

A utilização da Estatística na Ortodontia

Tien Li AN*, Osmar Aparecido CUOGHI**

Resumo

A estatística exerce um papel fundamental no método científico, a qual se preocupa em organizar, descrever, analisar e interpretar os dados obtidos a partir de uma observação ou de um experimento. Todavia, ela continua sendo pouco acessível na interpretação, seja por parte do leitor ou por parte daquele que a utiliza. Às vezes, a estatística é mal aplicada ou não compreendida devido aos termos, com significado próprio. Além disso, existem poucos artigos científicos da área de Ortodontia relacionados com este assunto. Assim, o presente artigo propõe realizar um levantamento sobre as frequências do uso da estatística, fazer considerações sobre a seleção dos seus métodos e explicar sobre as possíveis interpretações dos resultados.

Palavras-chave: Estatística. Ortodontia.

INTRODUÇÃO

Um conhecimento científico deve ser baseado em fundamentos comprovados, suplantando as qualidades de um julgamento arbitrário, da superstição e das tendências individuais. A comprovação, por sua vez, deve ser obtida sob a regência de um método minuciosamente elaborado, consistindo as seguintes etapas: 1) elaboração de uma hipótese; 2) consideração dos possíveis resultados atuantes sobre a hipótese escolhida; e, finalmente, 3) a avaliação da hipótese para a verificação da idéia original². Assim, uma idéia que parte do empirismo se torna um conhecimento; mesmo que a suposição original não corrobore com a hipótese, seja apenas casual.

A idade contemporânea está vivenciando uma grande afluência de tecnologias, de experiências e de idéias novas ou reintroduzidas, seja nas áreas tecnológicas ou biológicas. Mais do que nunca, um

profissional deve portar a habilidade para discernir um fato científico de uma opinião pessoal ou individual. Nas áreas de saúde, a estatística é uma ferramenta indispensável para auxiliar o indivíduo neste discernimento, a qual se preocupa com a organização, descrição, análise e interpretação dos dados experimentais. Sobre este aspecto, salienta-se que a elaboração de um trabalho com uma amostragem e um delineamento estatístico adequado pode proporcionar extrapolações práticas, mais seguras do que aquelas explicitadas aleatoriamente e sem amplitude suficientemente segura.

Embora os variados tipos de métodos estatísticos estejam descritos em livros específicos^{1,3,4,5,9,10} eles continuam distantes da nossa realidade devido à dificuldade no seu entendimento e à falta de exemplos clínico-científicos palpáveis que estimulem este entendimento. Por outro lado, poucos artigos voltados para a área de Ortodontia se

* Pós-graduado no Programa de Pós-Graduação em nível de Mestrado na área de Ortodontia - Faculdade de Odontologia de Araçatuba - UNESP.

** Professor Doutor da Disciplina de Ortodontia Preventiva - Departamento de Odontologia Infantil e Social - Faculdade de Odontologia de Araçatuba - UNESP.

preocuparam em descrever a aplicabilidade da estatística. Assim, o profissional freqüentemente fica desprovido de subsídio para interpretar os dados numéricos e/ou não-numéricos durante a leitura de um artigo, atendo-se apenas às conclusões oferecidas pelos autores.

Diante disso, o presente artigo propõe realizar um levantamento sobre a freqüência da utilização dos procedimentos estatísticos em artigos da área ortodôntica, fazer considerações sobre os cuidados que devem ser tomados na escolha dos métodos estatísticos e explicar sobre as possíveis interpretações dos resultados estatísticos, cujas terminologias diferem um pouco em termos de seus significados denotativos.

FREQÜÊNCIA DOS PROCEDIMENTOS ESTADÍSTICOS NOS ARTIGOS CIENTÍFICOS DA ORTODONTIA

Para verificar a freqüência dos tipos de procedimentos estatísticos, realizou-se um levantamento bibliográfico, num total de 123 artigos originais, da periódico *American Journal of Orthodontics Dentofacial Orthopedics* de circulação mensal do ano 2001. Optou-se por este periódico devido ao seu elevado conceito e também por fornecer grandes volumes de artigos científicos, que nos permitem reunir uma quantidade considerável de dados. As publicações verificadas consistem de artigos originais onde em cada um pode constar mais de um tipo de procedimento estatístico. Cada tipo de método estatístico é considerado e citado uma única vez mesmo que seja aplicado repetidamente com finalidades diferentes. Os resultados do levantamento se encontram nas tabelas 1 e 2.

Observou-se, por meio do levantamento, que 106 artigos, isto é 86,18%, tiveram necessidade do uso da estatística e somente 17 artigos (13,82%) não mostraram a participação da estatística (Tab. 1). Destes artigos, existem 250 citações de alguns tipos de procedimentos estatísticos e 5 citações do uso do modelo de elemento finito (Tab. 2).

A partir dos procedimentos estatísticos verifi-

cados, pode-se agrupá-los nas seguintes categorias: 1) Testes de médias e de variâncias (34,11%); 2) Testes de associação (16,86%); 3) Testes post hoc (11,75%); 4) Testes de confiança (11,35%); 5) Testes ordinais (10,58%); e, 6) Testes de proporções (3,92%). Os métodos que não apresentaram as características das categorias citadas foram considerados em 7) Outras categorias (11,35%).

Os testes de médias e de variâncias foram os mais utilizados (Tab. 3). Estes testes visam detectar se existem diferenças entre os grupos experimentais e controle. Desta categoria, os testes t foram os mais utilizados (59,77%), dos quais incluem teste t para uma amostra, teste t independente e teste t emparelhado. Outro teste de ampla utilização foi a análise de variância, a qual constituiu 39,08% dos métodos estatísticos encontrados.

Os testes de associação tiveram segunda maior freqüência, nos quais os autores tentam verificar os graus de associação e de previsibilidade entre as variáveis estudadas (Tab. 4). Os testes de regressão (48,84%), correlação de Pearson (41,86%) e de Spearman (9,30%) constituem esta categoria.

Os testes post hoc apresentaram terceira maior freqüência. Os testes desta categoria são aplicados logo após a detecção da existência de diferença entre os grupos pela análise de variância, indicando exatamente qual grupo apresenta diferença em relação aos outros grupos estudados. Entre os tipos de testes desta categoria, o teste de Tukey foi o mais aplicado, constituindo 53,33% dos testes desta categoria (Tab. 5).

A categoria dos testes de confiança e de erros foi a quarta categoria em termos de freqüência de utilização. Os procedimentos estatísticos desta categoria visam verificar o grau de confiabilidade das

Tabela 1 - Levantamento do uso da estatística nos 123 artigos revisados da revista *American Journal of Orthodontics Dentofacial Orthopedics* do ano 2001.

Aplicação de algum procedimento estatístico	106 (86,18%)
Sem estatística	17 (13,82%)

Tabela 2 - Frequência dos tipos de procedimentos estatísticos encontrados nos 106 artigos da revista American Journal of Orthodontics Dentofacial Orthopedics do ano 2001.

Tipos de testes realizados	Frequências	Tipos de testes realizados	Frequências
Teste t	52 (20,39%)	Correlação Intraclasse	2 (0,78%)
Análise de variância	34 (13,44%)	Modelos estatísticos multi-níveis	2 (0,78%)
Erro do método (Dahlberg)	16 (6,27%)	Friedman	1 (0,39%)
Teste de correlação	22 (8,63%)	Weibull	1 (0,39%)
Teste de regressão	21 (8,23%)	Fatorial	1 (0,39%)
Teste de Tukey	16 (6,27%)	Erro padrão	1 (0,39%)
Wilcoxon	11 (4,31%)	Cox-Mantel	1 (0,39%)
Teste de normalidade	10 (3,92%)	Teste dos sinais	1 (0,39%)
Kruskal-Wallis	8 (3,14%)	Duncan	1 (0,39%)
Mann-Whitney	7 (2,74%)	Kaplan-Meier	1 (0,39%)
Teste Qui-Quadrado	6 (2,35%)	Tabelas de contingência	1 (0,39%)
Bonferroni	6 (2,35%)	Fleiss-Cohen	1 (0,39%)
Erro do método (não especificado)	5 (1,96%)	Akritis	1 (0,39%)
Teste de Scheffé	4 (1,57%)	MANOVA	1 (0,39%)
Student-Newman-Keuls	3 (1,17%)	ANCOVA	1 (0,39%)
Coefficiente de confiança	3 (1,17%)	Porcentagem	5 (1,96%)
Teste de Fisher	2 (0,78%)	Elemento finito	5 (1,96%)
Kappa	2 (0,78%)		
		Total	255 (100%)

Tabela 3 - Frequências dos tipos de testes de médias e de variância.

Teste t	52 (59,77%)
Análise de variância	34 (39,08%)
ANCOVA	1 (1,15%)
Total	87 (100%)

Tabela 4 - Frequências dos tipos de testes de associação.

Regressão	21 (48,84%)
Correlação de Pearson	18 (41,86%)
Coefficiente de Spearman	4 (9,30%)
Total	43 (100%)

Tabela 5 - Frequências dos tipos de testes post hoc.

Tukey	16 (53,33%)
Bonferroni	6 (20,00%)
Scheffé	4 (13,33%)
Newman-Keuls	3 (10,00%)
Duncan	1 (3,34%)
Total	30 (100%)

Tabela 6 - Frequências dos tipos de testes de confiança e de erros.

Dahlberg	16 (55,17%)
Não-especificado	5 (17,25%)
Coefficiente de confiança	3 (10,34%)
Kappa	3 (10,34%)
Correlação intraclasse	2 (6,90%)
Total	29 (100%)

Tabela 7 - Frequências dos tipos de testes não-paramétricos para dados ordinais.

Wilcoxon	11 (40,74%)
Kruskal-Wallis	8 (29,63%)
Mann-Whitney - U	7 (25,93%)
Friedman	1 (3,70%)
Total	27 (100%)

Tabela 8 - Frequências dos tipos de testes não-paramétricos de proporção.

Qui-quadrado	6 (60,00%)
Fisher	2 (20,00%)
Teste dos sinais	1 (10,00%)
Mantel	1 (10,00%)
Total	10 (100%)

Tabela 9 - Outras categorias.

Teste de normalidade	10 (34,47%)
Porcentagem	5 (17,24%)
Elemento finito	5 (17,24%)
Modelos estatísticos multi-níveis	2 (6,90%)
MANOVA	1 (3,45%)
Akritas	1 (3,45%)
Weibull	1 (3,45%)
Erro padrão	1 (3,45%)
Tabelas de contingência	1 (3,45%)
Kaplan-Meier	1 (3,45%)
Fatorial	1 (3,45%)
Total	29 (100%)

variáveis mensuradas. Entre os métodos citados, o método preconizado por Dalhberg foi o mais citado, prevalecendo num total de 55,17% (Tab. 6).

Depois das categorias citadas, os teste não-paramétricos também foram encontrados, constituindo 14,50% de todos os métodos utilizados. Para a categoria dos testes dados não-paramétricos para dados ordinais, o teste de Wilcoxon (40,37%) foi o mais citado (Tab. 7). Enquanto que na categoria de testes para dados nominais, o mais freqüente foi o teste qui-quadrado, perfazendo 60% (Tab. 8).

Finalmente, outros métodos que não constituíram nas categorias já mencionadas perfazem 11,35% do total (Tab. 9).

É importante observar neste levantamento que os testes de significância mais utilizados (testes paramétricos) poderiam ser aplicados somente quando a distribuição dos dados for normal, ou seja, apresentem distribuição em forma de uma curva de Gauss. Por outro lado, o uso do teste de normalidade é pouco freqüente (10 citações) quando comparado com a ampla aplicação dos testes paramétricos (87 cita-

ções). Neste sentido, muitos artigos supõem que os dados sejam normais sem, entretanto, verificá-los, o que pode comprometer a veracidade dos resultados principalmente quando a amostra for pequena.

Para descrever melhor o conceito de normalidade ou de distribuição Gaussiana, o seguinte exemplo pode ser citado: sabe-se que o valor normativo do ângulo SNA é de 82° em adultos, porém, este é um valor médio, podendo encontrar valores menores ou maiores. Supondo-se que exista uma população com infinitos indivíduos, ao medir-se os valores do ângulo SNA de cada indivíduo, teríamos uma grande parte da população com o valor apresentando próximo da média, uma parte, em proporção menor, com valores um pouco afastados da média e teríamos também uma pequena parte com os valores deste ângulo extremamente afastados da média, seja para mais ou para menos. Assim, uma vez representados em gráfico os dados de cada indivíduo, a distribuição de freqüência teria a forma de um sino^{4,5,8}. Com o teste de normalidade tentar-se-ia verificar a distribuição dos dados estudados se estão de acordo com as características dessa curva de normalidade (Gráf. 1), em outras palavras, o teste verifica quão desviado está a distribuição da amostra selecionada em relação a uma curva normal⁸.

Os testes de confiança e de erros também deveriam apresentar uma freqüência maior daquela constatada (Tab. 6), uma vez que a maior parte dos dados trabalhados era dados contínuos, mensuráveis e, portanto, sujeitos a erros. Somente por meio dos testes de confiança e de erros que se obtém subsídios para verificar a reprodutibilidade dos métodos e a confiabilidade das conclusões estabelecidas.

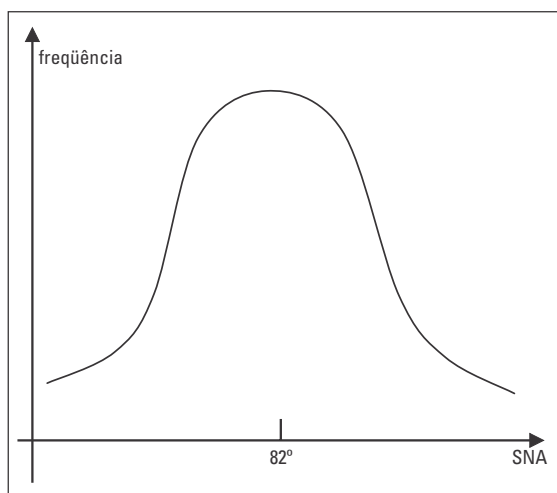


GRÁFICO 1 - Curva normal, também denominada de curva de Gauss, em forma de um sino.

De qualquer forma, com uma análise de 123 artigos, foi possível demonstrar a importância da estatística no meio científico da área ortodôntica e, indiretamente, ressaltar a relevância no entendimento deste assunto para a interpretação dos resultados disponíveis na literatura.

USO DOS MÉTODOS ESTATÍSTICOS

Existem muitos livros específicos que explicam detalhadamente sobre a estatística, portanto, não seria interessante repetir estas descrições, entretanto, vale a pena discutir sobre o uso de alguns testes frequentemente citados, servindo de orientação para que o seu emprego seja o mais acessível e adequado possível^{1,3,4,5,9,10}.

De acordo com os tipos de procedimentos estatísticos encontrados no levantamento, a maioria pode ser separada em três grandes conjuntos: 1) os testes de significância, utilizados para comparar as amostras e verificar as suas diferenças, os quais incluem testes de médias e de variância, os testes post hoc, os testes ordinais e os testes de proporções; 2) os testes de associação, para a verificação da correlação ou da previsibilidade; e 3) os testes de confiança, para determinar o nível de erro das mensurações. Os testes que não apresentaram os perfis citados foram incluídos em outras categorias.

Testes de significância

Para efetuar os testes de significância, a dificuldade reside na escolha dos tipos de testes, que podem ser subdivididos em dois grandes grupos: os paramétricos e os não-paramétricos⁴.

Os testes denominados paramétricos são aqueles aplicados para a comparação de variáveis contínuas, que apresentam variâncias iguais e distribuição normal ou próxima da normalidade também conhecida como distribuição Gaussiana, ou seja, variáveis quantitativas cujas distribuições, quando representadas em gráfico, apresentam a forma de um “sino”^{4,5,8}. Na utilização destes testes, pode ser necessário optar entre teste unicaudal ou bicaudal. O teste unicaudal é apropriado quando a direção de uma diferença pode ser pré-determinada ou quando se pode afirmar o grupo com maior média⁸. Em situações onde não é possível especificar a direção da hipótese, isto é, a diferença pode variar para mais ou para menos, um teste bicaudal deve ser escolhido⁸. De fato, a segunda situação é mais comum de ocorrer, portanto, é razoável sempre que possível, optar pelo teste bicaudal.

É importante notar que, nem sempre as variáveis quantitativas que apresentam uma distribuição normal numa população implicam na distribuição normal da amostra selecionada. Isto pode ocorrer em circunstâncias de amostras pequenas^{5,8}. Assim, pode-se considerar como um procedimento preliminar, a execução do teste de normalidade para definir a amostra segundo sua distribuição de frequências.

Poucos artigos científicos revisados relataram o emprego do teste de normalidade, mas aplicaram diretamente testes paramétricos para a comparação das variáveis. Isto pode perder a veracidade quando a amostra é pequena. Às vezes, duas amostras com variáveis contínuas que não apresentam distribuição normal podem ser submetidas a um teste paramétrico desde que suas variáveis sejam transformadas^{5,8}. A transformação consiste em realizar uma aproximação das variáveis à curva de normalidade⁸. O procedimento de transformação pode ser relevante, visto que,

comparativamente, os testes paramétricos são mais eficientes para encontrar diferenças, isto é, eles apresentam maior poder de análise⁸. Portanto, sempre que possível, deve-se realizar testes paramétricos para variáveis quantitativas. Os testes paramétricos mais utilizados são o teste t e a análise de variância.

Os tipos de testes não-paramétricos são utilizados quando as variáveis das amostras são nominais ou ordinais, onde geralmente aparecem em classes ou escores em vez de valores contínuos^{4,5,8,9,10}. Às vezes, as variáveis contínuas coletadas podem estar em pequena quantidade desprovidas de uma distribuição normal e impossibilitadas de serem transformadas, nestas situações, os testes não-paramétricos são melhores aplicados do que os paramétricos^{4,8}.

Diferentemente dos testes paramétricos, onde a amostra testada constitui um parâmetro da população da qual extraiu, as variáveis utilizadas nos testes não-paramétricos estão diretamente relacionadas à amostra e indiretamente à população^{2,4}. Assim, o que realmente está sendo testado seria apenas a amostra e as conclusões estabelecidas se restringem à mesma. Os testes não-paramétricos comumente observados são os testes de Wilcoxon, Mann-Whitney, Kruskal-Wallis e Qui-quadrado. É interessante salientar que existe certa equivalência entre os tipos de testes, isto é, para um determinado teste paramétrico, existe a sua equivalência no teste não-paramétrico. Isto pode facilitar bastante na escolha do tipo de teste a ser aplicado^{2,5}.

Uma vez determinadas as características das variáveis, deve verificar quantos grupos serão comparados e se existe emparelhamento entre as variáveis⁵. O conceito de emparelhamento, erroneamente referido como pareamento, consiste em relacionar os dados da amostra dois a dois segundo algum critério que introduz uma influência marcante entre os diversos pares, que supomos, influir igualmente sobre os valores de cada par⁵. O emparelhamento pode ser natural, como ocorre entre gêmeos idênticos; artificial quando se procura formar pares semelhantes em relação a uma série de

características, como por exemplo, mesma idade, mesmo gênero, mesma altura e peso⁵. Ainda pode ser auto-emparelhamento, onde cada elemento atua como seu próprio controle, freqüentemente utilizado quando o estudo é longitudinal⁵.

Resumidamente, na escolha de um teste de significância, as seguintes etapas podem ser seguidas: 1) Determinar os tipos de dados disponíveis, que podem ser nominais, ordinais ou contínuos, isto é, determinar se os dados são paramétricos ou não-paramétricos; 2) Determinar quantos grupos serão comparados, que podem ser um, dois ou mais de dois grupos; e 3) Deve verificar se existe emparelhamento entre os dados⁵. O fluxograma apresentado no quadro 1 esquematiza a escolha de um teste de significância.

Testes de associação

Além dos testes de significância, os testes de associação também são freqüentemente utilizados, os quais incluem os testes de correlação e de regressão. Os testes de correlação visam estabelecer uma relação linear entre duas variáveis e avaliar a sua intensidade^{2,4,5,8,9,10}. O coeficiente de correlação pode variar entre -1 e 1 . A correlação é mais forte quanto mais o coeficiente se aproxima do 1 ou do -1 ^{2,5}. Se o coeficiente for zero, não existe uma correlação linear entre as variáveis. Isto não implica que as duas variáveis não estão relacionadas, mas apenas que elas não apresentam uma relação linear^{2,5}. Quando o coeficiente apresenta sinal positivo, significa que as variáveis variam em conjunto no mesmo sentido, isto é, se uma aumenta a outra também aumenta e se uma diminui, a outra também diminui^{4,5,8,9}. Por outro lado, se o sinal for negativo, a relação é inversa, ou seja, as variáveis variam em sentidos opostos^{4,5,8,9}.

O teste de correlação não faz julgamento sobre se uma variável é causa ou consequência da outra, apenas estuda quão consistentemente duas variáveis mudam em conjunto. Muitos fatos estão associados e não têm uma relação de causa e efeito⁵.

Analogamente aos testes de significância, para

a realização dos testes de correlação, deve-se iniciar pela verificação da distribuição das amostras. O teste de correlação de Pearson é destinado para dados onde ocorre distribuição normal, enquanto que o de Spearman é destinado para dados sem distribuição normal⁸.

Os testes de regressão visam estabelecer uma função matemática para descrever a relação entre as variáveis estudadas⁸, sendo freqüentemente utilizados quando se deseja verificar a previsibilidade do comportamento de uma determinada variável em relação à outra. Geralmente o teste de regressão é aplicado quando existe uma correlação significativa entre as variáveis definidas pelo coeficiente de determinação que consiste do coeficiente de correlação elevado ao quadrado (r^2). Este coeficiente de determinação deve ser o mais próximo possível do 1^{4,5,8}.

Tanto o teste de correlação como o de regressão não devem ser aplicados para dados nominais em forma de escores ou escalas, pois estes números são arbitrários e não apresentam um significado numérico real. Se for realmente de interesse a associação entre duas variáveis nominais, o teste Qui-quadrado pode ser considerado².

Para um melhor entendimento, pode-se exemplificar o emprego dos testes estatísticos citando um artigo extraído do levantamento bibliográfico. Num estudo efetuado por Kim e Muhl⁷, estes autores propuseram: 1) avaliar como o AEB cervical afeta na direção do crescimento mandibular durante e após o tratamento quando comparado com o grupo não-tratado; 2) analisar a previsibilidade da rotação mandibular durante o tratamento a partir de várias medidas de padrão esquelético; 3) veri-

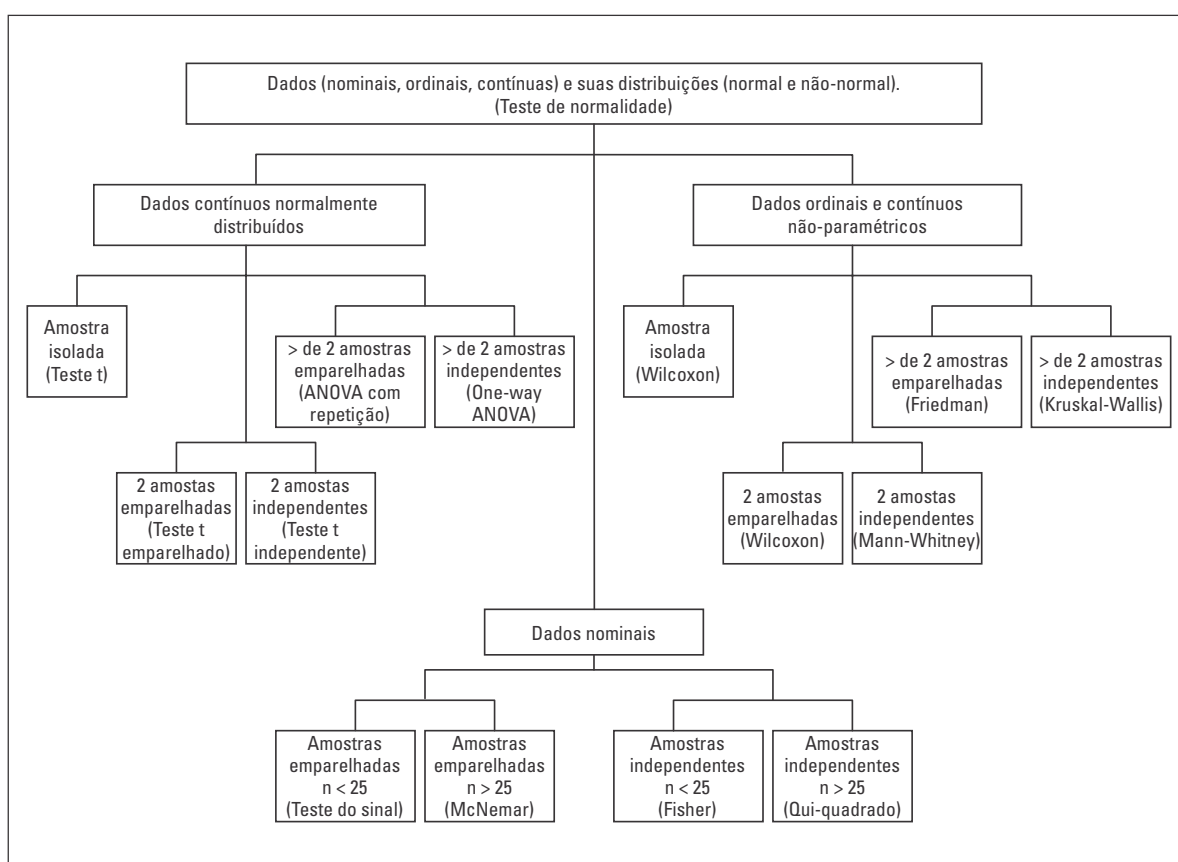


FIGURA 2 - Fluxograma esquematizando os passos na seleção de um determinado teste estatístico.

ficar a relação entre a alteração do tratamento e após o tratamento para cada variável; e 4) avaliar a relação entre as alterações das variáveis durante o tratamento e no período de acompanhamento.

Para o estudo, Kim e Muhl⁷ utilizaram telerradiografias tomadas a partir de um grupo de 30 pacientes tratados e um grupo controle de 26 pacientes não-tratados. As radiografias foram tomadas antes e no final do tratamento, bem como dois anos após a remoção da contenção. As variáveis utilizadas para o estudo foram várias grandezas cefalométricas, as quais foram comparadas entre o grupo tratado e o não-tratado durante os três períodos estudados.

Pela primeira proposição, os autores tentaram verificar a significância entre o grupo tratado e o controle, cuja hipótese nula foi estabelecida como se o tratamento não afetasse a direção do crescimento.

Como primeiro passo na escolha dos tipos de testes, deve-se perguntar qual tipo de dado está sendo utilizado? No estudo referido, os dados são as medidas cefalométricas, portanto, trata-se de dados contínuos. Desde que os dados contínuos apresentem distribuição normal, testes paramétricos podem ser utilizados. Embora não esteja constatado no artigo, seria interessante aplicar o teste de normalidade para prosseguir a escolha dos tipos de testes. Talvez seja pela quantidade razoável de amostras ($n = 30$) foi suposto que os dados estavam normalmente distribuídos.

A segunda pergunta a ser levantada é quantos grupos serão comparados? No caso em questão, dois grupos foram comparados, escolhendo assim, o teste t para verificar a significância. A terceira pergunta a ser respondida é existe emparelhamento entre os dados? A resposta é não. Pois os grupos comparados são independentes entre si. Embora o estudo seja longitudinal, as comparações foram feitas entre dois grupos distintos, um independe do outro. Por outro lado, se utilizarmos o mesmo grupo tratado para uma avaliação longitudinal sem a participação do controle, isto é, o grupo estudado serviria como controle dele mesmo, avaliando-se antes, durante e após o tratamento, neste caso existiria emparelha-

mento de acordo com a definição já introduzida.

Assim, seguindo o fluxograma, o teste de significância a ser aplicado é o teste t para dados independentes, o qual foi escolhido pelos autores (Tab. 10). Aqui, pela tabela 10, pode-se verificar que os autores compararam as alterações ocorridas durante os três períodos de estudos (T1, T2 e T3) entre o grupo tratado e o controle, isto é, as diferenças das alterações das grandezas citadas ($T2 - T1$, $T3 - T2$ e $T3 - T1$). Pode-se verificar que foram as diferenças entre os três períodos que estavam sendo comparadas: as alterações entre pré-tratamento e final do tratamento ($T2 - T1$); as alterações entre final do tratamento e a fase sem a contenção ($T3 - T2$) (Tab. 11); e as alterações entre fase sem a contenção e o período pré-tratamento ($T3 - T1$) (Tab. 12). A comparação de várias médias duas a duas usando o teste t não seria satisfatória. Isso se deve ao fato de que o nível de significância seria desvirtuado, pois, quanto maior o número de comparações feitas, maior a probabilidade de se obterem rejeições de hipótese por mera casualidade⁴. No exemplo, o teste mais apropriado seria a análise de variância em associação com algum tipo de teste post hoc, onde o primeiro detectaria a presença de diferença significativa e o segundo determinaria onde estaria exatamente a diferença.

Para as outras proposições, são necessários os testes de associação. Na segunda proposição, com a finalidade de verificar a previsibilidade da rotação mandibular em relação às outras variáveis, foi escolhido o teste de regressão. O teste de regressão determinaria uma equação matemática com a qual seria possível prever, neste caso, a rotação da mandíbula em função de uma outra variável (Tab. 13). Aqui, as duas variáveis que os autores escolheram foram o ângulo do plano oclusal e o ângulo do plano palatino, ou seja, quantos graus de rotação do plano mandibular alterariam se essas duas variáveis se alterassem numa determinada quantidade?

Para a terceira e a quarta proposição, testes de correlação de Pearson foram utilizados considerando que os dados estavam normalmente distribuídos. Os testes de correlação foram utilizados na tentati-

va de verificar a relação existente entre determinadas variáveis, isto é, verificar o comportamento de uma variável em função da outra (Tab 14).

Além dos testes citados, os autores também aplicaram o teste t emparelhado para estudar a reprodutibilidade do método. Para isso, repetiram os traçados e suas respectivas mensurações em dez telerradiografias tomadas antes e dez tomadas após o tratamento num intervalo de duas semanas. Depois, encontraram as diferenças entre as grandezas cefalométricas antes e após o tratamento tanto para o primeiro como para o segundo traçado e mensuração. Posteriormente, essas diferenças foram comparadas pelo teste t emparelhado.

De acordo com Houston⁶, o uso do teste t serve para estudar o erro sistemático, também conhecido como o erro causado pelas tendenciosidades. Para tornar um trabalho ainda mais confiável,

seria interessante verificar também o erro aleatório, principalmente quando o estudo é cefalométrico, o que corresponde às diferenças cometidas quando na identificação das estruturas e nas mensurações das grandezas cefalométricas.

INTERPRETAÇÃO DOS RESULTADOS ESTADÍSTICOS

O item anterior objetivou servir de orientação para o procedimento das análises estatísticas. Porém, tão importante quanto saber aplicar os métodos estatísticos é interpretar os dados e talvez seja a parte mais interessante para qualquer leitor.

A estatística pode ser dividida em descritiva e analítica durante o processamento dos resultados^{4,5,9}. A descritiva consiste em apenas descrever os resultados, ordenando-os em tabelas, gráficos, quadros e outros para um melhor entendimento.

Tabela 10 - Comparação entre os períodos T1 e T2 de algumas grandezas angulares cefalométricas utilizadas no artigo de Kim e Muhl⁷ por meio do teste *t*.

	T2-T1				P
	Grupo tratado		Controle		
	Média	Desvio-padrão	Média	Desvio-padrão	
Rotação mandibular	0,25	3,0	-1,7	2,4	0,94
Rotação maxilar	2,6	1,6	-0,7	1,3	0,00
Ângulo facial	0,5	1,8	1,4	1,4	0,04
ANB	-2,4	1,9	-0,5	1,0	0,00
Ângulo do plano oclusal	-2,0	3,2	-3,4	4,6	0,16
Ângulo do plano palatino	2,2	1,5	-0,7	1,9	0,00

Tabela 11 - Comparação entre os períodos T2 e T3 de algumas grandezas angulares cefalométricas utilizadas no artigo de Kim e Muhl⁷ por meio do teste *t*.

	T3-T2				P
	Grupo tratado		Controle		
	Média	Desvio-padrão	Média	Desvio-padrão	
Rotação mandibular	-1,5	1,3	-0,7	1,2	0,03
Rotação maxilar	-0,7	1,2	-0,2	0,9	0,11
Ângulo facial	0,5	1,0	0,5	0,7	0,94
ANB	-0,3	1,0	-0,1	0,8	0,45
Ângulo do plano oclusal	-1,7	2,7	-2,0	2,6	0,64
Ângulo do plano palatino	-0,5	1,0	0,5	1,2	0,00

Tabela 12 - Comparação entre os períodos T3 e T1 de algumas grandezas angulares cefalométricas utilizadas no artigo de Kim e Muhl⁷ por meio do teste *t*.

	T3-T1				P
	Grupo tratado		Controle		
	Média	Desvio-padrão	Média	Desvio-padrão	
Rotação mandibular	-1,3	3,5	-2,9	3,1	0,10
Rotação maxilar	1,8	2,1	-1,3	1,9	0,00
Ângulo facial	1,0	2,3	2,2	1,9	0,07
ANB	-2,7	2,1	-0,8	1,0	0,00
Ângulo do plano oclusal	-3,6	2,0	-5,1	3,7	0,15
Ângulo do plano palatino	1,7	1,7	-0,5	1,4	0,00

Tabela 13 - Resultado da análise de regressão para a rotação mandibular durante o tratamento a partir dos valores das variáveis independentes na fase T1⁷.

Seqüência	Variável	r ²	r ² cumulativo	F	P
1	Ângulo do plano oclusal (APO)	0,2286	0,2286	5,88	0,0076
2	Ângulo do plano palatino (APP)	0,0750	0,3035		

Fórmula para prever a rotação mandibular durante o tratamento:

$$\text{Rotação mandibular}_{T1-T2} = -2,72 + 0,37 \cdot \text{APO}_{T1} + 0,27 \cdot \text{APP}_{T1}$$

Tabela 14 - Coeficiente de correlação (r) para verificar as relações das alterações de determinadas variáveis durante a fase de tratamento no grupo tratado⁷.

Variáveis	r	P
Rotação mandibular e distância horizontal do \underline{G}	-0,67	0,0001
Nperp-A e ângulo do plano mandibular	-0,77	0,0001
Nperp-A e rotação mandibular	-0,73	0,0001
Rotação maxilar e rotação mandibular	0,39	0,0339
Rotação maxilar e distância horizontal do \underline{G}	-0,44	0,0154
Eixo Y e ângulo do plano mandibular	0,72	0,0001

A analítica ou indutiva se responsabiliza pela parte da interpretação e inferência dos resultados, isto é, efetuar uma extrapolação para estabelecer conclusões mais objetivas possíveis^{4,5}. Esta segunda parte geralmente leva a certa confusão, sendo, portanto, o foco da explanação.

A maioria das análises aplicadas visa comparações entre as amostras e, conseqüentemente, responder às seguintes questões: nas amostras estudadas, existem diferenças entre elas? Essas diferenças encontradas são verdadeiras ou são apenas casuais devido à variabilidade da amostra? Diante

disso, duas hipóteses podem ser aventadas: aquela denominada de hipótese experimental supõe que existe uma diferença entre as amostras, enquanto que a hipótese denominada nula supõe que não existe diferença entre as amostras^{4,5,8,9}.

Às vezes, a diferença não existe de fato, ou seja, a hipótese nula é verdadeira. Por outro lado, a diferença pode realmente existir, isto é, a hipótese nula é falsa. No entanto, dependendo da condução do experimento e da sua análise, dois tipos de erros podem ser cometidos: o erro tipo I ocorre quando a hipótese nula é verdadeira, porém, pela análise,

a mesma é rejeitada; em outras palavras, o experimento demonstra uma diferença entre as amostras a qual não existiria na realidade; enquanto que o erro tipo II ocorre quando a hipótese nula é falsa, porém, a análise foi aceita, ou seja, realmente existe a diferença embora não seja detectada^{4,5,8,9}.

Assim, toda vez que a decisão é tomada para aceitar ou rejeitar alguma hipótese, implica certa probabilidade de cometer erros. A probabilidade de cometer o erro tipo I é descrita pela letra " α "^{4,5}, sendo conhecida como o nível de significância. A probabilidade de cometer erro tipo II é descrita pela letra " β " e está relacionada ao poder de estudo, isto é, a capacidade do estudo em detectar diferenças⁵. É razoável sempre determinar quanto de erro é admissível: para o erro tipo I, considera-se geralmente que até 5% é aceitável e para o erro tipo II, comumente se aceita de 10% a 20%⁵.

Em termos de nível de significância, comumente descrito na parte de resultado como diferença estatisticamente significativa ao nível de 5%, 1%, 0,1% e assim por diante, implica que ao decidir em rejeitar a hipótese nula (amostras iguais) existe uma probabilidade de 5%, 1% ou 0,1% de estar rejeitando a hipótese nula mesmo que a mesma seja verdadeira. Logicamente, quanto menor for esta probabilidade, a decisão de rejeitar a hipótese nula estaria bem mais fundamentada^{5,8}. Vale salientar que quando a diferença não é estatisticamente significativa, pode-se concluir somente que não existe evidência suficiente para rejeitar a hipótese nula, em vez de concluir que a hipótese nula seja verdadeira, ou seja, não existe diferença entre as amostras⁸.

Por outro lado, o termo significativo às vezes pode ser mal interpretado. Uma significância estatística não implica significância clínica. Uma diferença estatisticamente significativa não é sinônima de clinicamente significativa e tão pouco implica que seja biologicamente ou clinicamente importante. Por outro lado, um resultado não significativo pode se tornar muito importante pelo seu contexto⁸.

Às vezes, o valor de " α " é bem pequeno, por

exemplo 0,01%, intuitivamente, este valor é mais significativo do que 4%, porém, no seu sentido estrito, isto não é uma interpretação correta. Uma vez que o nível de significância é estabelecido, qualquer resultado somente pode ser significativo ou não-significativo, pois o grau de significância estatística não é distinto⁸. Este ponto de vista é bastante destacado pelos estatísticos, embora os cientistas não sejam tão rígidos quanto a isso, e têm referido seus resultados como sendo "pouco significativo", "muito significativo" ou "extremamente significativo"⁸.

Para a interpretação dos dados estatísticos oferecidos pelo artigo citado⁷, voltemos às tabelas 10, 11 e 12. Pode-se considerar estatisticamente significativo quando as comparações entre o grupo tratado e o controle apresentar o valor de p menor do que 0,05 (5% de significância). No exemplo da tabela 11, a diferença do valor do ângulo do plano palatino foi estatisticamente significativa entre o grupo tratado e o controle. Porém, se considerarmos essa diferença em valor clínico, em torno de 1°, a diferença talvez não seja significativa. Assim, vale enfatizar que nem sempre a diferença estatisticamente significativa implica na diferença clinicamente significativa.

Em termos da análise de regressão (Tab. 13), deve-se verificar primeiramente o valor do r^2 , o qual pode variar de 0 a 1 (0% a 100%). No exemplo, o valor foi de 0,2286 e 0,0750, sendo considerado muito pequeno para poder fazer uma inferência. O valor de 0,2286 e 0,075 significa que apenas 22,86% e 7,5%, respectivamente, dos casos podem ser previstos pela fórmula estabelecida pela análise para as variáveis citadas em relação à rotação mandibular, o que é muito baixa. Assim como o coeficiente de correlação (r), os valores de r^2 pequenos não implicam que as variáveis não apresentem relação entre si, apenas indicam que as suas alterações não variam de forma consistentes.

Com relação aos coeficientes de correlação (Tab. 14), pode-se notar que existem valores com sinais negativos e valores com sinais positivos. O valor positivo indica uma variação no mesmo sentido, para mais ou para menos, e o sinal negativo

indica uma variação em sentidos opostos, isto é, se um aumenta o outro diminui e vice-versa. Quanto mais próximos de 1 ou -1, mais perfeita é a correlação linear.

Finalmente, a estatística deve ser encarada como um meio auxiliar de grande valia para discriminar as tendências de um determinado

pesquisador em vez de um “bicho de sete cabeças”. Além disso, os resultados estatísticos são apenas números, podendo ser significativamente diferentes ou não quando submetido a tratamentos, os seus significados clínicos devem ser mensurados e avaliados em conjunto com a experiência e o bom senso do profissional.

Enviado em: Agosto de 2002
Revisado e aceito: Agosto de 2003

The application of Statistics in Orthodontics

Abstract

Statistics plays a fundamental role in scientific method, which aims to organize, to describe, to analyze and to interpret the data obtained from an observation or an experiment. However, it continues unavailable, in interpretation, by the readers as well as by those who employ it. Sometimes, statistics is mistakenly applied or comprehended due to some terms, which, statistically, present their own meanings. Additionally, there are few articles in Orthodontics that describe this subject. Thus, the present article aims to perform a survey about the frequencies of the use of statistics, to make considerations about the selection of its methods and to explain about the possible interpretations of the data.

Key words: Statistics. Orthodontics.

REFERÊNCIAS

1. ARMITAGE, P.; BERRY, G.; MATTHEWS, J. N. S. **Statistical methods in medical research**. 4th ed. Oxford: Blackwell Science, 2002.
2. BEGOLE, E. A. Statistics for the orthodontist. In: GRABER, T.; VANARSDALL, R. L. **Orthodontics: current principles and techniques**. 3rd ed. St. Louis: Mosby-Year Book, 2000. p. 339-352.
3. BERQUÓ, E. S.; SOUZA, J. M. P.; GOTLIEB, S. L. D. **Bioestatística**. 2. ed. São Paulo: Pedagógica e Universitária, 1981. p. 350.
4. COSTA NETO, P. L. O. **Estatística**. São Paulo: E. Blücher, 1977. p. 264.
5. DORIA FILHO, U. **Introdução à estatística: para simples mortais**. São Paulo: Negócio, 1999. p.152.
6. HOUSTON, W. J. B. The analysis of errors in orthodontic measurements. **Am J Orthod**, St. Louis, v. 83, p. 382-390, 1983.
7. KIM, K. R.; MUHL, Z. F. Changes in mandibular growth direction during and after cervical headgear treatment. **Am J Orthod Dentofacial Orthop**, St. Louis, v. 119, p. 522-530, 2001.
8. MOTULSKY, H. J. **Analyzing data with graphpad prism**. San Diego: Graphpad software, 1999. p. 357.
9. SOUNIS, E. **Bioestatística: princípios fundamentais: metodologia estatística: aplicação às ciências biológicas**. 2. ed. Barueri: McGraw-Hill do Brasil, 1975. p. 230.
10. VIEIRA, S. **Introdução à bioestatística**. 3. ed. Rio de Janeiro: Ed. Campus, 1980. p.196.

Endereço para correspondência

Osmar Aparecido Cuoghi
Disciplina de Ortodontia Preventiva
Faculdade de Odontologia de Araçatuba – UNESP
Rua José Bonifácio, 1193 - CEP: 16015-050 - Araçatuba - SP
e-mail: litien@hotmail.com / osmar@foa.unesp.br