

**Universidade Federal de Minas Gerais**  
**Instituto de Ciências Exatas**  
**Departamento de Estatística**

**TESTES BOOTSTRAP PARA A FRAÇÃO DE  
NÃO-DISJUNÇÃO MEIÓTICA EM PACIENTES  
COM SÍNDROME DE DOWN**

**Paula Arantes Barros**  
**Glaura da Conceição Franco**

**Relatório Técnico**  
**RTP- 01/2002**  
**Série Pesquisa**

***TESTES BOOTSTRAP PARA A FRAÇÃO DE NÃO-DISJUNÇÃO  
MEIÓTICA EM PACIENTES COM SÍNDROME DE DOWN***

**Paula Arantes Barros<sup>\*</sup>  
Glaura da Conceição Franco**

**Departamento de Estatística - UFMG**

\* Departamento de Estatística – ICEX – UFMG – Caixa Postal 702  
Belo Horizonte – MG - Brazil – 31.270-901  
[pab@est.ufmg.br](mailto:pab@est.ufmg.br) - Tel: 55(31)3499-5910 – Fax: 55(31)3499-5924

**AGRADECIMENTOS**

Este trabalho contou com o apoio do CNPq através da bolsa PIBIC para a aluna Paula Arantes Barros. Gostaríamos também de agradecer ao Prof. Dr. Sérgio Danilo Pena e Flávia C. Parra, do Departamento de Bioquímica e Imunologia da UFMG, pelo material da pesquisa.

## RESUMO

As trissomias cromossômicas ocorrem principalmente devido ao processo de não-disjunção na meiose. Com o objetivo de entender melhor o mecanismo subjacente da não-disjunção, é necessário estabelecer a proporção de casos que ocorrem na primeira ou segunda divisão meiótica. Pacientes trissômicos apresentam, em estudos de microssatélites empregando a Reação em Cadeia da Polimerase (RCP), três fragmentos de igual intensidade, dois fragmentos a uma taxa de 2:1 ou um único fragmento. Neste trabalho, é apresentado um modelo de probabilidade para o número de picos em um locus de microssatélite polimórfico, que é uma função da fração de não-disjunção na meiose I,  $F$ . Baseado neste modelo, obtém-se um estimador de máxima verossimilhança para  $F$ , assim como sua variância assintótica, usando a proporção observada de um, dois e três alelos em indivíduos com trissomia do cromossomo 21. Devido ao fato de que as trissomias cromossômicas são eventos raros, o uso da teoria assintótica pode ficar comprometida. Assim, a técnica bootstrap é empregada para a construção de testes de hipóteses simples para  $F$ . Neste caso, duas estatísticas de teste são propostas e o tamanho dos testes comparados através de simulações Monte Carlo.

**Palavras-Chaves:** Trissomia, não-disjunção, Máxima Verossimilhança, Bootstrap, Testes de Hipóteses.

## ABSTRACT

Chromosome trisomy primarily arises by the process of nondisjunction in meiosis. In order to increase our understanding of the mechanisms underlying nondisjunction, it is necessary to establish the proportion of cases occurring in first or second division of meiosis. Trisomic patients will display, in study of microsatellite by employing Polymerase Chain Reaction (PCR) based approach, three fragment peak of equal intensity, two fragments at an average 2:1 dosage or one individual fragment. In this work we present a probability model for the number of peaks in a polymorphic microsatellite locus, which is a function of the fraction of meiosis I nondisjunction,  $F$ . Based on this model, we obtain a maximum likelihood estimator for  $F$ , using the observed proportion of one, two and three allele patterns in individuals with chromosome trisomy. Relying on the properties of maximum likelihood theory, we also calculate the asymptotic variance of  $F$ . To test simple hypothesis about  $F$ , we propose two test statistics. Due to the fact that chromosome trisomies are rare events, the use of asymptotic theory may be compromised. Thus, we employ the bootstrap technique to build hypothesis tests for  $F$ .

**Keywords:** Trisomy, Nondisjunction, Maximum Likelihood, Bootstrap, Hypothesis Tests.

# 1. INTRODUÇÃO

A trissomia do cromossomo 21 é o resultado da não-disjunção primária, que pode ocorrer em ambas as divisões meióticas e em ambos os pais. Os cromossomos pareados não se separam de forma apropriada na meiose fazendo com que um dos gametas receba dois cromossomos 21 e o outro nenhum. Esta anomalia cromossômica numérica produz um fenótipo conhecido como síndrome de Down.

Um dos pontos importantes para o entendimento da epidemiologia desta trissomia é determinar se a não-disjunção ocorre na primeira ou na segunda divisão meiótica. Até agora a estimativa de  $F$  (a probabilidade de que, em indivíduos trissômicos, a não-disjunção ocorra na meiose I) tem feito uso do estudo comparativo de marcadores genéticos na criança afetada e em seus pais. A necessidade do estudo dos pais complica estes estudos e impossibilita o uso de material de arquivo. Uma outra alternativa utilizada pelos pesquisadores é fazer a análise do fracionamento por densitometria a laser utilizando a amplificação através da Reação em Cadeia da Polimerase (PCR). (Parra, 1999).

No estudo de locus de microssatélites polimórficos pela reação em cadeia da polimerase (PCR), pacientes trissômicos apresentam: 1) a presença de três picos de mesma intensidade, 2) dois picos com uma dosagem relativa 2:1 ou 3) apenas um pico (casos não-informativos). Para a ocorrência do padrão de 3 picos em um feto trissômico, é necessário que a não-disjunção ocorra na meiose I e que sejam transmitidos 3 alelos diferentes dos pais.

Se conhecermos a proporção de casos de não-disjunção na primeira divisão meiótica e as frequências relativas ( $p_1, p_2, \dots, p_m$ ) dos  $m$  alelos de um loco multialélico de microssatélites, podemos aplicar o equilíbrio de Hardy-Weinberg para calcular a frequência relativa esperada dos padrões de 1, 2 e 3 picos, já que as frequências genótípicas têm distribuição multinomial (Guo e Thompson, 1992).

Um dos objetivos deste trabalho é estimar a proporção  $F$  de não-disjunção na meiose I, usando a proporção observada de padrões de um, dois ou três picos em uma amostra de indivíduos trissômicos, através de um estimador de máxima verossimilhança (EMV). O cálculo deste estimador envolve derivadas de funções complicadas, que exigem o uso de um software matemático para sua solução. Utilizando propriedades assintóticas do EMV é possível calcular a variância assintótica destes estimadores através da matriz de informação observada (Welsh, 1996).

O outro objetivo é desenvolver testes bootstrap (Efron e Tibshirani, 1993) para testar hipóteses simples sobre  $F$ . Duas estatísticas de teste são propostas e comparadas através de simulações Monte Carlo para calcular o tamanho dos testes. Como as trissomias cromossômicas são eventos raros, as amostras obtidas são geralmente pequenas, o que pode colocar em risco a validade da utilização de testes assintóticos. Assim, para determinar a distribuição empírica da estatística de teste sob a hipótese de igualdade das frações de não-disjunção será utilizada a técnica bootstrap (Efron e Tibshirani, 1993).

Uma aplicação a dados reais, com amostras de sangue de 34 indivíduos trissômicos, é apresentada. Para esta amostra, o estimador de máxima verossimilhança de  $F$  é calculado e testes de hipóteses são realizados.

## 2. GENÉTICA

### 2.1 A TRISSOMIA DO CROMOSSOMO 21

A meiose é um processo de reprodução sexual através do qual o número cromossômico das células germinativas diplóides ( $2n$ ) é reduzido à metade ( $n$ ) na formação das células reprodutivas maduras ou gametas. Em um indivíduo normal, essa redução é obtida através de duas divisões celulares sucessivas, durante as quais os cromossomos são duplicados apenas uma vez. Procedendo à primeira das duas divisões, os cromossomos homólogos são pareados e duplicados. Esses cromossomos pareados são separados na primeira divisão meiótica da célula. Na segunda divisão meiótica, as duas cromátides de cada cromossomo separam-se, produzindo um total de quatro células sexuais maduras, cada uma capaz de fertilização. Uma explicação mais detalhada do processo de meiose pode ser vista em Gardener e Swstad (1986).

A trissomia do cromossomo 21 é o resultado da não-disjunção primária, que pode ocorrer em ambas as divisões meióticas e em ambos os pais. Os cromossomos pareados não se separam de forma apropriada para os pólos na anáfase; um dos gametas receberá dois cromossomos 21 e o outro nenhum. As figuras 1a e 1b apresentam processos de meiose com a não-disjunção ocorrendo na 1ª e 2ª fases, respectivamente. Sem perda de generalidade, ilustraremos o processo com a não-disjunção ocorrendo na mãe de um indivíduo trissômico.

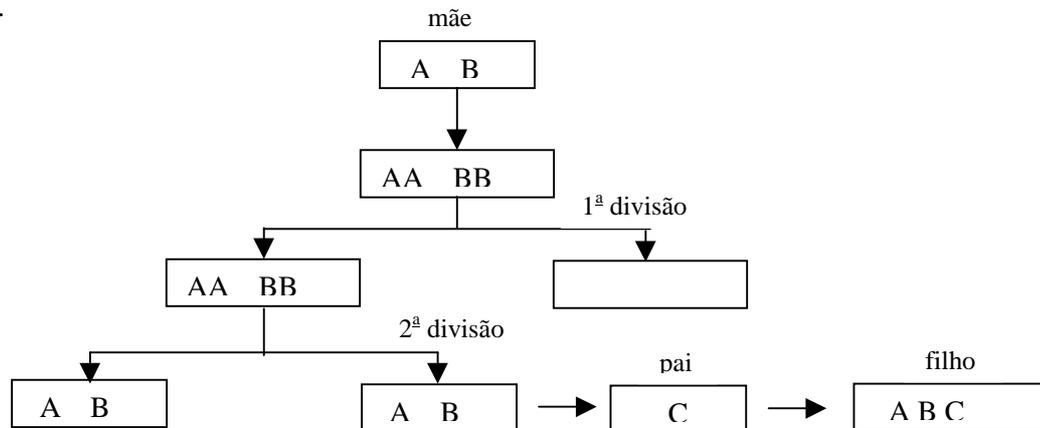


Figura 1a: Não-disjunção na 1ª fase da meiose em indivíduos trissômicos

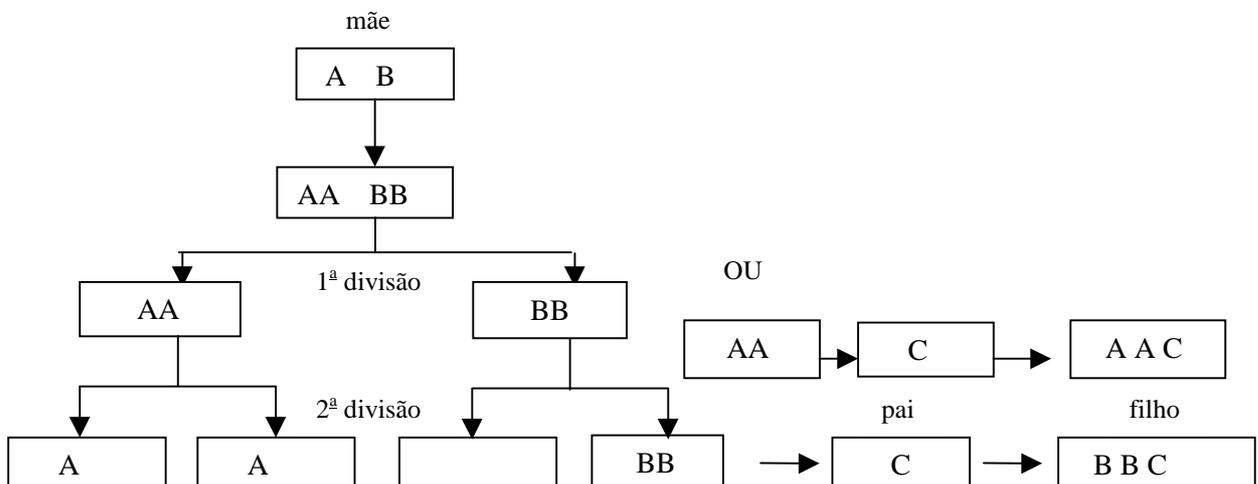


Figura 1b: Não-disjunção na 2ª fase da meiose em indivíduos trissômicos

## 2.2 ANÁLISE DA ORIGEM DAS TRISSOMIAS ATRAVÉS DA PCR

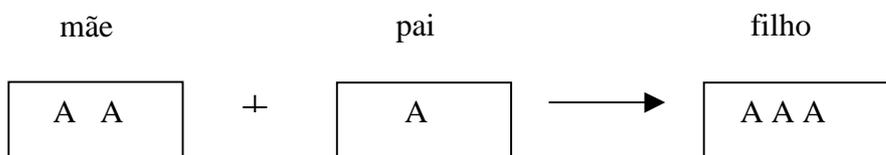
A análise da origem das trissomias é feita através da utilização de marcadores moleculares dos cromossomos, explorando regiões polimórficas do DNA. Por serem altamente informativos, os polimorfismos de DNA permitem a determinação da origem das trissomias na maioria dos casos, além de propiciarem uma análise mais objetiva.

A introdução da técnica de amplificação do DNA através da Reação em Cadeia da Polimerase (PCR) permitiu a identificação de um grande número de marcadores denominados microssatélites, mapeados ao longo dos diversos cromossomos (Pena, 1998). Os microssatélites são unidades curtas de DNA formadas por 1 a 5 pares de bases que se repetem seguidamente. Estes marcadores puderam ser estudados em conjunto através da PCR, possibilitando um resultado mais rápido e ainda mais informativo.

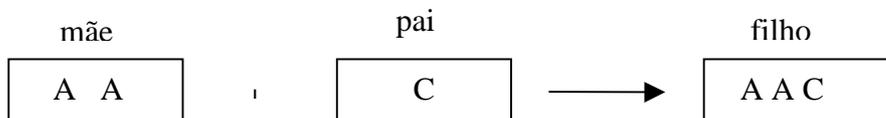
## 2.3 PADRÕES DE PICOS EM INDIVÍDUOS TRISSÔMICOS

Em estudos com microssatélites polimórficos, indivíduos portadores de trissomias apresentam padrões de um, dois ou três picos (ver figura 2). O padrão de um pico indica a igualdade entre os três alelos herdados pelo indivíduo. Padrões de dois picos, normalmente, estão em uma dosagem relativa de 2:1, indicando a presença de dois alelos iguais e um diferente e o padrão de três picos de mesma intensidade equivale à presença de três alelos com tamanhos diferentes.

I) 1 PICO (três alelos iguais):



II) 2 PICOS (dois alelos iguais):



III) 3 PICOS (três alelos diferentes):

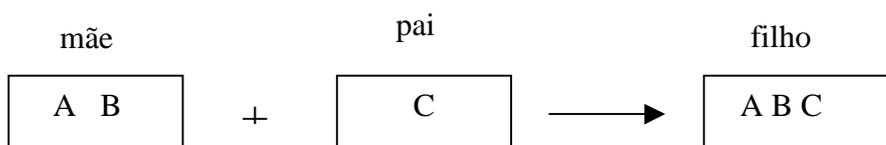


Figura 2: Desenho esquemático do número de picos

A proporção relativa de cada um dos padrões de um, dois e três picos vai depender, principalmente, do momento do acidente meiótico, além do índice de heterozigosidade do loco na população. Ignorando recombinações, para que ocorra um padrão de três picos, é necessário que o erro na segregação dos cromossomos tenha ocorrido na primeira divisão meiótica e que sejam transmitidos 3 alelos diferentes dos pais, provenientes de dois cromossomos homólogos de um dos genitores e um cromossomo diferente, herdado pelo outro genitor. A não ser quando o genitor que transmitiu os dois cromossomos homólogos seja homozigoto, o padrão de dois picos retrata uma situação em que a não-disjunção acontece na segunda divisão, quando as cromátides irmãs permanecem unidas.

A detecção molecular de trissomias humanas usando PCR e densitometria a laser (Pena, 1998) surgiu, portanto, como uma poderosa ferramenta alternativa para o diagnóstico aplicado em tecidos humanos que não estão em divisão, o que não é permitido na citogenética convencional.

### 3. DETERMINAÇÃO DO MODELO DE PROBABILIDADE

Como visto na Seção 2.3, os padrões de picos detectados pela técnica de PCR em indivíduos trissômicos estão totalmente relacionados com o momento em que a não-disjunção ocorreu na meiose. Assim, estabelecendo o modelo de probabilidade associado a estes padrões, que são a única resposta a que se tem acesso, é possível obter maiores informações sobre a probabilidade de não-disjunção. Ao longo de todo este trabalho, assume-se que o espaço amostral é formado somente por indivíduos trissômicos, ou seja, por indivíduos para os quais a não-disjunção ocorreu em uma das duas fases da meiose.

Seja então  $X$  a variável aleatória que denota o número de picos em indivíduos portadores de trissomias cromossômicas. Para determinar a distribuição de probabilidade de  $X$ , seja a variável  $ND$  definida abaixo:

$$ND = \begin{cases} 1 & \text{se a não - disjunção ocorre na 1ª fase da meiose} \\ 0 & \text{se a não - disjunção ocorre na 2ª fase da meiose} \end{cases}$$

Assim,

$$\begin{aligned} P_X(x) &= P[\{X = x, ND = 1\} \cup \{X = x, ND = 0\}] \\ &= P(X = x, ND = 1) + P(X = x, ND = 0) - P[(X = x, ND = 1) \cap (X = x, ND = 0)] \end{aligned}$$

Como a não-disjunção não pode ocorrer em ambas as fases ao mesmo tempo,

$$P[(X = x, ND = 1) \cap (X = x, ND = 0)] = 0.$$

Então,  $p_X(x) = P(X = x, ND = 1) + P(X = x, ND = 0)$

$$\Rightarrow p_X(x) = P(X = x / ND = 1) \times P(ND = 1) + P(X = x / ND = 0) \times P(ND = 0).$$

Chamando  $P(ND = 1) = F$  e  $P(ND = 0) = 1 - F$ , temos

$$p_X(x) = P(X = x / ND = 1) \times F + P(X = x / ND = 0) \times (1 - F)$$

Logo, vemos que a função de distribuição de  $X$  é função da fração desconhecida,  $F$ , de não-disjunção na meiose I.

Seja  $p_i$  a probabilidade de o alelo  $i$  estar presente em um indivíduo,  $i=1, \dots, m$ . Para a ocorrência do padrão de 1 pico, a mãe deve ser homocigota e o alelo do pai deve ser igual ao da mãe. Assim,

$$\begin{aligned} P(X = 1) &= F \times P(X = 1 / ND = 1) + (1 - F) \times P(X = 1 / ND = 0) \\ &= F \sum_{i=1}^m p_i^3 + (1 - F) \sum_{i=1}^m p_i^2 \end{aligned}$$

Para ocorrer o padrão de 2 picos há duas possibilidades:

- a) se a mãe for homocigota, o alelo do pai, necessariamente, terá que ser diferente dos alelos da mãe;
- b) se a mãe for heterocigota, o alelo do pai terá que ser igual a qualquer um dos alelos da mãe.

Logo,

$$\begin{aligned} P(X = 2) &= F \times P(X = 2 / ND = 1) + (1 - F) \times P(X = 2 / ND = 0) \\ &= F \times 3 \sum_{i=1}^m \sum_{j=1}^m p_i^2 p_j + (1 - F) \times 2 \sum_{i=1}^m \sum_{j=1}^m p_i p_j, \text{ com } i \neq j \end{aligned}$$

Para a ocorrência do padrão de 3 picos, a mãe deverá ser, obrigatoriamente, heterocigota e o alelo do pai deverá ser diferente dos alelos da mãe. Além disto, como o padrão de 3 picos nunca vai ocorrer se a não-disjunção meiótica for na 2ª fase, temos que  $P(X=3 / ND_2) = 0$ . Logo,

$$\begin{aligned} P(X = 3) &= P(X = 3 / ND = 1) \times F + P(X = 3 / ND = 0) \times (1 - F) \\ &= F \times 6 \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m p_i p_j p_k, \text{ com } i \neq j \neq m. \end{aligned}$$

Finalmente, seja a variável aleatória  $Y_i$  denotando o número de pessoas, em uma amostra de tamanho  $n$ , que apresentam  $i$  picos,  $i=1,2,3$ . Pode-se concluir então que o modelo usado para descrever o comportamento de  $\mathbf{Y}=(Y_1, Y_2, Y_3)$  será uma Multinomial  $(n, \theta_1, \theta_2, \theta_3)$ , onde  $\theta_1, \theta_2$  e  $\theta_3$  são dados respectivamente por:

$$\theta_1 = P(X = 1) = F \sum_{i=1}^m p_i^3 + (1 - F) \sum_{i=1}^m p_i^2 \quad (3.1a)$$

$$\theta_2 = P(X = 2) = F \times 3 \sum_{i=1}^m \sum_{j=1}^m p_i^2 p_j + (1 - F) \times 2 \sum_{i=1}^m \sum_{j=1}^m p_i p_j \quad (3.1b)$$

$$\theta_3 = P(X = 3) = F \times 6 \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m p_i p_j p_k \quad (3.1c)$$

onde  $0 \leq F \leq 1$ .

#### 4. ESTIMAÇÃO DE $F$ E DA VARIÂNCIA ASSINTÓTICA

##### 4.1 ESTIMADOR DE $F$

Vimos anteriormente que  $f_Y(y, \theta) \sim$  Multinomial ( $n; \theta_1, \theta_2, \theta_3$ ). Devemos notar que os  $\theta_i$ 's,  $i = 1, 2, 3$ , são funções do parâmetro  $F$  (ver equação 3.1) que desejamos estimar.

Assim, podemos calcular a função de verossimilhança, que é dada por:

$$L(y, F) = f_Y(y, \theta) = \frac{n!}{n_1! n_2! n_3!} \theta_1^{n_1} \times \theta_2^{n_2} \times (1 - \theta_1 - \theta_2)^{n - n_1 - n_2} \quad (4.1)$$

onde:  $n$  = número de indivíduos com trissomia

$n_1$  = número de indivíduos com o padrão de 1 pico

$n_2$  = número de indivíduos com o padrão de 2 picos

$n_3$  = número de indivíduos com o padrão de 3 picos

e os  $\theta_i$ 's são dados na equação (3.1).

O estimador de máxima verossimilhança (EMV) de  $F$  é encontrado maximizando-se a função de verossimilhança  $L$  em relação a  $F$  (Mood, *et al.* (1974)).

Utilizando-se o software Mathematica (Blachman, 1992) para maximizar esta função, encontramos o EMV de  $F$ , que é dado por:

$$\hat{F} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \quad (4.2)$$

onde  $A = n(a - b)(3c - 1 + b)$ ,  $B = (n - n_2)(a - b)(1 - b) + b(n - n_1)(3c - 1 + b)$ ,

$C = b(n - n_1 - n_2)(1 - b)$ ,  $a = \sum_{i=1}^m p_i^3$ ,  $b = \sum_{i=1}^m p_i^2$  e  $c = \sum_{i=1}^m \sum_{j=1}^m p_i^2 p_j$ ,  $i \neq j$ .

## 4.2 VARIÂNCIA ASSINTÓTICA

Como foi definido no item anterior  $\hat{F} = \hat{F}(Y_1, \dots, Y_n)$  é o estimador de máxima verossimilhança de  $F$  baseado numa amostra aleatória de tamanho  $n$ . Pelas propriedades da teoria assintótica (Mood, *et al.*, 1974), temos que, para  $n$  suficientemente grande, a distribuição do EMV de  $F$  é aproximadamente

$$\hat{F} \sim N(F, I^{-1}(F))$$

onde  $I = \left[ E \left[ \frac{\partial^2 \log L(F)}{\partial F^2} \right] \right] \Big|_{F=\hat{F}}$  é conhecido como a matriz de informação esperada de Fisher.

Como na maioria das vezes o cálculo desta matriz é muito trabalhoso, então utilizamos um estimador para  $I$  conhecido como a matriz de informação observada ( $\mathbf{I}$ ) (Welsh, 1996), dada por:

$$I = \frac{\partial^2 \log L(F)}{\partial F^2} \Big|_{F=\hat{F}}$$

No nosso caso,  $I$  é dada por<sup>1</sup>:

$$I = \frac{\partial^2 \log L(F)}{\partial F^2} \Big|_{F=\hat{F}} = \frac{(a-b)^2 n_1}{(b + (a-b)\hat{F})^2} + \frac{n - n_1 - n_2}{\hat{F}^2} + \frac{(3c + b - 1)^2 n_2}{(1 - b + (-1 + b + 3c)\hat{F})^2}$$

onde  $\hat{F}$  é calculado como em (4.2).

Assim a variância assintótica de  $F$  é dada por:

$$\text{Vass}(F) = 1 / I \quad (4.3)$$

<sup>1</sup> A derivada 2ª em relação a  $F$  foi calculada utilizando o software Mathematica (Blachman, 1992).

## 5. TESTE DE HIPÓTESES BOOTSTRAP PARA A FRAÇÃO DE NÃO-DISJUNÇÃO MEIÓTICA

Neste capítulo apresentamos duas estatísticas de teste que serão utilizadas para testar hipóteses sobre a fração de não-disjunção  $F$ . Através da técnica de bootstrap, obtemos as distribuições empíricas destas estatísticas e calculamos as probabilidades de significância associadas.

### 5.1 ESTATÍSTICAS DE TESTE

Supondo um valor fixo  $F_0$ , as hipóteses simples a serem testadas são:

$$\left\{ \begin{array}{l} H_0: F = F_0 \\ H_1: F \neq F_0 \end{array} \right.$$

Para se testar as hipóteses acima serão utilizadas duas estatísticas de teste:

- **Est<sub>1</sub>**: O próprio EMV de  $F$ ,  $\hat{F}$ , dado em (4.2);
- **Est<sub>2</sub>**: A estatística  $T^* = \frac{\hat{F} - F_0}{d.p.(\hat{F})}$ , onde  $d.p.(\hat{F})$  é o desvio-padrão de  $\hat{F}$  sob  $H_0$ .

Como no estudo de trissomias cromossômicas o tamanho das amostras obtidas é, na maioria das vezes, muito pequeno, serão utilizados testes bootstrap (Efron e Tibshirani, 1993) para testar hipóteses acerca de  $F$ .

### 5.2 TESTES BOOTSTRAP

A idéia básica de um teste bootstrap é a de se retirar um grande número de amostras bootstrap, obedecendo à hipótese nula e que, tanto quanto possível, se assemelhem à amostra real, e assim comparar a estatística de teste observada com as estatísticas calculadas das amostras bootstrap.

É possível usar o bootstrap tanto para calcular um valor crítico para a estatística de teste  $T$ , quanto para calcular a probabilidade de significância, ou P-valor, do teste, definido como

$$p = \Pr(T \geq t / H_0)$$

onde  $t$  é o valor observado da estatística de teste.

Em muitos problemas a distribuição da estatística de teste  $T$  sob  $H_0$  dependerá de parâmetros de perturbação que não podem ser condicionados. Então a aproximação natural é ajustar o modelo nulo  $\hat{M}_0$  e usar  $p = \Pr(T \geq t / \hat{M}_0)$  para calcular o P-valor.

Se a expressão  $p = \Pr(T \geq t / \hat{M}_0)$  não pode ser calculada exatamente ou se não há aproximação satisfatória (normal ou outra), deve-se proceder pela simulação. Isto é,  $B$  amostras independentemente

replicadas  $x_1^*, \dots, x_n^*$  são geradas de  $\hat{M}_0$  e, para a  $b$ -ésima amostra, o valor da estatística de teste  $t_b^*$  é calculado. Então a probabilidade de significância será aproximada por:

$$p_{boot} = \frac{1 + \#\{t_b^* \geq t\}}{B + 1} . \quad (5.1)$$

Maiores detalhes sobre testes bootstrap podem ser encontrados em Davison and Hinkley (1997).

### 5.3 CÁLCULO DO TAMANHO DO TESTE

Como visto em (3.1), a distribuição do número de picos segue uma multinomial, com as probabilidades sendo funções da fração de não-disjunção meiótica,  $F$ . Assim, para se determinar a distribuição empírica das estatísticas de teste, deve-se gerar  $B$  amostras bootstrap da multinomial sob  $H_0$  (ver programa em Anexo). Para cada uma das  $B$  amostras bootstrap calcula-se o EMV para  $F$ .

Para determinar a precisão dos testes obtidos utilizando-se as estatísticas propostas na seção 5.1 realizamos simulações Monte Carlo para calcular o tamanho do teste em ambos os casos. Os testes baseados em simulações Monte Carlo (Davison e Hinkley, 1997) comparam estatísticas observadas  $t$  para  $R$  valores independentes de  $T$ , que são obtidas das amostras correspondentes simuladas independentemente sob o modelo da hipótese nula. Sob  $H_0$  os valores simulados, denotados por  $t_1^*, \dots, t_R^*$ , e a estatística observada  $t$  são igualmente prováveis. Se exatamente  $k$  dos  $t^*$  valores simulados excedem  $t$ , então

$$\alpha_0 = \Pr(T \geq t \mid H_0) \cong \frac{k + 1}{R + 1} \quad (5.2)$$

O resultado acima fornece o Tamanho do Teste para a estatística de interesse.

## 6. APLICAÇÃO

Neste estudo um estimador de máxima verossimilhança para a fração de não-disjunção na meiose I é calculado, baseado em uma amostra aleatória composta do sangue de 34 indivíduos brasileiros com trissomia do cromossomo 21. Estes indivíduos foram aleatoriamente selecionados entre os pacientes atendidos no GENE – Núcleo de Genética Médica, um serviço de genética médica, localizado em Belo Horizonte, MG.

Para se determinar os padrões do número de picos foi escolhido o microssatélite D21S215, devido ao seu polimorfismo e, principalmente, à sua localização na região pericentromérica do cromossomo 21, o que indica que não há evidências da ocorrência de recombinação com regiões do centrômero (Shen *et. al.*, 1998). Foram observados neste microssatélite, composto de 14 repetições de CA (par de bases), 6 alelos, que apresentaram um polimorfismo com frequências de  $p_1=0,12$ ,  $p_2= 0.45$ ,  $p_3=0.09$ ,  $p_4=0.31$ ,  $p_5=0.01$  e  $p_6=0.02$ , respectivamente, para os alelos CA-13, CA-14, CA-15, CA-16, CA-17 e CA-19.

Através da técnica da PCR, é possível determinar o padrão do número de picos de cada indivíduo. Analisando a amostra, verifica-se que a frequência para o padrão de 1 pico é aproximadamente 17,65%, 2 picos é 64,7% e 3 picos também é 17,65%.

O EMV de  $F$  é calculado utilizando-se a fórmula (4.2). Com o número de picos pode-se calcular o estimador de  $F$ ,  $\hat{F}$ , e a variância assintótica ( $V_{ass}$ ) para esta amostra, que são dados respectivamente por,  $\hat{F} = 0,65$  e  $V_{ass} = 0.0481$ .

O próximo passo é construir testes bootstrap para a fração de não-disjunção meiótica,  $F$ . Trabalhos anteriores estimavam  $F$  através de estudos comparativos de marcadores polimórficos pericentroméricos de DNA em crianças afetadas e seus pais (Parra, 1999). Utilizando alguns destes estudos, o valor de  $F_0$  foi escolhido como sendo igual a 0.68 (Petersen et al., 1992).

Assim, deseja-se testar:

$$\begin{cases} H_0: F = 0,68 \\ H_1: F \neq 0,68 \end{cases} \quad (6.1)$$

Para se determinar a distribuição empírica das duas estatísticas de teste, foi gerada uma amostra da distribuição multinomial, com  $F = 0,68$  e os valores de  $p_i$ ,  $i=1,\dots,6$ , como dados acima. Portanto, sob  $H_0$  foi gerada uma amostra que tem distribuição multinomial ( $N=34$ ;  $\theta_1=0,19$ ,  $\theta_2=0,62$ ,  $\theta_3=0,18$ ). Os valores de  $\theta_1$ ,  $\theta_2$  e  $\theta_3$  são obtidos utilizando-se a equação (3.1). A partir daí foram introduzidas as técnicas bootstrap.

Da amostra gerada sob  $H_0$  foram geradas  $B=1000$  amostras com reposição. Para cada amostra foi calculado o estimador de máxima verossimilhança da proporção  $F$  de não-disjunção na meiose I, denotados por  $\hat{F}_1^*, \dots, \hat{F}_{1000}^*$ .

Para a primeira estatística de teste,  $Est_1$ , as estimativas  $\hat{F}_1^*, \dots, \hat{F}_{1000}^*$  foram utilizadas para se determinar a distribuição empírica de  $\hat{F}$  (Figura 4.1). Pela figura pode-se ver que a distribuição é assimétrica e está mais concentrada acima do valor 0,5.

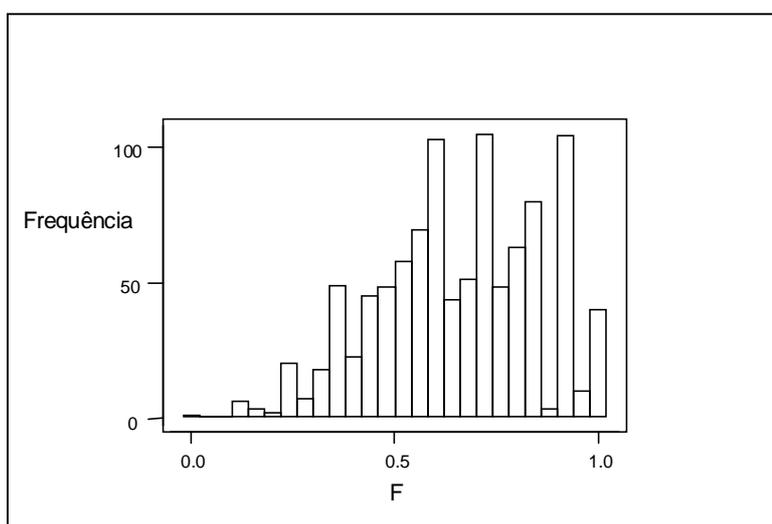


Figura 4.1: Histograma da Distribuição Empírica de  $\hat{F}$  sob  $H_0$

Para a segunda estatística de teste, **Est<sub>2</sub>**, a distribuição empírica de  $T^*$  foi determinada a partir das estatísticas  $t_1^* = \frac{\hat{F}_1^* - F_0}{sd(\hat{F}_1^*)}, \dots, t_B^* = \frac{\hat{F}_B^* - F_0}{sd(\hat{F}_B^*)}$ . Para cada uma das  $B=1000$  amostras foram geradas 200 amostras com reposição para se calcular o desvio-padrão de cada uma das estimativas bootstrap de  $F$ . Em seguida foi encontrada a distribuição empírica das  $t_1^*, \dots, t_{B=1000}^*$  (ver Figura 4.2). Pela figura nota-se que a distribuição parece ser assimétrica à esquerda sendo distribuída em torno do valor 0.

Através de um programa implementado na linguagem C++, geramos amostras de uma multinomial  $(N, \theta_1, \theta_2, \theta_3)$  sob  $H_0$ , e calculamos as estatísticas de teste para testar a hipótese de interesse (6.1). Estas distribuições empíricas foram utilizadas para calcular valores críticos para as estatísticas de testes propostas.

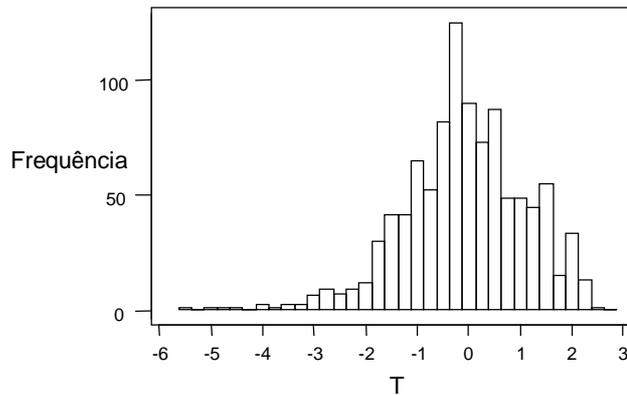


Figura 4.2: Histograma da Distribuição Empírica de  $\hat{T}$  sob  $H_0$

Em um estudo do Tamanho do Teste ( $\alpha_0$ ), ao nível de 5% de significância, foram feitas  $R=1000$  simulações Monte Carlo com  $B=1000$  replicações bootstrap. Para este caso, o Tamanho do Teste foi igual a:

$$\mathbf{Est}_1: \hat{F} \Rightarrow \alpha_0 = 0,108$$

$$\mathbf{Est}_2: T^* \Rightarrow \alpha_0 = 0,047.$$

Pelos resultados acima, a estatística de teste  $T^*$  parece ser mais adequada, pois apresentou um valor de  $\alpha_0$  mais próximo do nível nominal de 5%.

Para a amostra de 34 indivíduos, utilizando-se 200 replicações bootstrap para o cálculo do desvio-padrão de  $\hat{F}$ , temos os seguintes resultados:

$$\mathbf{Est}_1: \hat{F} \Rightarrow \text{valor-p} = 0,94$$

$$\mathbf{Est}_2: T^* \Rightarrow \text{valor-p} = 0,94$$

A probabilidade de significância neste caso foi 0.94 para ambos os testes e, portanto, concluímos que a hipótese nula não será rejeitada. Em outras palavras, não parece haver diferença no valor da fração de não-disjunção quando comparada a resultados obtidos através de estudos comparativos de marcadores de DNA.

## 7. CONCLUSÕES

Nesse trabalho mostrou-se que o número de picos em um locus de microsatélite polimórfico tem distribuição multinomial, com os parâmetros sendo uma função da fração de não-disjunção,  $F$ , na meiose I. Baseado em uma amostra de sangue de indivíduos trissômicos, foi calculada a estimativa de máxima verossimilhança para  $F$ , assim como sua variância assintótica. Como trissomias cromossômicas são eventos raros, a técnica bootstrap foi utilizada para construir testes de hipóteses para  $F$ . Duas estatísticas de teste foram propostas,  $\hat{F}$  e  $T^* = \frac{\hat{F} - F_0}{d.p.(\hat{F})}$ , com a segunda apresentando uma melhor performance, quando da comparação dos níveis nominais empíricos.

Até agora, estimativas de  $F$  têm sido obtidas através de estudos comparativos de marcadores polimórficos de DNA pericentroméricos de crianças afetadas e seus pais. O estimador de máxima verossimilhança obtido neste trabalho, tem a vantagem de poder ser obtido sem a necessidade da informação dos pais, fazendo deste um método simples e elegante para determinar a origem da não-disjunção meiótica. Isto abre a possibilidade de usar material de arquivo para comparar estudos epidemiológicos da síndrome de Down e outras aneuploidias em várias populações e diferentes países.

## Anexo

### Geração da Distribuição Multinomial

Para gerar a distribuição multinomial foi utilizado o seguinte algoritmo:

Seja  $X \sim \text{mult}(N, \theta_1, \dots, \theta_m)$

Repita

Gera  $u \sim U(0,1)$

Se  $u \leq \theta_1$  então  $k_1 = k_1 + 1$

Se  $\theta_1 < u \leq \theta_1 + \theta_2$  então  $k_2 = k_2 + 1$

Se  $\theta_1 + \theta_2 < u \leq \theta_1 + \theta_2 + \theta_3$  então  $k_3 = k_3 + 1$

.

$\theta_1 + \dots + \theta_{m-2} < u \leq \theta_1 + \dots + \theta_{m-1}$  então  $k_{m-1} = k_{m-1} + 1$

$k_m = k_m + 1$

tantas vezes quanto  $N$ .

No caso, sob  $H_0$ ,  $X =$  número de picos tem distribuição multinomial( $N; \theta_1, \theta_2, \theta_3$ ) onde os  $\theta$ 's são calculados como em (3.1), com  $F = F_0$ .

## REFERÊNCIAS BIBLIOGRÁFICAS

Blachman, N.R., (1992) *Mathematica: A Practical Approach*. New Jersey: Prentice-Hall.

Davison, A.C., Hinkley, D.V., (1997). *Bootstrap Methods and their Application*. New York: Cambridge University Press.

Efron, B., Tibshirani, R.J., (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Gardener, E e Swstad, D. P. (1986) *Genética*. 4ª edição, Rio de Janeiro: Editora Guanabara.

Guo, S., Thompson, E. (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics*, 48:361-372.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) *Introduction to the Theory of Statistics*. Third Edition, Singapore: McGraw-Hill.

Parra, F.C. (1999) *Estudo da Origem de Não-disjunção Meiótica em Indivíduos Brasileiros com Trissomia do Cromossomo 21 Utilizando um Estimador de Máxima Verossimilhança*. Belo Horizonte: Instituto de Ciências Biológicas, UFMG. 100 p . (Dissertação de Mestrado).

Pena, S.D.J. (1998) Molecular Cytogenetics I: PCR-based diagnosis of human trisomies using computer-assisted laser densitometry. *Genet Mol Biol.* **3**: 317-322.

Petersen MB *et al.* (1992) Comparative study of microsatellite and cytogenetic markers for detecting the origin of the nondisjoined chromosome 21 in Down syndrome. *Am J Hum Genet* **51**: 516-525.

Shen JJ, Sherman SL, Hassold TJ. Centromeric genotyping and direct analysis of nondisjunction in humans: Down syndrome. *Chromosoma* 1998; **107**: 166-172.

Welsh, A.H., (1996) *Aspects of Statistical Inference*. New York : John Wiley.