marcadores do cromossomo 7 estão mais associados à densidade óssea (BMD) do fêmur na escala logarítmica;

Análise Descritiva

Inicialmente, foi feita uma análise descritiva que nos permitiu identificar as possíveis conclusões que poderiam ser obtidas a respeito dos objetivos com base na análise dos dados da amostra observada.

Análise Descritiva - Genótipos

genotipos.jpeg

Análise Descritiva - Genótipos

Tabela: Contagem do genótipo para sete marcadores genéticos de $n = 651$ indivíduos.

| Genótipo | MK- 1-1 | MK- 1-2 | MK- 1-3 | MK- 1-4 | MK- 1-5 | MK- 1-6 | MK- 1-7 |
|----------|------------|------------|------------|------------|------------|------------|------------|
| 1 | 165 | 152 | 158 | 171 | 174 | 186 | 151 |
| 0 | 319 | 335 | 299 | 308 | 317 | 301 | 348 |
| -1 | 167 | 164 | 194 | 172 | 160 | 164 | 152 |

Tabela: Medidas de tendência central dos genótipos dos sete marcadores genéticos.

| | Genótipos | | |
|---------|-----------|-----|-----|
| | 1 | 0 | -1 |
| média | 165 | 319 | 167 |
| mediana | 165 | 319 | 167 |

Genótipos - Conclusões

Com base nos gráficos das frequências genotípicas dos marcadores do cromossomo 7, na tabela da contagem dos genótipos dos 7 marcadores genéticos e na tabela de medidas de tendência central, temos indicativos de que:

- Genótipo Aa (0) é predominante em todas as regiões estudadas, ou seja, há predominância de indivíduos heterozigotos na amostra.
- Os genótipos aa (-1) e AA (1) têm frequências próximas na maioria dos marcadores.
- No entanto, no marcador 3, o genótipo aa tem uma frequência maior (29,8%) do que o genótipo AA (24,3%).
- A predominância do genótipo Aa em todas as regiões estudadas pode indicar que os genótipos estão em equilíbrio de Hardy-Weinberg.
- As medidas de tendência central nos dão indícios de que a distribuição dos genótipos é simétrica, uma vez que a média e mediana de cada genótipo, são iguais.

Análise Descritiva - Alelos

alelos.jpeg

Análise Descritiva - Alelos

Tabela: Contagem de alelos para sete marcadores genéticos de $n = 651$ indivíduos.

| Alelo | MK- 1-1 | MK- 1-2 | MK- 1-3 | MK- 1-4 | MK- 1-5 | MK- 1-6 | MK- 1-7 |
|-----------|------------|------------|------------|------------|------------|------------|------------|
| Dominante | 649 | 639 | 615 | 650 | 665 | 673 | 650 |
| Recessivo | 653 | 663 | 687 | 652 | 637 | 629 | 652 |

Tabela: Medidas de tendência central dos alelos dos sete marcadores genéticos.

| | Alelos | |
|---------|-----------|-----------|
| | Recessivo | Dominante |
| média | 653 | 649 |
| mediana | 653 | 649 |

Alelos - Conclusões

Com base nos gráficos das frequências alélicas dos marcadores do cromossomo 7, na tabela da contagem dos alelos dos 7 marcadores genéticos e na tabela de medidas de tendência central, temos indicativos de que:

- Há semelhança nas frequências alélicas entre os marcadores para a maioria dos locus, exceto pelo terceiro marcador, que tem uma frequência um pouco maior no alelo recessivo;
- As medidas de tendência central nos dão indícios de que a distribuição dos alelos é simétrica, uma vez que a média e mediana de cada alelo são iguais;

Análise Descritiva - BMD

bmd freq.png

Tabela: Medidas de resumo

| Min. | 1 Quartil. | Median | Mean | 3 Quartil. | Max. |
|-------|------------|--------|-------|------------|-------|
| 6.486 | 6.667 | 6.692 | 6.690 | 6.720 | 6.799 |

BMD - Conclusões

Com base na análise do gráfico da frequência de indivíduos por densidade óssea do cromossomo 7 e na tabela das medidas de resumo, temos que:

- A densidade óssea observada varia de 6.486 a 6.799;
- Há uma maior concentração de indivíduos em torno da média;
- Há um indivíduo que apresenta densidade óssea atípica;
- Com exceção dos valores extremos que distorcem simetria, a distribuição da densidade óssea parece ser simétrica;

Análise Descritiva - Gametas

Tabela: Tabela de contingência para a contagem dos gametas aos pares.

| | MK-1-2 | | MK-1-3 | | MK-1-4 | | MK-1-5 | | MK-1-6 | | MK-1-7 | |
|-----------|--------|-----|--------|-----|--------|-----|--------|-----|--------|-----|--------|-----|
| MK-1-1 | D | R | D | R | D | R | D | R | D | R | D | R |
| Dominante | 495 | 154 | 456 | 193 | 446 | 203 | 422 | 227 | 414 | 235 | 412 | 237 |
| Recessivo | 144 | 509 | 159 | 494 | 204 | 449 | 243 | 410 | 259 | 394 | 238 | 415 |

Tabela: Tabela de contingência para a contagem dos gametas aos pares.

| | MK-1-3 | | MK-1-4 | | MK-1-5 | | MK-1-6 | | MK-1-7 | |
|-----------|--------|-----|--------|-----|--------|-----|--------|-----|--------|-----|
| MK-1-2 | D | R | D | R | D | R | D | R | D | R |
| Dominante | 573 | 66 | 560 | 79 | 508 | 131 | 480 | 159 | 420 | 219 |
| Recessivo | 42 | 621 | 90 | 573 | 57 | 506 | 193 | 470 | 230 | 433 |

Análise Descritiva - Gametas

Tabela: Tabela de contingência para a contagem dos gametas aos pares.

| | MK-1-4 | | MK-1-5 | | MK-1-6 | | MK-1-7 | |
|-----------|--------|-----|--------|-----|--------|-----|--------|-----|
| MK-1-3 | D | R | D | R | D | R | D | R |
| Dominante | 549 | 66 | 497 | 118 | 467 | 148 | 403 | 212 |
| Recessivo | 101 | 586 | 168 | 519 | 206 | 481 | 247 | 440 |

Tabela: Tabela de contingência para a contagem dos gametas aos pares.

| | MK-1-5 | | MK-1-6 | | MK-1-7 | |
|-----------|--------|-----|--------|-----|--------|-----|
| MK-1-4 | D | R | D | R | D | R |
| Dominante | 568 | 82 | 529 | 121 | 444 | 206 |
| Recessivo | 97 | 555 | 144 | 508 | 206 | 446 |

Análise Descritiva - Gametas

Tabela: Tabela de contingência para a contagem dos gametas aos pares.

| | MK-1-6 | | MK-1-7 | |
|-----------|--------|-----|--------|-----|
| MK-1-5 | D | R | D | R |
| Dominante | 612 | 53 | 502 | 163 |
| Recessivo | 61 | 576 | 148 | 489 |

Tabela: Tabela de contingência para a contagem dos gametas aos pares.

| | MK-1-7 | |
|-----------|--------|-----|
| MK-1-6 | D | R |
| Dominante | 525 | 148 |
| Recessivo | 61 | 576 |

Gametas - Conclusões

Com base na análise da tabela de gametas, é possível identificar possíveis indícios de desequilíbrio de ligação entre os marcadores. Essa conclusão é derivada da observação de padrões inconsistentes e desproporcionais nos números observados de cada combinação de alelos.

Coeficiente de desequilíbrio

O coeficiente de desequilíbrio pode ser estimado quando verificamos que o EHW não acontece. Esta é simplesmente a diferença entre uma frequência e seu valor esperado quando não há associações entre os alelos. Considerando para o nosso problema, temos o mais simples de dois alelos A e a, $D_{AA} = D_{aV} = D_{aa}$. Logo, simplificamos a notação de D_{AA} por D_A e temos um único coeficiente de desequilíbrio.

Coeficiente de desequilíbrio

Tabela: Coeficiente de desequilíbrio dos marcadores do cromossomo 7

| Marcador | D_A |
|----------|--------|
| MK-1-1 | 0 |
| MK-1-2 | -0.01 |
| MK-1-3 | 0.0019 |
| MK-1-4 | 0.01 |
| MK-1-5 | 0.0099 |
| MK-1-6 | 0.0196 |
| MK-1-7 | -0.02 |

- O coeficiente de desequilíbrio para o marcador 1 é igual a 0, o que indica que a presença de indivíduos heterozigotos e homozigotos nesse marcador é a mesma;
- O coeficiente de desequilíbrio para os marcadores 3,4,5,6 apresentaram valores positivos, o que indica maior presença de indivíduos homozigotos nesses marcadores;
- O coeficiente de desequilíbrio para os marcadores 2,7 apresentaram valores negativos, o que indica maior presença de indivíduos heterozigotos nesses marcadores;

Lei de Hardy-Weinberg

Após a estimativa da frequência alélica e genotípica, uma das primeiras análises de interesse dos dados genéticos populacionais consiste em investigar associações entre dois alelos em um determinado locus. Quando não há forças perturbadoras e há o acasalamento aleatório em grandes populações, os pares de alelos são considerados como não associados ou aleatórios.

Uma consequência importante desta lei (se não houver influências naturais ou artificiais) é que as proporções dos genótipos não mudam de uma geração para outra. Na geração que sucede a população cujos genótipos AA, Aa, aa estão presentes nas proporções $p^2 : 2pq : q^2$, os genótipos aparecerão nas mesmas proporções relativas.

Quando as frequências alélicas não mudam de geração para geração dizemos que elas estão em equilíbrio. E quando esse equilíbrio combina (se assemelha) com as proporções de H-W dizemos que há equilíbrio de Hardy-Weinberg (EHW) nesse específico locus.

Verificando o equilíbrio

Um dos métodos utilizados para verificar o equilíbrio de H-W é o teste qui-quadrado.

Suposições:

- A amostra é aleatória;
- A unidade de medida da variável é no mínimo nominal;

Verificando as suposições:

- Como não há forças perturbadoras, o acasalamento entre os pais foram aleatórias, então, nossa amostra é aleatória;
- Estamos usando variáveis nominais, pois as características que observamos são genótipos;

Teste de hipóteses

- **Hipóteses:**
- H_0 = Os genótipos do locus em questão estão em EHW;
- H_1 = Os genótipos do locus em questão não estão em EHW;
- **Nível de significância:** $\alpha = 0,05$

Teste de hipóteses

- Estatística de teste sob H_0 :

$$Q = \sum_{j=A}^a \sum_{i=B}^b \frac{(n_{ji} - e_{ji})^2}{e_{ji}}$$

em que e_{ji} é o número de observações esperado sob H_0 na linha j e na amostra i . Sob H_0 , Q tem aproximadamente a distribuição χ^2_1 .

Nesse caso, perdemos mais um grau de liberdade, pois estimamos a frequência alélica p_A .

- Regra de decisão:

Rejeitamos H_0 ao nível α de significância se o valor observado da estatística teste Q for maior ou igual a $\chi^2_{1-\alpha, K-1}$ em que $\chi^2_{1-\alpha, K-1}$ é o quantil $1 - \alpha$ da distribuição χ^2_{K-1} , ou seja,

$$RC = \{ Q : Q \geq \chi^2_{1-\alpha, K-1} \}$$

e valor-p = $\Pr_0(Q \geq Q_{obs})$.

Teste de hipóteses

- Conclusão

Tabela: P-valores dos testes qui-quadrado.

| Marcadores | p-valor |
|------------|---------|
| MK-1-1 | 0.6666 |
| MK-1-1 | 0.4996 |
| MK-1-3 | 0.05402 |
| MK-1-4 | 0.1959 |
| MK-1-5 | 0.5644 |
| MK-1-6 | 0.06952 |
| MK-1-7 | 0.09192 |

Com base na amostra observada, e usando $\alpha = 0.05$ de significância, temos evidências de que as regiões estão sob equilíbrio de Hardy-Weinberg, ou seja, os alelos maternos e paternos estão sendo herdados de forma independente.

Desequilíbrio de ligação

Um outro interesse comum na análise de dados genéticos é verificar a associação entre loci e analisar suas proporções alélicas, pois elas mostram a associação entre essas regiões genéticas. Se existir associação genética entre duas regiões do DNA dizemos que elas estão em desequilíbrio de ligação, mesmo que seja associação espúria (sem uma explicação natural ou razoável). O estado de associação aleatória (ou não associação) genética entre essas regiões é chamado de equilíbrio de ligação.

Importância: Geralmente genes ou regiões genéticas que estão em desequilíbrio de ligação afetam uma mesma característica física (fenótipo) de indivíduos, e isso ajuda muito na identificação de regiões que regulam ou afetam a característica física em estudo.

Desequilíbrio de ligação via gametas

Em uma primeira situação, os dados estão disponíveis em gametas (combinações dos alelos de diferentes regiões) e não há necessidade de conhecer as associações genotípicas. Para um par de alelos em dois loci, os procedimentos para definir, estimar e testar o desequilíbrio são muito parecidos com o que vimos para um par de alelos em um único locus. O coeficiente de desequilíbrio para os alelos $A(a)$ e $A(a)$ nas duas regiões comparadas é calculado como $D_{AA} = p_{AAP}p_{APa}$. Um teste qui-quadrado pode ser realizado para testar o desequilíbrio,

Hipóteses:

$$\begin{cases} H_0 : D_{AA} = 0 \\ H_1 : D_{AA} \neq 0 \end{cases}$$

- H_0 : As duas regiões estão em equilíbrio de ligação quando a frequência do alelo A na primeira região é independente da frequência do alelo A na segunda região e não estão associadas.
- H_1 : As duas regiões não estão em equilíbrio de ligação quando a frequência do alelo A na primeira região depende da frequência do alelo A na segunda região e estão associadas.

Teste de hipóteses

- **Nível de significância:**

Para controlar a probabilidade do erro tipo I global (considerando todas as regiões genéticas em estudo), precisamos aplicar a correção de Bonferroni no nível de significância de cada teste individual, isto é, seja $\alpha = 0,05$ o nível de significância desejado para o estudo todo, cada teste individual é feito com um nível de significância $\frac{\alpha}{21} = \frac{0,05}{21} = 0.002380952$, onde 21 é o número de testes que serão realizados.

Teste de hipóteses

- Estatística de teste sob H_0 :

$$Q = \sum_{j=A}^a \sum_{i=B}^b \frac{(n_{ji} - e_{ji})^2}{e_{ji}}$$

em que e_{ji} é o número de observações esperado sob H_0 na linha j e na amostra i . Sob H_0 , Q tem aproximadamente a distribuição χ^2_1

- Regra de decisão:

Rejeitamos H_0 ao nível α de significância se o valor observado da estatística teste Q for maior ou igual a $\chi^2_{1-\alpha, K-1}$ em que $\chi^2_{1-\alpha, K-1}$ é o quantil $1 - \alpha$ da distribuição χ^2_{K-1} , ou seja,

$$RC = \{ Q : Q \geq \chi^2_{1-\alpha, K-1} \}$$

e valor-p = $\Pr(Q \geq Q_{obs})$.

Conclusão

- Conclusão

Tabela: P-valores dos testes qui-quadrado

| Marcadores | p-valor | Marcadores | p-valor | Marcadores | p-valor |
|------------|---------|------------|---------|------------|---------|
| MK-1-2 | 2.2e-16 | MK-2-4 | 2.2e-16 | MK-3-7 | 2.2e-16 |
| MK-1-3 | 2.2e-16 | MK-2-5 | 2.2e-16 | MK-4-5 | 2.2e-16 |
| MK-1-4 | 2.2e-16 | MK-2-6 | 2.2e-16 | MK-4-6 | 2.2e-16 |
| MK-1-5 | 2.2e-16 | MK-2-7 | 2.2e-16 | MK-4-7 | 2.2e-16 |
| MK-1-6 | 2.2e-16 | MK-3-4 | 2.2e-16 | MK-5-6 | 2.2e-16 |
| MK-1-7 | 2.2e-16 | MK-3-5 | 2.2e-16 | MK-5-7 | 2.2e-16 |
| MK-2-3 | 2.2e-16 | MK-3-6 | 2.2e-16 | MK-6-7 | 2.2e-16 |

Com base na amostra observada e a um nível de significância de $\alpha = 0,002380952$, rejeitamos H_0 . Portanto, temos evidências que os marcadores do cromossomo 7, aos pares, estão em desequilíbrio de associação.

Intervalo de confiança

Desejamos estimar a relação entre a frequência do alelo dominante p_A e a frequência do alelo recessivo p_a em nossa amostra. Embora tenhamos os estimadores pontuais para essas frequências, não conseguimos obter uma solução analítica para a variância desses estimadores. Portanto, pretendemos utilizar a técnica de bootstrap para aproximar a variância e, conseqüentemente, construir um intervalo de confiança sem assumir uma distribuição.

Intervalo de confiança

Intervalos de confiança bootstrap usando o estimador $\hat{\theta} = \frac{\hat{p}A}{\hat{p}a}$. Com 95% de confiança, obtemos:

Tabela: Intervalos de confiança bootstrap para os marcadores do cromossomo 7.

| Marcadores | Intervalo |
|------------|----------------|
| MK-1-1 | (0.889, 1.110) |
| MK-1-2 | (0.868, 1.070) |
| MK-1-3 | (0.798, 1.000) |
| MK-1-4 | (0.895, 1.113) |
| MK-1-5 | (0.938, 1.166) |
| MK-1-6 | (0.952, 1.192) |
| MK-1-7 | (0.900, 1.106) |

Intervalo de confiança

Com base nos intervalos apresentados para razão das frequências alélicas em cada marcador, podemos observar que todos eles contiveram o valor 1. Assim, os intervalos de confiança indicam que não há evidências estatísticas de desequilíbrio significativo entre os alelos A e a em todas as regiões do cromossomo analisadas para a característica de densidade óssea das ratinhas. Isso significa que a relação entre as frequências dos alelos dominante e recessivo é aproximadamente igual em todas as regiões, sugerindo uma distribuição equilibrada desses alelos em relação à característica estudada.

Marcadores associados a densidade do fêmur

Uma maneira simples de analisar a relação entre as variações genéticas em uma região específica do genoma e características físicas (fenótipos), como a presença de uma doença (fenótipo binário) ou uma medida numérica (fenótipo quantitativo), é através do teste qui-quadrado e medidas de associação derivadas dele.

Nesse tipo de estudo, a ideia é testar se existe uma associação entre o genótipo de loci específicos e os diferentes estados do fenótipo em questão (por exemplo, doente ou não doente para um fenótipo binário ou diferentes níveis de uma característica para um fenótipo quantitativo). O teste qui-quadrado é aplicado para verificar se as frequências observadas dos genótipos são diferentes das frequências esperadas sob a hipótese nula de não associação.

Essas análises podem fornecer informações importantes sobre quais regiões genéticas estão envolvidas na regulação das características estudadas.

Marcadores associados à densidade do fêmur

Como os genótipos das regiões genéticas do cromossomo 7 são variáveis categóricas (do tipo 1, 0, -1), para medirmos associação entre eles e a densidade do fêmur foi necessário categorizar (discretizar) essas variáveis antes de medir a associação delas com o fenótipo (através do teste qui-quadrado).

As variáveis foram categorizadas em percentis, em que:

- Percentil 1: percentil em que 25% dos dados estão abaixo de 6.67 e 75% dos dados estão acima de 6.67.
- Percentil 2: percentil em que 50% dos dados estão abaixo de 6.69 e 75% dos dados estão acima de 6.69.
- Percentil 3: percentil em que 75% dos dados estão abaixo de 6.72 e 25% dos dados estão acima de 6.72.

bmd.png

Marcadores associados à densidade do fêmur

Com base na análise do gráfico da densidade óssea (BMD) do fêmur na escala logarítmica, vemos que:

- 168 indivíduos possuem uma BMD menor que 6.67;
- 159 indivíduos possuem uma BMD entre 6.67 e 6.69;
- 161 indivíduos possuem uma BMD entre 6.69 e 6.72;
- 163 indivíduos possuem uma BMD maior que 6.72;

Portanto, uma vez que as variáveis estão categorizadas, podemos dar continuidade com o teste qui-quadrado.

Teste de hipóteses

- **Hipóteses:**
- H_0 : O marcador não está em equilíbrio de ligação com a variável resposta densidade do fêmur
- H_1 : O marcador está em equilíbrio de ligação com a variável resposta densidade do fêmur.

Teste de hipóteses

- **Nível de significância:**

Para controlar a probabilidade do erro tipo I global (considerando todas as regiões genéticas em estudo), precisamos aplicar a correção de Bonferroni no nível de significância de cada teste individual, isto é, seja $\alpha = 0,05$ o nível de significância desejado para o estudo todo, cada teste individual é feito com um nível de significância $\frac{\alpha}{7} = \frac{0,05}{7} = 0.007142857$, onde 7 é o número de testes à serem realizados.

Teste de hipóteses

Estatística de teste sob H_0 :

$$Q = \sum_{j=A}^a \sum_{i=B}^b \frac{(n_{ji} - e_{ji})^2}{e_{ji}}$$

em que e_{ji} é o número de observações esperado sob H_0 na linha j e na amostra i . Sob H_0 , Q tem aproximadamente a distribuição χ^2_1

- **Regra de decisão:**

Rejeitamos H_0 ao nível α de significância se o valor observado da estatística teste Q for maior ou igual a $\chi^2_{1-\alpha, K-1}$ em que $\chi^2_{1-\alpha, K-1}$ é o quantil $1 - \alpha$ da distribuição χ^2_{K-1} , ou seja,

$$RC = \{ Q : Q \geq \chi^2_{1-\alpha, K-1} \}$$

e valor-p = $\Pr_0(Q \geq Q_{obs})$.

Teste de hipóteses

- Conclusão:

| Marcador e BMD | p-valor | Estatística teste |
|----------------|--------------|-------------------|
| MK-1-1 e BMD | 0.349924080 | 6.695531 |
| MK-1-2 e BMD | 0.567503653 | 4.816892 |
| MK-1-3 e BMD | 0.264436389 | 7.655595 |
| MK-1-4 e BMD | 0.4497935986 | 5.766974 |
| MK-1-5 e BMD | 0.170946432 | 9.047101 |
| MK-1-6 e BMD | 0.004472395 | 18.823467 |
| MK-1-7 e BMD | 0.011314303 | 16.498492 |

Com base na amostra observada do cromossomo 7 e considerando um nível de significância de $\alpha = 0.00714285$, concluímos que a hipótese nula H_0 não foi rejeitada na maioria dos casos, com exceção do MK-1-6 e BMD. Em outras palavras, temos evidências de que o marcador 6 está em equilíbrio de ligação com a densidade óssea do fêmur e outros marcadores analisados não estão em equilíbrio de ligação a densidade óssea do fêmur.

Concluimos então que, o marcador 6 explica a densidade do fêmur.

Conclusões

Com base na análise dos marcadores do cromossomo 7, podemos concluir que:

- Com base nas frequências obtidas para cada marcador genético, determinamos qual alelo é o mais raro em cada caso.
Para o marcador 1, o alelo mais raro foi o dominante. Para o marcador 2, o alelo mais raro foi o dominante. Para o marcador 3, o alelo mais raro foi o dominante. Para o marcador 4, o alelo mais raro foi o dominante. Para o marcador 5, o alelo mais raro foi o recessivo. Para o marcador 6, o alelo mais raro foi o recessivo. Para o marcador 7, o alelo mais raro foi o recessivo.
- A relação entre as frequências alélicas p_A e p_a para cada marcador pode estar associada a uma baixa variação genética na população estudada. Isso pode ter ocorrido porque analisamos indivíduos com características muito semelhantes entre si.
- Os marcadores do cromossomo 7 estão em equilíbrio de Hardy-Weinberg.
- Os marcadores do cromossomo 7 estão em desequilíbrio de associação entre si. Concluimos também que, os marcadores estão em EHW, ou seja, os alelos maternos e paternos estão sendo herdados de forma independente.

Conclusão

- O marcador genético que parece apresentar mais desbalanceamento entre os alelos dominantes e recessivos é o marcador 7, uma vez que o seu coeficiente de desequilíbrio é maior.
- O marcador que está mais associado com a densidade óssea do fêmur (BMD) é o marcador 6.

Referências Bibliográficas

- [1] WEIR, B. S. Genetic data analysis: methods for discrete population geneticdata, 2a edição, Sunderland: Sinauer Associates. 1996.
- [2] Daiane Aparecida Zuanetti Notas de aula da disciplina Tópicos em Estatística Genética.