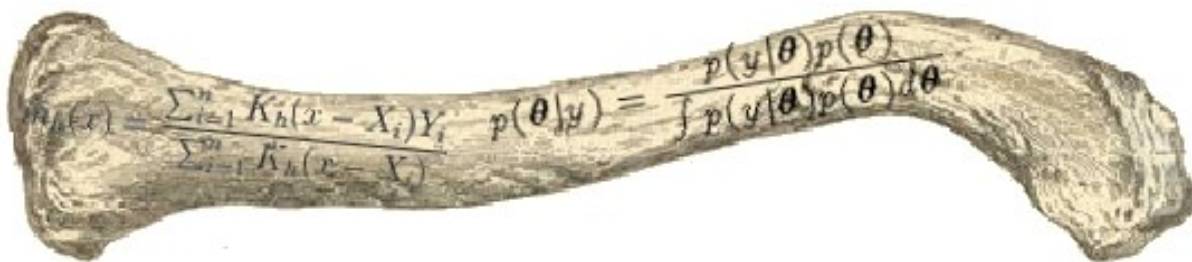




METODOLOGIA DE ESTIMATIVA DA IDADE À MORTE PELA CLAVÍCULA:

Uma abordagem Bayesiana



Ricardo Filipe Mendes Belo Vicente

Tese de Mestrado em Medicina Legal e Ciências Forenses

Metodologia de Estimativa da Idade à Morte pela Clavícula

Uma abordagem Bayesiana

Dissertação apresentada à Faculdade de Medicina da Universidade de Coimbra para obtenção do grau de Mestre em Medicina Legal e Ciências Forenses

Orientação

Professora Doutora Eugénia Cunha (Faculdade de Ciências e Tecnologia da Universidade de Coimbra)

Co-orientação

Professor Doutor Paulo Oliveira (Faculdade de Ciências e Tecnologia da Universidade de Coimbra)
Professor Doutor Duarte Nuno Vieira (Faculdade de Medicina da Universidade de Coimbra)

Ricardo Filipe Mendes Belo Vicente

Setembro de 2012

Agradecimentos

Findo este trabalho é incontornável o agradecimento a várias pessoas que, de forma directa ou indirecta, contribuíram e tornaram possível a sua realização.

À Professora Doutora Eugénia Cunha, pela confiança, por tudo o que me ensinou, pela constante energia que tem e transmite e pela total disponibilidade mesmo quando tempo é algo que quase não existe, tudo isto foi essencial não só para este trabalho mas para toda a minha formação.

Ao Professor Doutor Paulo Oliveira por ter aceite co-orientar este trabalho e por toda a paciência que sempre teve neste processo de “Introdução às Probabilidades e Estatística em tempo *record*”. Pelas sugestões e ensinamentos que em muito contribuíram para o meu enriquecimento académico.

Ao Professor Doutor Duarte Nuno Vieira por ter aceite prontamente co-orientar este trabalho assim como a disponibilidade tida quando algo lhe era pedido.

À Professora Doutora Ana Luísa Santos e ao Doutor Hugo Cardoso pela ajuda e disponibilidade para o estudo das colecções do departamento de Ciências da Vida da Universidade de Coimbra e do Museu Bocage – Lisboa.

Ao David Navega e ao Gonçalo Carnim pelas inúmeras discussões e sugestões que aconteceram ao longo deste percurso académico e que foram essenciais a uma melhor compreensão do que é afinal isto de “fazer” Antropologia Forense e ser Antropólogo Forense (eventualmente).

A todos os amigos, colegas, professores e funcionários do Departamento de Ciências da Vida da FCTUC que no meu percurso académico me ajudaram, ouviram, apoiaram e ensinaram, sem dúvida todos ajudaram de uma forma ou de outra a que pudesse chegar a este ponto.

À Joana por me ter ouvido e aturado em momentos em que parecia que “estava tudo mal” ou que “os resultados não vão nada de jeito”, pela presença e paciência constante.

Um obrigado especial aos meus pais e à minha irmã, nada do que até aqui aconteceu, este trabalho inclusive, tinha sido possível sem a presença, compreensão e amizade constantes.

Índice

Resumo.....	viii
Abstract.....	x
Capítulo 1 – Introdução.....	1
1.1 – História da Estimativa da Idade à Morte.....	4
1.2 – O interesse da Estimativa da Idade à Morte.....	5
1.3 – Metodologias de Estimativa de Idade à Morte.....	7
1.4 – Fontes de Erro na Estimativa da Idade à Morte.....	9
1.5 – Revisão da Literatura: Metodologias de Estimativa da Idade à Morte pela Clavícula.....	12
1.6 – Objectivos.....	19
Capítulo 2 – A Estatística e a Antropologia Forense.....	21
2.1 – A Importância da Estatística nas Ciências Forenses e na Estimativa da Idade à Morte.....	21
2.2 – Evolução Matemática dos Métodos Estatísticos.....	23
2.2.1 – <i>Regressão Linear</i>	25
2.2.2 – <i>Calibração Clássica</i>	31
2.2.3 – <i>Abordagem Bayesiana</i>	33
2.2.3.1 – Estimação pelo método dos núcleos.....	36
Capítulo 3 – Amostra.....	40
Capítulo 4 – Metodologia.....	43
4.1 – Procedimentos de Recolha dos Dados.....	43
4.2 – Análise de Clusters.....	46
4.3 – Medidas de Avaliação dos Erros e Qualidade dos Modelos.....	48
4.4 – Reamostragem por Jackknife.....	52
4.5 – Descrição da Metodologia Estatística Utilizada para a Criação dos Modelos.....	53
4.5.1 – <i>Regressão Linear e Calibração Clássica</i>	53
4.5.2 – <i>Versão discreta do Teorema de Bayes</i>	54
4.5.3 – <i>Modelos criados com o método de Regularização local sem e com distribuição marginal conhecida</i>	55

Capítulo 5 – Resultados	57
5.1 – Modelo de Regressão Linear	58
5.2 – Modelo de Calibração Clássica	60
5.3 – Versão Discreta do Modelo Bayesiano	61
5.4 – Modelo 1 (Regularização local sem distribuição marginal conhecida)	62
5.5 – Modelo 2 (Regularização local com distribuição marginal conhecida)	63
5.6 – Modelo com Amostra Reduzida (Regularização local com distribuição marginal conhecida)	63
5.7 – Resumo e Comparação de todos os Modelos e respectivas Medidas de Erro.....	64
Capítulo 6 – Discussão	65
Capítulo 7 – Considerações finais	69
Capítulo 8 – Aplicação prática dos Modelos	72
Bibliografia	76
Apêndice A	88
Apêndice B	91
Apêndice C	105
Apêndice D	107

Índice de tabelas

Tabela 1.1. Resumo dos dados recolhidos e resultados obtidos por autores de estudos de idade à morte com base na clavícula	15
Tabela 2.2. Distribuição dos indivíduos da amostra pelas Fases de União epifisial e Grupos Etários	35
Tabela 3.1. Distribuição dos indivíduos da amostra por Coleção, Sexo e Idade à Morte ...	41
Tabela 3.2. Distribuição dos indivíduos da amostra por Ano de morte Mínimo, Máximo e Médio e por Coleção	41
Tabela 3.3. Teste t para as diferenças das médias dos Rácios por Coleção	42
Tabela 4.1 Distribuição dos indivíduos da amostra de acordo com a Fase de união epifisial e Grupo Etário criados	47
Tabela 4.2. Distribuição dos indivíduos da amostra reduzida por Fase de união epifisial e Grupo Etário	56
Tabela 5.1. Teste t para as diferenças entre as médias dos Rácios obtidos para ambos os sexos	58
Tabela 5.2. Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o modelo Bayesiano discreto	62
Tabela 5.3. Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 1	62
Tabela 5.4. Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 2	63

Tabela 5.5. Probabilidades <i>a posteriori</i> consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 2 para a amostra reduzida	63
Tabela 5.6. Valores das medidas de avaliação dos modelos	64
Tabela 8.1. Probabilidades <i>a posteriori</i> consoante a Fase de união epifisial e Grupo etário de acordo com o modelo Bayesiano discreto	74
Tabela 8.2. Probabilidades <i>a posteriori</i> consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 1	74
Tabela 8.3. Probabilidades <i>a posteriori</i> consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 2	74

Índice de figuras

Figura 2.1. Regressão quadrática: exemplo de uma relação inaceitável	24
Figura 2.2. Função de regressão com 3 hipotéticos pontos e respectivos erros assinalados ..	26
Figura 2.3. Regressão dos Erros sobre a Idade Real	29
Figura 2.4. Regressão dos Erros sobre a Idade Real	32
Figura 2.5. Exemplo de um gráfico de uma Distribuição Normal	37
Figura 4.1. Exemplo da colocação dos pontos para determinação da área da extremidade da clavícula (Ae) no <i>software</i> ImageJ	44
Figura 4.2. Exemplo da colocação dos pontos para determinação da área de ossificação da clavícula (Ao) no <i>software</i> ImageJ	45
Figura 4.3. Ilustração das medidas de avaliação do erro “Intervalo de Confiança 95%” e “Erro”	49
Figura 5.1. Regressão da Idade sobre o indicador Rácio	58
Figura 5.2. Regressão dos Erros sobre a Idade Real	59
Figura 5.3. Gráfico da regressão do Rácio sobre a Idade	60
Figura 5.4. Recta dos Erros provenientes da Calibração. Recta de Regressão assinalada a tracejado vermelho	61

Resumo

Uma amostra composta por 158 clavículas (85 Femininas, 73 Masculinas) provenientes da Coleção de Esqueletos Identificados do Museu Antropológico da Universidade de Coimbra (n = 105) e da Coleção de Esqueletos Identificados do Museu Bocage – Lisboa (n = 53) – foi estudada com o intuito de ser criada uma nova metodologia de estimativa de idade à morte através da quantificação da evolução do processo de união da epífise medial à diáfise da clavícula.

Os objectivos do estudo passaram pela criação de uma metodologia capaz de responder às necessidades actuais da Antropologia Forense, no que à estimativa da idade à morte diz respeito, e também por aplicar aos dados recolhidos várias abordagens estatísticas já utilizadas em estudos distintos (Regressão linear, Calibração Clássica) e introduzir uma nova (Regularização local sem e com distribuição marginal conhecida). Desta forma pudemos compreender, através da comparação dos resultados, qual a melhor abordagem estatística ao tratamento deste tipo de dados. Foi ainda criado um modelo através de uma sub-amostra retirada da nossa amostra inicial composta apenas por 19 indivíduos com o intuito de testar o comportamento da nova abordagem estatística quando aplicada a amostras de tamanho reduzido.

A análise preliminar dos dados recolhidos permitiu verificar que o primeiro indivíduo a concluir o processo de união da epífise à diáfise em estudo era do sexo feminino e tinha 23 anos, sendo que toda a amostra já tinha o processo de união concluído aos 28 anos de idade. Dos 5 modelos criados o que mostrou resultados mais promissores foi o modelo tratado com a Regularização local com distribuição marginal conhecida, permitindo estimar a idade com uma grau de certeza de 100% na resposta dada e ainda associando aos intervalos etários uma probabilidade de o indivíduo em análise se encontrar nos respectivos grupos etários. Este tipo de resposta vem trazer uma maior credibilidade à análise do perito perante uma entidade judicial, à imagem do que acontece noutras áreas da ciência forense, como a Genética, por exemplo. Outro resultado de salientar refere-se ao comportamento do método de Regularização local com distribuição marginal conhecida quando aplicado à amostra reduzida, tendo os resultados ficado muito acima da expectativa, dando-nos este modelo um grau de credibilidade de 89%.

Este estudo teve também como objectivo realçar algumas questões que podem introduzir erro tanto na criação de metodologias de estimativa da idade à morte como na sua aplicação e sugerir algumas soluções para as mesmas.

Abstract

In this study, we sought out to develop a new methodology age at death estimation through the quantification of the process of union between the medial epiphysis and diaphysis of the clavicle. A sample comprised by 158 clavicles (85 females, 73 males) from the Collection of Identified Skeletons of the Anthropological Museum of the University of Coimbra (n = 105) and from the Collection of Identified Skeletons of Museu Bocage – Lisbon – (n = 53) was used for this purpose.

Also, we aimed to test in our sample several statistical approaches that have been used across different studies (Linear Regression, Classical Calibration) and also a new one (Local smoothing without and with given margins). We compared the results of these different statistical methodologies in order to understand which method would be more suitable for this type of data. Furthermore, from a subsample of 19 individuals selected from the original sample, we created a model to test the performance of the new statistical approach when applied to small samples.

A preliminary analysis of the data showed that the first individual to conclude the process of union between the epiphysis and diaphysis was a 23 year old female, and that the whole sample had this process concluded by the age of 28. From the 5 models created, the model with Local smoothing with given margins showed the best results, allowing to estimate age at death with 100% certainty. Also, it allowed us to associate a probability to age intervals of an individual being in the age groups. We believe this brings more credibility to the expert analysis in forensic contexts, as it occurs in other areas of forensic sciences, such as Genetics. Another important result refers to the performance of the Local smoothing with given margins method when applied to the reduced sample, providing a degree of credibility of 89%, which was above our expectations.

This study also aimed to highlight several questions that might bring error in the development of methodologies to estimate age at death or in their application, and to suggest possible solutions.

Capítulo 1 – Introdução

“(...) that branch of physical anthropology, which, for forensic purposes, deals with the identification of more or less skeletonized remains known to be, or suspected of being human”

Stewart, TD (1979)

Esta é apenas uma das várias definições que, ao longo do tempo, foram atribuídas à Antropologia Forense. Cada uma delas reflecte de certa forma o pensamento e contexto da época em que foi criada, sendo assim possível traçar uma linha evolutiva nesta disciplina através de uma análise das definições que lhe foram sendo propostas.

Apesar desta definição datar de 1979, já muitos anos antes o conhecimento da antropologia física servia para ajudar a resolver questões de índole médico-legal. Como exemplos podemos referir o trabalho do especialista em desenho anatómico, o francês Jean-Joseph Sue (1710 - 1792), que através da medição de cadáveres de não adultos (de fetos a jovens adultos) forneceu as ferramentas para a estimativa da estatura da população Francesa. Apesar do seu objectivo ser o de recolher informações para artistas acerca das proporções corporais, o seu trabalho foi aproveitado por Matthieu-Joseph-Bonaventure Orfila (1823) que incluiu os dados de Sue no seu trabalho dando assim as bases necessárias para que este tipo de análises se iniciasse (Ubelaker, 2006). O trabalho de Paul Broca (1824-1880) é também incontornável em qualquer texto introdutório de Antropologia física e forense. O seu contributo para o conhecimento das medições ósseas, assim como a criação de alguns instrumentos para o mesmo propósito e as discussões que iniciou em torno do tema, foram um grande impulso para o conhecimento da variabilidade humana e trouxeram maior cientificidade à análise óssea (Spencer, 1997). Não será de todo errado também incluir aqui o trabalho de Karl Pearson (1857 – 1936) e a forte influência que a sua monografia sobre equações de regressão (Pearson, 1899) teve e tem em inúmeros trabalhos de antropologia forense, inclusivamente no presente estudo.

Como exemplo de um dos primeiros casos em que a antropologia física foi utilizada em contexto legal podemos referir o de Jeffries Wyman (1814 – 1874), professor de anatomia de Harvard, que conseguiu identificar os restos esqueléticos queimados encontrados no

laboratório de John W. Webster como sendo pertencentes a George Parkman, um médico que terá sido morto e posteriormente desmembrado por Webster que queimou parte dos restos cadavéricos de Parkman no seu laboratório com o intuito de esconder o crime que cometera (Ubelaker, 2006).

Três pontos-chave para a Antropologia Forense que não podem deixar de ser mencionados foram a criação da American Academy of Forensic Sciences (AAFS) em 1972, da American Board of Forensic Sciences (ABFS) em 1977 e da Forensic Anthropology Society of Europe (FASE) em 2003.

A definição apresentada no início do capítulo foi dada por T. D. Stewart em 1979 num texto de grande influência para a Antropologia Forense: *Essentials of Forensic Anthropology, Especially as Practiced in the United States*, que Stewart escreve enquanto curador no Museu Smithsonian onde verificou que a maioria dos ossos que lhe chegavam ao laboratório vinham embrulhados em papel ou cartão, completamente separados do contexto em que foram encontrados. Esta definição representa o que se pode dizer ser o início da distinção entre antropologia forense e a já bem estabelecida medicina / patologia forense (Ubelaker, 2006) e Stewart acrescentou no mesmo ano (Stewart, 1979b) que o papel do antropólogo forense será o de “*para além da eliminação dos elementos não humanos, o processo de identificação tem que informar acerca do sexo, idade, raça, estatura, assim como outras características de cada indivíduo envolvido que possam levar à sua identificação*”.

Analisar todas as definições seria mais exaustivo do que esta introdução pretende ser, pelo que para se conseguir compreender de forma correcta, mas sucinta, a evolução desta disciplina nos últimos 35 anos, podemos analisar as diferenças entre a definição dada por Stewart em 1979 e a definição actualmente dada pela American Board of Forensic Anthropology (2006):

“Forensic Anthropology is the application of the science of physical anthropology to the legal process. The identification of skeletal, badly decomposed, or otherwise unidentified human remains is important for both legal and humanitarian reasons. Forensic anthropologists apply standard scientific techniques developed in physical anthropology to identify human remains, and to assist in the detection of crime. Forensic anthropologists frequently work in conjunction with forensic pathologists, odontologists, and homicide investigators to identify a decedent, discover evidence of foul play, and/or the postmortem

interval. In addition to assisting in locating and recovering suspicious remains, forensic anthropologists work to suggest the age, sex, ancestry, stature, and unique features of a decedent from the skeleton.”

Da análise comparativa destas duas definições, algumas diferenças saltam imediatamente à vista. Primeiro, é claramente visível que o papel de um antropólogo forense actualmente tem um rol de objectivos mais amplo do que tinha há 35 anos atrás. Enquanto Stewart faz apenas referência à identificação, actualmente é já do âmbito de estudo e análise da antropologia forense assistir, dentro do que é possível, no esclarecimento da causa e circunstância da morte, estudar o intervalo *post mortem*, trabalhar a par com médicos em casos de suspeita de traumas repetidos (Dirkmaat et al., 2002), participar activamente em investigações de violações de direitos humanos, entre outros. Outra diferença diz respeito aos casos em que é solicitado o envolvimento do antropólogo forense. Stewart refere que este deve participar em casos de “*more or less skeletonized remains*” o que contrasta com o actual “*skeletal, badly decomposed, or otherwise unidentified human remains*”. Reichs (1995) reporta que em 1992 o envolvimento dos antropólogos se distribuía da seguinte forma: 63% dos casos envolviam esqueletos ou crânios, 13% cadáveres queimados, 11% cadáveres decompostos, 11% cadáveres frescos e 2% mumificados. Já no ano de 1996, Ubelaker (2000) reporta que a percentagem de casos com cadáveres decompostos e frescos sobe para 38,7% e 23%, respectivamente. Apesar desta tendência certamente não agrada a antropólogos com estômagos mais sensíveis, é certamente indicativa da importância da disciplina e do reconhecimento dessa importância por parte do sistema médico-legal.

Uma visão mais aprofundada da história da Antropologia Forense, pode ser encontrada em Bass (1969, 1979), Snow (1982), Ubelaker (1996, 1997, 2000a, 2000b, 2000c) e Buikstra et al. (2003).

Como é possível ler na definição de Antropologia Forense avançada em 2006 pela ABFA, um dos objectivos do antropólogo passa pela diagnose sexual, pela estimativa da idade à morte e da estatura e pela avaliação da afinidade biológica do indivíduo em análise, isto é, pela elaboração do perfil biológico do indivíduo. O presente estudo visa a criação de uma metodologia de estimativa da idade à morte, pelo que no seguimento desta introdução o foco de interesse recairá sobre essa mesma temática.

1.1 – História da Estimativa da Idade à Morte

Desde os primeiros estudos que foram feitos neste âmbito, há mais de 450 anos (Santos, 1995), muita coisa mudou no que respeita à escolha das amostras, dos indicadores ósseos a que se dá maior importância, dos processos de recolha dos dados, das abordagens estatísticas com que se tratam os dados, entre outros. Escrever sobre a história e evolução do estudo e das metodologias da estimativa da idade à morte é uma tarefa complexa e quase enciclopédica pelo que aqui serão focados apenas os principais tópicos e desenvolvimentos desta temática.

Por entre os vários estudos que podem ser citados para uma breve descrição da evolução deste tema o *The Identification of the Human Skeleton, A Medico-Legal Study* de Thomas Dwight (1878) é possivelmente o mais indicado para iniciar a história da estimativa da idade à morte. No capítulo dedicado à estimativa da idade, Dwight deixa algumas notas que ainda hoje são merecedoras de atenção. Um exemplo são as 4 classes em que divide o esqueleto humano adulto, estabelecendo logo na primeira a diferença entre indivíduos masculinos e femininos, o que deixa transparecer o conhecimento do processo de desenvolvimento do esqueleto humano e o facto de este não ocorrer sempre ao mesmo ritmo durante toda a vida, assim como da influência do sexo neste processo. É hoje aceite que esta pré-classificação do indivíduo em análise num sub-grupo, é importante para melhor se poder escolher o método mais correcto a utilizar (Baccino & Schmitt, 2006). Outro ponto de interesse salientado pelo autor são os possíveis indicadores que os antropólogos têm à disposição na estimativa da idade à morte, tais como a união das epífises às diáfises, a obliteração de linhas de união, o encerramento das suturas cranianas e, em estádios “senis” (Dwight, 1878) a alteração na forma e constituição dos ossos. De salientar também o facto de o referido autor realçar que o grau de dificuldade da estimativa da idade à morte aumenta com a idade do indivíduo assim que este deixa de ser “imaturado (< 25 anos)”. Este trabalho de Dwight peca por não fazer praticamente referência nenhuma aos dentes, com a excepção da referência ao 3º molar comparando-o aos comboios de algumas estações “*due when they arrive*” (1878:39) salientando o carácter variável deste indicador no que respeita à idade da sua erupção.

Outro trabalho que merece referência é o de Aleš Hrdlička “*Anthropometry*” (1920), em que o autor salienta a natureza variável do desenvolvimento do esqueleto humano não só entre sexos mas também entre diferentes ascendências biológicas e grupos sociais. Este

trabalho merece referência também pelo facto de Hrdlička apresentar o que pode ser considerado uma abordagem multivariada do esqueleto para a estimativa da idade à morte, nomeadamente através da conjugação da obliteração das suturas cranianas com o desgaste dentário.

Ambos os autores acima referidos fizeram também referência à grande variação que está por detrás do desenvolvimento do esqueleto humano e, conseqüentemente, às análises e inferências que se fazem a partir deste, como a estimativa da idade à morte. Foi possivelmente com esta ideia em mente que T.D. Stewart reconheceu a necessidade de informação proveniente de novas amostras (Ubelaker, 2010), com o propósito de melhor se conseguir estudar essa variabilidade. Tendo como base um total de 450 esqueletos pertencentes a soldados Americanos mortos em combate durante a guerra na Coreia do Norte, Mckern e Stewart recolheram dados referentes às alterações esqueléticas decorrentes da idade e publicaram a clássica monografia “*Age Changes in Young American Males*” (1957). Este trabalho, apesar de ter sido realizado apenas com esqueletos do sexo masculino com idades compreendidas entre os 17 e os 50 anos (idade militar) tem uma enorme importância devido ao grande detalhe com que as alterações esqueléticas decorrentes da idade são examinadas. Há ainda que fazer referência ao “*The Human Skeleton in Forensic Medicine*” de Krogman (1962), que é a primeira grande síntese sobre metodologias de estimativa da idade à morte.

1.2 – O interesse da Estimativa da Idade à Morte

A estimativa da idade à morte é um tema de interesse comum a várias áreas de estudo, desde a Antropologia à Arqueologia, passando pela Biologia, Medicina, Ciências Criminais, História e Paleontologia. (Santos, 1995), tendo em cada uma destas áreas finalidades distintas consoante o objectivo a que cada uma se propõe. Para a antropologia física e forense uma correcta estimativa da idade à morte apresenta-se imprescindível na construção de um perfil biológico.

Sendo os ossos e os dentes os tecidos mais resistentes do corpo humano, são muitas vezes estes os únicos indicadores biológicos que nos chegam de populações pretéritas ou surgem em casos forenses. No entanto, uma vez que estes dois contextos antropológicos têm objectivos distintos, uma correcta estimativa da idade à morte pode servir tanto para construir o perfil demográfico de uma população encontrada num contexto arqueológico, como para

ajudar na obtenção de um perfil biológico que leve à identificação de um indivíduo. De uma forma geral, podemos afirmar que num contexto arqueológico a estimativa da idade à morte serve para informar acerca de uma população (construção de tábuas de sobrevivência, associação entre idades e maior incidência de determinadas patologias, etc.) e num contexto forense para tirar elações acerca do indivíduo em análise, de forma a permitir a correcta identificação deste (por exemplo).

Para além dos objectivos serem distintos existem também diferenças a nível do material ósseo a analisar (devido a processos tafonómicos, por exemplo), de requerimentos feitos pelas entidades judiciais e de tempo dado para a análise, o que leva a que os métodos utilizados num estudo de cariz arqueológico não possam na maioria das vezes ser utilizados numa perícia forense (Cunha et al, 2009). No entanto, ainda que os métodos não possam (ou não devam) ser utilizados em ambos os contextos, a criação destes baseia-se no conhecimento estrutural comum da Antropologia Física, o que leva a que em não raras vezes, um método criado para utilização em casos forenses e outro para ser utilizado em contexto arqueológico tenham apenas de diferente a amostra base utilizada, isto é, a amostra identificada a partir da qual se extraem parâmetros que permitem posteriormente realizar inferências acerca de indivíduos de identidade desconhecida.

Uma correcta estimativa da idade à morte é uma parte essencial do processo de identificação de um desconhecido uma vez que a idade de uma pessoa é um dos factores cruciais na sua identidade, razão pela qual é sistematicamente solicitada pelas autoridades judiciais quando se deparam com um cadáver de identidade desconhecida. Uma estimativa incorrecta vai comprometer de forma significativa um processo que tem consequências não só a nível administrativo e ético, mas também legais, uma vez que sem uma correcta identificação uma pessoa não pode judicialmente ser declarada morta o que no âmbito do direito civil implica que o estatuto jurídico da pessoa em causa e também dos seus familiares vai estar alterado de uma forma incorrecta e com as previsíveis consequências que daí advêm (a nível de seguros, heranças, etc..) (Ritz-Timme et al., 2000). No âmbito do direito penal, as consequências de uma incorrecta identificação ou falta desta podem conduzir a um processo judicial que vai ocorrer de forma enviesada podendo resultar numa detenção de alguém inocente ou na declaração de inocência de alguém culpado.

1.3 – Metodologias de Estimativa de Idade à Morte

Existem diversas metodologias de estimativa da idade à morte. Estas variam consoante o indicador utilizado, o grupo etário alvo para a aplicação do método, o contexto do qual o indivíduo ou população em que se quer aplicar o método provem, e até mesmo a população em que o método é desenvolvido e deve ser aplicado. A escolha do melhor método está por isso sempre dependente de algumas variáveis, entre elas: o indivíduo em análise (não adulto, jovem adulto, adulto), material ósseo disponível para análise, população de origem, instrumentos e recursos disponíveis para a análise, experiência do antropólogo, entre outros.

Para os indivíduos não adultos o processo de desenvolvimento dentário é o que apresenta melhor correlação com a idade (Ubelaker, 1990). Contudo, nem sempre a análise das peças dentárias é possível, ou, ainda que seja, é sempre importante recolher o máximo de informação disponível no esqueleto em análise, pelo que a observação deste deve-se estender para lá do exame dentário, complementando-se com métodos que têm por base o crescimento e os processos de maturação do esqueleto. Por crescimento entende-se o aumento do tamanho dos ossos que pode ser acompanhado por alterações na forma e na função, o que corresponde aos processos de maturação (Klepinger, 2006). A análise recai assim na observação de centros de ossificação e fusão (processo de maturação), no tamanho dos ossos (particularmente dos longos) e nos estádios de união epifisial (Ubelaker, 2010). Quando este processo de desenvolvimento termina é atingida a maturidade. Atingida a maturidade, isto é, chegando à idade adulta, os indicadores através dos quais se realiza a estimativa estão relacionados com processos degenerativos e de remodelação dentária e óssea (Baccino & Schmitt, 2006).

Como tinha antecipado Dwight (1878), a estimativa da idade à morte nos adultos é mais difícil, uma vez que os indicadores utilizados apresentam uma menor correlação com a idade (Klepinger, 2006), variando também em função de outros factores (Jackes, M., 2000). Como facilmente se depreende, uma maior variabilidade no processo biológico de envelhecimento produz efeitos na estimativa da idade à morte (Baccino & Schmitt, 2006), fazendo com que a diferença entre a idade cronológica – que se pretende estimar – e a idade biológica – a estimada – seja maior, ou seja, o erro na estimativa vai ser maior.

Para uma melhor compreensão das diferentes metodologias existentes para estimar a idade à morte em adultos, opta-se aqui por fazer uma divisão entre métodos macroscópicos e microscópicos.

No que diz respeito aos métodos macroscópicos para indivíduos adultos, a maioria baseia-se na análise de articulações não sinoviais. Nestas, são analisados alguns indicadores avançados pelos autores do respectivo estudo, tais como osteofitoses, micro e macroporosidades, granulações, entre outros e, de acordo com a avaliação total das características presentes no indivíduo em análise, estas vão estar relacionadas com classes etárias predefinidas para o respectivo método. São exemplos deste tipo de métodos Lovejoy et al (1985), que analisa a superfície auricular do *ilium* e que foi posteriormente revisto e simplificado por Buckberry e Chamberlain (2002); Todd (1920), que estudou a sínfise púbica e cujo método já foi revisto por diversos autores, entre eles Suchey - Brooks (Brooks & Suchey, 1990); Iscan et al (1984), num método de análise da ossificação da extremidade esternal das costelas. Com uma abordagem um pouco diferente das metodologias acima citadas há ainda que fazer referência ao *Two Step Procedure* (TSP) (Baccino & Zerili, 1997) que combina de forma simples o método de Suchey – Brooks (1990) e o de Lamendin et al (1992), aumentando comprovadamente a precisão da estimativa comparativamente aos dois métodos utilizados isoladamente (Baccino & Schmitt, 2006).

Os métodos microscópicos de estimativa da idade à morte baseiam-se no pressuposto de que a substituição de osso cortical primário por secundário acontece a um ritmo que pode ser predito (Pfeiffer, 1992). Este tipo de estudos iniciaram-se com Kerley (1965) que, através de um exame microscópico do osso cortical humano, contou o número de osteons e canais não haversianos em 4 pontos distintos da margem periosteal do córtex femoral e efectuou a estimativa da percentagem de osso lamelar ainda presente nesses 4 locais. Posteriormente alterações a este método foram propostas (Ahlqvist & Damsten, 1969; Singh & Gunberg, 1970). Ao reverem o estudo inicial de Kerley, Bouvier e Ubelaker (1977) encontraram alguns erros, que foram emendados num estudo posterior realizado por Kerley e Ubelaker (1978). Este tipo de metodologias têm-se mostrado bastante útil principalmente nas idades em que é mais difícil realizar a estimativa pelos métodos macroscópicos (acima dos 50) e por reduzirem a amplitude dos intervalos de idade fornecidos (Thompson, 1979; Stout & Gehlert, 1980; Ubelaker, 1986).

Apesar de sucinta, esta introdução às metodologias de estimativa da idade à morte deixa já transparecer que o número de métodos disponíveis é bastante grande, sendo esta área de estudo da antropologia forense aquela em que possivelmente mais publicações e estudos são realizados (Ritz-Timme et al., 2000; Cunha et al., 2009). Ainda assim, esta é também a área da antropologia forense em que mais dificuldades continuam a existir. Estas prendem-se com diversas razões e serão discutidas no tópico seguinte.

1.4 – Fontes de Erro na Estimativa da Idade à Morte

Apesar de ser, como já referido, dos temas mais estudados no âmbito da antropologia física e forense, a estimativa da idade à morte continua a ser das temáticas com menor consenso entre os investigadores. Esta falta de acordo é possivelmente a origem de muitas das fontes de erro que surgem em métodos utilizados para estimar a idade. Ao referirmo-nos a “Método” aqui utilizamos a definição dada por Schmeling et al (2007) – “*a transformação das descobertas de um processo ontogénico para uma escala cronológica*”.

Factos como a referência a métodos “preferidos” em livros e artigos sem estar esta escolha justificada por qualquer evidência estatística; a falta de estudos que testem o mesmo método em várias amostras independentes obedecendo a um protocolo fixo; a utilização de diferentes medidas de avaliação da precisão dos métodos de estudo para estudo e nem sempre utilizadas de forma correcta (Ritz-Timme et al., 2000); o não cumprimento de *guidelines* existentes (Ritz-Timme et al., 2000; Schmeling et al., 2007; Cunha et al., 2009) para a construção de novas metodologias; a má utilização de testes estatísticos (apontada como uma das principais limitações à aplicação prática de alguns métodos) (Ritz-Timme et al., 2000), são algumas das fontes que introduzem erro tanto na criação de metodologias como na aplicação das já existentes.

Mas as fontes de erro na estimativa da idade à morte não derivam todas de erros de investigação ou da falta de uniformização da forma de investigar e aplicar resultados. Existe um tipo de erro que será certamente muito mais difícil (se possível) de corrigir e que está inerente ao próprio indivíduo. Este erro é a diferença entre a idade cronológica e a idade biológica. De facto, esta diferença é crucial para a correcta compreensão desta temática.

Por idade cronológica compreende-se o tempo decorrido desde o nascimento do indivíduo até um determinado dia; por sua vez a idade biológica é uma idade estimada através de um conjunto de indicadores ósseos e/ou dentários que se alteram ao longo do tempo e têm por isso uma maior ou menor correlação (dependendo do indicador) com a idade cronológica. No entanto, esta relação não é simples, uma vez que os indicadores de idade reflectem sempre uma adaptação contínua dos ossos a stresses biomecânicos, metabolismo, crescimento, remodelações e a doença (Nawrocki, 2010). A ideia geral de que restrições biomecânicas influenciam a estrutura óssea é comumente referida como “Lei de Wolff” (Ruff, 2008).

É por isto de esperar que, dois indivíduos com uma idade cronológica precisamente igual tenham uma idade biológica diferente, basta para isso que tenham estado sujeitos durante as suas vidas a pressões biomecânicas e/ou a contextos socioeconómicos e/ou culturais distintos, o que é praticamente inevitável que aconteça, mesmo em ambientes muito semelhantes. Isto acontece porque à medida que o tempo passa alterações biomecânicas e fisiológicas vão sendo acumuladas de uma forma que não é regular ao longo da vida e que está dependente de inúmeras variáveis. Uma simples analogia com o que acontece com um lápis ao longo do tempo consegue clarificar esta relação entre alterações nos indicadores ósseos e/ou dentários e idade cronológica. Sendo que o lápis, desde o início da sua utilização, vai diminuindo de tamanho com o tempo, pode dizer-se que o tempo “causa” uma diminuição do tamanho do lápis. O problema com este raciocínio é que se comprarmos o lápis e não o utilizarmos durante um ano, este vai ficar precisamente do mesmo tamanho, apesar do tempo passar. Neste caso, o tempo é irrelevante para a diminuição do tamanho quando comparado com os reais factores que fazem com que o lápis se gaste, como a pressão exercida na sua utilização, a textura das superfícies em que é utilizado, etc. O que acontece é que estes factores são claramente mais difíceis de medir e quantificar do que o passar do tempo (a idade cronológica do lápis) e, neste sentido, a idade torna-se num substituto mais simples (uma variável proxy) dos complexos factores que na realidade provocam alterações e cujos efeitos se acumulam apenas de forma secundária com o passar do tempo.

Na estimativa da idade à morte de um esqueleto a situação é semelhante e o adágio popular “o tempo passa e nós ficamos mais velhos”, não tem implícito uma relação tão linear entre estas duas variáveis (tempo e idade) como à partida possa parecer. A variação nos indicadores ósseos e/ou dentários (desgaste do lápis) não é provocada pela passagem do tempo *per se*, mas sim pelo acumular de processos fisiológicos e biomecânicos (factores

como a pressão exercida no lápis e a superfície em que é utilizado) que pela sua natureza variável não estão relacionados com o tempo medido por um calendário de forma linear. As nossas articulações estão constantemente a ser submetidas a stresses biomecânicos, mas o grau e a natureza desse stress e o conseqüente constrangimento sentido pelo osso varia consoante a nossa massa muscular, padrões de actividade, composição dos tecidos conectivos (Nawrocki, 2010). Para além disso, estes efeitos biomecânicos acontecem sob um *background* de sistemas e de processos fisiológicos localizados que estão constantemente a acontecer no corpo, tal como o metabolismo do oxigénio, a absorção de nutrientes, o funcionamento dos sistemas de excreção, actividade endócrina e a secreção hormonal, a activação do sistema imunitário, entre outros processos que variam ao longo da vida e entre indivíduos (Nawrocki, 2010). É por esta razão que apelidámos a idade cronológica de uma proxy ou substituta de toda a complexidade de fenómenos que acontecem durante o desenvolvimento e maturação óssea.

Estimar a idade cronológica é um processo que envolve a transformação de indicadores ósseos e/ou dentários numa idade biológica, que traduz o grau de desenvolvimento, maturação ou degeneração do esqueleto e que pode ser utilizada para inferir a idade cronológica que permite informar a identificação de um indivíduo de identidade desconhecida. Este processo de transformação é chamado, neste contexto, de calibração – processo através do qual o valor de uma variável que não pode ser facilmente observável, pode ser estimado através do conhecimento do valor de outras variáveis relacionadas com a primeira e que podem ser observadas facilmente. O que acontece é que nenhum indicador tem uma correlação perfeita com a idade cronológica, pelo que será de esperar que todo e qualquer método baseado na análise de indicadores ósseos e/ou dentários tenha associado uma margem de erro que, na melhor das hipóteses, advém apenas desta diferença inevitável entre a idade cronológica e a biológica.

O erro acima mencionado é praticamente eliminado numa metodologia recentemente utilizada por Ubelaker e Parra (2011) que, ao efectuarem uma análise de radiocarbono ao esmalte dentário, conseguem estimar não a idade biológica de um indivíduo, que levaria ao referido erro, mas sim o ano de nascimento e de morte, o que permite facilmente obter a estimativa da idade à morte do indivíduo.

1.5 – Revisão da Literatura: Metodologias de Estimativa da Idade à Morte pela Clavícula

A clavícula é um osso longo que faz parte da cintura escapular, articulando medialmente com o *manubrium* e lateralmente com o processo acromial da escápula. Por ser um osso compacto e de certa forma robusto, tende a preservar-se bem mesmo após longos períodos de inumeração ou depois de submetido a condições tafonómicas adversas, o que faz com que seja um osso que desperte interesse para estudos antropológicos (Lin, 1991 *in* Scheuer & Black, 2000). De facto, existem vários estudos a atestar a sua utilização na diagnose sexual (Thieme & Schull, 1957; Iordanidis, 1961; Steel, 1966), na estimativa da estatura (Jit e Singh, 1956) e até mesmo na identificação da lateralidade dominante (Steele & Mays, 1995) ou a informar em casos de identificação (Sanders et al., 1972). Ainda assim, a sua utilização mais comum é para a estimativa da idade à morte, e é esta que o presente estudo vai analisar.

O facto de a clavícula ser o primeiro osso a iniciar o processo de ossificação, entre a 5ª e a 6ª semana de vida intra-uterina, e o último a completar a união epifisial (Scheuer & Black, 2000) em muito contribui para que seja um bom indicador para a estimativa da idade à morte, nomeadamente através do comprimento da sua diáfise enquanto esta se encontra em crescimento (até o indivíduo atingir a puberdade) e, após o crescimento cessar, através da fusão da sua epífise medial à diáfise, um processo que demora, em média, 10 anos (Scheuer & Black, 2000). Após um período inicial de ossificação da epífise na altura da puberdade (Ogden, 1979) a fusão da epífise à diáfise inicia-se entre os 16 e os 21 anos estando completa antes dos 30. No entanto estas idades apresentam alguma variabilidade que pode ser explicada pelo facto de indivíduos diferentes terem tempos de maturação diferentes devido a factores genéticos e ambientais (Scheuer & Black, 2000).

Como ficou patente no parágrafo anterior, podemos dividir o processo de ossificação e união da epífise medial à diáfise da clavícula (processo que daqui para a frente passaremos a chamar apenas de união da epífise à diáfise por uma questão de melhor leitura) em 3 fases distintas:

1. Crescimento da diáfise
2. Ossificação da epífise
3. União da epífise à diáfise

Tendo isto e a forma de análise da clavícula em mente, podemos agrupar os vários métodos de estimativa da idade à morte pela clavícula existentes da seguinte forma:

- Métodos Anatómicos
 - Comprimento da diáfise
 - Análise da Epífise
 - Ossificação e união da epífise à diáfise
 - União da epífise à diáfise
- Métodos Imagiológicos
 - Ultra-som
 - Radiografia
 - CT-Scans

Esta divisão pretende elucidar as várias formas de analisar a clavícula através das quais se criam as metodologias de estimativa da idade à morte. Os “Métodos Anatómicos” são aqueles que partem de uma análise da clavícula com o osso já seco, livre de tecidos moles. Através destes métodos é possível realizar medições do “Comprimentos da Diáfise” através do qual se estabelece uma correlação com a idade estimada (Scheuer & Black, 2000). Outra forma de análise, talvez a mais comum, é através da “Análise da epífise”. Esta análise pode ser feita em duas situações: quando, no indivíduo em análise ou na colecção em estudo estão preservadas as epífises que, já tendo iniciado o processo de ossificação, ainda não se encontram fundidas com a diáfise (Webb & Suchey, 1985); ou quando estas epífises não estão presentes e apenas é possível observar o processo de “União da epífise à diáfise” (Schaefer, 2008; Mckern & Stewart, 1957; Todd & D’Errico, 1928; Stevenson, 1924).

O presente estudo enquadra-se neste último grupo, que acreditamos ser o mais comumente encontrado num caso prático de antropologia forense, uma vez que a epífise é uma peça óssea muito frágil e de reduzidas dimensões pelo que, caso não se encontre unida à diáfise, facilmente se perde, o que leva a uma maior necessidade de métodos que tenham em consideração apenas o processo de união.

Os “Métodos Imagiológicos” também contemplam as três situações acima descritas para os métodos anatómicos, com a diferença que este tipo de metodologias emprega técnicas de imagiologia para a observação das referidas características. As principais vantagens destes métodos estão relacionadas com o facto de se poderem desenvolver e aplicar em indivíduos vivos e de permitirem uma comparação de imagens obtidas em vida (Raio-X ao tórax por

exemplo) com imagens realizadas *post mortem* que, desde que cumpridas as mesmas normas de realização da imagem, podem possibilitar uma identificação positiva (Stephan et al., 2011). Os estudos em que são realizados ultra-sons estão relacionados na sua maioria com a estimativa da idade pré-natal, e a idade é estimada através da sua correlação com o comprimento da diáfise da clavícula (Fazekas & Kosa, 1978; Sherer et al., 2006). No entanto, esta técnica também pode ser utilizada para estimar a idade pós-natal em indivíduos jovens adultos (Quirnbach et al., 2009). Os estudos que utilizam radiografias (Flecker, 1932; Galstaun, 1937; Jit & Kulkarni, 1976; Walker & Lovejoy, 1985; Kreitner et al., 1998; Schmeling et al., 2004) estão reportados como sendo os que podem levar a uma estimativa mais enviesada devido à sobreposição de outras estruturas, como costelas, vertebrae e sombras mediastinais (Kreitner et al., 1998), que impedem uma correcta leitura da fase de ossificação. Este problema parece ficar resolvido com a utilização de CT-scans, que dão uma imagem mais detalhada dos tecidos moles e por isso permitem uma melhor leitura da fase de ossificação ou união da epífise à diáfise, desde que realizados cortes com a devida espessura na imagem (Schmeling et al., 2004). Exemplos de estudos realizados com CT-scans são Kellinghaus, et al., (2010), Bassed, et al., (2010), Kreitner et al., (1998) e Schultz et al., (2008).

Todos estes estudos apresentam o processo de ossificação e união da epífise à diáfise dividido em várias fases. Na tabela 1.1 é possível ver um resumo de alguns dos estudos referidos, indicando os resultados que foram obtidos pelos autores assim como algumas características específicas de cada um.

Do estudo da clavícula, para além da aplicação prática de cada método e da resposta a um caso específico, resulta também outro tipo de informação que merece a atenção dos antropólogos que estudam este indicador ósseo. Saber a idade mínima a que é esperado observar a ossificação completa da epífise à diáfise é uma informação de grande importância para a antropologia forense, mais concretamente para a estimativa da idade no vivo. Nesta área a questão mais importante a que normalmente se pretende responder é se o indivíduo já atingiu ou não a maioridade (Schmeling et al., 2004) e é aqui que surgem as principais diferenças entre os resultados obtidos pelos vários autores. Os estudos existentes definem o estágio final de ossificação como aquele em que a cartilagem da epífise já está completamente ossificada mas sem reportar se a linha de ossificação está ou não visível (Schmeling et al., 2004). Para esta fase são reportadas várias idades distintas, consoante o estudo que temos por base

Tabela 1.1

Resumo dos dados recolhidos e resultados obtidos por autores de estudos de idade à morte com base na clavícula

Grupo de estudo	Nt	Nm	Nf	Intervalo etário	Nº de Fases	Método	Idade mínima	Idade mínima (amostra total)
Black e Scheuer (1996)	143	82	61	0 a 30	5	Anatómico (C. e U.)	25	28
Webb e Suchey (1985)	859	605	254	11 a 40	4	Anatómico (O. e U.)	21	34
Mckern e Stewart (1957)	374	374	0	17 a 31	5	Anatómico (O. e U.)	23	31
Todd e D'Errico (1928)	166	130	36	17 a 29	4	nd	22	28
MacLaughlin (1990)	30	20	12	11 a 39	5	Anatómico (U.)	24	29
Stevenson (1924)	110	49	61	15 a 28	4	Anatómico (U.)	nd	28
Quirnbach et al. (2009)	77	77	0	18 a 24	4	Ultra som	nd	nd
Flecker (1933)	437	206	231	11 a 26	nd	Raio-X	22	26
Galstaun (1937)	654	446	209	12 a 25	nd	Raio-X	19	25
Jit e Kulkarni (1976)	684	391	193	11 a 30	nd	Raio-X	22	25
Kreitner (1998)	380	229	151	0 a 29	nd	CT-scan	22	27
Schmeling (2004)	699	243	456	16 a 30	5	Raio-X	20	26
Kellinghaus et al. (2010)	185	104	81	13 a 26	5	CT-scan	21	26
Bassed et al. (2010)	674	455	219	15 a 25	5	CT-scan	18	≥27
Kelling (2009)	502	288	214	10 a 35	5	CT-scan	21	26
Presente estudo	157	73	85	10 a 30	4	Anatómico (U.)	23	28

Nota. Nt = Número total de indivíduos na amostra; Nm = Número de indivíduos do sexo masculino na amostra; Nf = Número e indivíduos do sexo feminino na amostra; Anatómico (C. e U.) = Método Anatómico com estudo do Comprimento e da União; Anatómico (O. E U.) = Método Anatómico com estudo da Ossificação e da União; Anatómico (U.) = Método Anatómico com estudo da União; nd = informação não disponível

Uma vez que neste estudo o método que vai ser desenvolvido se enquadra nos Anatómicos com análise da união, importa fazer uma revisão e discussão mais detalhada de outros métodos com esta abordagem.

A grande maioria dos estudos realizados com este tipo de metodologia baseia-se nos *standards* que foram apresentados por Stevenson (1924), Todd e D’Errico (1928) e McKern e Stewart (1957). É prática comum em antropologia física e forense apenas serem criadas algumas versões de um tipo de metodologia (neste caso 3) e posteriormente serem feitas adaptações ou apenas aplicações desses mesmos estudos a populações diferentes. Esta abordagem parece-nos correcta e até bastante útil, uma vez que poderia facilitar o processo de aplicação dos métodos e a comparação de resultados, que nos daria uma visão global do comportamento dos indicadores em análise em diversas situações. Mas para que tal fosse possível vários pressupostos deveriam ser cumpridos.

Os estudos que criam a metodologia e que vão servir de base aos seguintes têm que ser robustos quanto ao tipo de amostra utilizada, às medidas de avaliação dos erros utilizadas e muito claros quanto aos passos seguidos para a elaboração do método. Tal não parece ser o caso quando olhamos para os 3 estudos acima referidos. A amostra utilizada por Todd e D’Errico (1928) e por Stevenson (1924) é proveniente de uma colecção – *The Western reserve Collection* – criada com dados de indivíduos provenientes de sala de autópsia e identificados. Ainda assim, Todd (1920) realça o facto de nem sempre os registos municipais serem de confiança, o que leva a que as idades com que os indivíduos estão identificados não sejam com 100% de certeza as idades reais, tendo mesmo sido eliminados alguns esqueletos quando as idades com que estavam identificados não correspondiam aos *standards* existentes à data para o desenvolvimento do esqueleto (Webb & Suchey, 1985). No que diz respeito à amostra utilizada por McKern e Stewart (1957), apesar de ser uma amostra com uma dimensão significativa e de ter identificações fiáveis, é apenas constituída por indivíduos do sexo masculino com idades entre os 17 e os 30 anos que foram mortos em combate durante a guerra da Coreia. Este tipo de amostra não é claramente representativa de uma distribuição real dos casos que surgem num contexto normal de Antropologia Forense, pelo que pode levar ao estabelecimento de uma distribuição *a priori* incorrecta (Konigsberg et al., 2008) o que tem como consequência que os resultados não sejam fidedignos quando o método é aplicado a uma amostra ou indivíduo proveniente de um outro contexto. Um ponto comum entre os estudos referidos é o facto de utilizarem divisões por fases que foram criadas tendo por base apenas a análise morfológica do processo de união. Este tipo de abordagem, apesar

de simplificar este passo metodológico levanta questões quanto à exactidão da posterior análise que vai ser feita, uma vez que nunca aparece associada a qualquer modo de quantificação do processo de evolução da união da epífise à diáfise. Como consequência, temos a criação e posterior utilização de fases que podem levantar dúvidas aquando da tomada de decisão entre uma fase e outra como refere Stevenson (1924) em relação à Fase 3 do seu método. A opção por este tipo de critério para a criação das fases pode também levar a um erro inter observador elevado, apesar de não nos podermos referir com certeza a este ponto pois nenhum dos estudos analisados reporta os dados referentes ao erro intra e /ou inter observador. No entanto, é conhecido que as análises morfológicas estão mais sujeitas a este tipo de erro e também muito dependentes do grau de experiência do utilizador do método.

Opta-se, por vezes, para eliminar este tipo de erro, pela escolha de um menor número de fases, o que faz com que a imprecisão da escolha de uma fase em detrimento de outra seja praticamente eliminada por completo uma vez que as fases são bastante distintas entre elas. Apesar da eliminação deste erro ser positiva, este tipo de metodologias podem levar a que sejam dados intervalos etários maiores e nem sempre muito uteis num caso de antropologia forense, em que uma diferença de poucos anos face à idade real pode ser crítica para propósitos de identificação.

Para uma melhor compreensão do que está escrito acima, encontram-se a seguir as descrições das formas de divisão utilizadas em metodologias com 3, 4 e 5 fases:

- 3 Fases (Johnston, 1961)
 - Fase 1 – Não existe união
 - Fase 2 – União Parcial
 - Fase 3 – União completa

- 4 Fases (Stevenson, 1924)
 - Fase 1 – Não existe união
 - Fase 2 – Início da união
 - Fase 3 – União recente
 - Fase 4 – União completa

- 4 Fases (Webb e Suckey, 1985)
 - Fase 1 – Não existe união nem epífise separada
 - Fase 2 – Não existe união mas existe epífise separada
 - Fase 3 – União parcial
 - Fase 4 – União completa

- 5 Fases Mckern e Stewart (1957)
 - Fase 0 – Não existe união
 - Fase 1 – Início de união
 - Fase 2 – União activa
 - Fase 3 – União recente
 - Fase 4 – União completa

- 5 Fases (MacLaughlin, 1990)
 - Fase 1 – Distintas cristas onduladas e sulcos presentes
 - Fase 2 – Cristas e sulcos presentes mas não muito distintos.
 Início do alisamento da superfície epifisial
 - Fase 3 – Início da união epifisial
 - Fase 4 – União já bastante distinta mas ainda com a linha de união visível
 - Fase 5 – União epifisial completa

Outro ponto crucial para que um método seja reconhecido é a possibilidade de permitir que seja testado por vários autores em várias populações (Cunha et al., 2009) de forma a que os resultados possam ser comparados e assim melhor discutidos. No entanto, a comparação entre os diversos estudos torna-se difícil pela ausência de uma definição de certos parâmetros, como por exemplo, em que idade é que se considera que ocorre a união epifisial. Enquanto Stevenson (1924) diz que é “ a idade em que a transição entre estes dois grandes estádios de nenhuma união e união completa acontece que determina a verdadeira idade de união da epífise em questão”, Flecker (1942) diz que podemos referir à idade de união quando 50% ou mais da amostra mostrar o processo de união completo.

Outro ponto que é de notar nestes estudos é a falta de esclarecimento no que diz respeito aos intervalos etários correspondentes às diversas fases de união, uma vez que não são esclarecidos como é que esses intervalos etários são criados, se é de acordo com a idade mínima e máxima em que é observada essa fase, ou se é de acordo com outro qualquer critério (Cardoso, 2008).

Do ponto de vista das necessidades provenientes dos objectivos da Antropologia Forense existe, como já referimos e esta revisão ajudou a clarificar, uma clara necessidade de *standards* que possam ser seguidos para a criação e aplicação de métodos de estimativa da idade à morte pela união da epífise medial à diáfise da clavícula em particular e para a estimativa da idade em geral. Para além disso, são bastante importantes estudos de avaliação de várias metodologias como o de Baccino et al. (1999) ou de revisão de literatura e estado da arte como o de Cunha et al. (2009). É também importante que estes *standards* informem sobre as estatísticas mais correctas para cada tipo de dados. Trabalhos como o de Konigsberg et al. (2008) e o de Aykroyd et al. (1996), são sem dúvida importantes para melhor compreender a estatística da estimativa da idade à morte que é, como já referimos, uma das principais fontes de erro ligada a estas metodologias (Ritz-Timme et al., 2000).

1.6 – Objectivos

Este estudo propõe-se a apresentar uma nova metodologia de estimativa da idade à morte pela análise da união da epífise medial à diáfise da clavícula. Para tal, foi desenvolvido um protocolo para a recolha dos dados que visa conseguir uma quantificação rigorosa da evolução do processo de união em estudo.

Através da aplicação de diversas abordagens estatísticas pretendemos também dar resposta a que tipo de metodologia apresenta melhores resultados para este tipo de dados, criando também uma bateria consistente de medidas de avaliação do erro da metodologia e da sua capacidade de reprodutibilidade.

Pretendemos, também, através da aplicação de uma abordagem probabilística ao tratamento dos dados ir ao encontro das *guidelines* apresentadas no caso Daubert (1993) trazendo assim uma maior capacidade de resposta da Antropologia Forense a um processo judicial.

Uma vez que este estudo visa apresentar algumas alternativas no que diz respeito às abordagens estatísticas a aplicar, com o intuito de testar quais as mais adequadas quanto aos resultados que cada uma delas produz e de que forma nos podem ajudar a eliminar alguns erros presentes neste tipo de trabalhos, o capítulo seguinte visa fazer uma curta exposição da importância da estatística nas ciências forenses e da evolução matemática das abordagens aplicadas aos problemas de estimativa da idade à morte.

Capítulo 2 – A Estatística e a Antropologia Forense

“Statistics may be defined as a body of methods for making wise decisions in the face of uncertainty”

W. A. Wallis

2.1 – A Importância da Estatística nas Ciências Forenses e na Estimativa da Idade à Morte

Em Janeiro de 2003, Timucin ‘O’, um jovem de nacionalidade Turca imigrante na Alemanha – Berlin – entrou num centro de fitness feminino e matou a tiro três membros do *staff* do centro. Timucin ‘O’ foi rapidamente detido e alegou que tinha sido contractado por um homem que lhe tinha pago para matar a esposa que se ia encontrar nesse dia a essa hora nesse ginásio. Alegou também que as outras duas vítimas foram mortas com o intuito de não deixar testemunhas (Boyes, 2003).

Quando em Julho do mesmo ano foi levado a tribunal, foi levantada uma questão relativamente à idade do alegado homicida. Timucin ‘O’ tinha mais ou menos de 21 anos à data dos assassinatos? O facto de ter mais ou menos de 21 anos era a diferença entre ser julgado como um adulto ou como menor, era a diferença entre prisão perpétua ou pena máxima de 10 anos. A defesa argumentava que era comum os imigrantes criarem documentos falsos que lhes aumentassem a idade para enganar as autoridades e que os verdadeiros registos na Turquia eram já impossíveis de obter. O arguido tinha nos seus documentos uma data de nascimento de 26 de Janeiro de 1981, que implicava que à data dos homicídios já tinha mais que 21 anos, mas alegava que a data de nascimento real era 26 de Janeiro de 1982, tendo assim menos de 21 anos à data do crime. A questão que se coloca neste caso ao perito forense é se o indivíduo em causa tem mais ou menos de 21 anos, uma questão que implica um grau de precisão muito grande na sua resposta. A incerteza associada aos indicadores de idade de Timucin ‘O’ ia-se traduzir numa incerteza da decisão judicial. É precisamente na quantificação deste grau de incerteza que a estatística pode dar uma ajuda imprescindível às ciências forenses.

O tipo de casos de Timucin ‘O’ é apenas um exemplo em que é pedido a um perito forense que dê o seu testemunho relativamente à estimativa da idade de um indivíduo, e em que invariavelmente são exigidos graus de assertividade muito grandes uma vez que uma estimativa tal como, “tem entre os 20 e os 30 anos” não tem qualquer tipo de utilidade nestes casos. São exemplos de casos deste género a participação de atletas em classes desportivas inferiores, validação da idade em crianças adoptadas, falsificação da idade para uma educação favorecida, falsificação da idade para pedidos de reformas antes do tempo ou ainda casos com uma gravidade maior como abuso sexual de menores, trabalho forçado de menores ou mesmo a utilização das chamadas “crianças soldado” (Lucy, 2010).

Chamar um perito a prestar esclarecimentos quanto a uma determinada questão e aceitar esses esclarecimentos feitos com base em nada mais do que o *curriculum* do mesmo, torna o sistema bastante permeável a erros de peritagem que até o melhor dos peritos em qualquer área pode cometer ou a manipulações de decisões por agentes que podem ter parte interessada em que o juiz faça determinada interpretação da prova, interpretação essa que está a ser auxiliada pelo perito. É por isso importante que a peritagem seja feita com base em observações que sigam um método que permita que, caso necessário, a mesma observação feita por outro perito leve à mesma conclusão. O caso de *Daubert vs Merrill Dow Pharmaceutical, Inc.* (1993) em muito contribuiu para que os métodos empregues pelos peritos para analisar provas fossem revistos, alterados ou pelo menos repensados. Neste caso foram criadas *guidelines* que o conteúdo de um testemunho deve seguir:

1. Ter sido testado e ser passível de teste no futuro de acordo com o método científico
2. Ter sido sujeito a revisão pelos pares
3. Ter *standards* conhecidos
4. Ter uma taxa de erro conhecida
5. Ser aceite pela comunidade científica relevante

No caso específico das peritagens em antropologia forense sobre a estimativa da idade, pode ser então necessário que o antropólogo seja capaz de informar com grande precisão a idade de um indivíduo de identidade desconhecida. Como já foi referido é aqui que a estatística pode fazer a diferença, tanto na estimativa da idade no vivo como no morto, uma vez que de um ponto de vista estatístico não há qualquer tipo de diferença. Uma correcta aplicação da estatística pode ser a diferença entre ser apenas possível dizer “o indivíduo tem

entre 20 e 30 anos” e “ existe uma probabilidade de x de o indivíduo ter mais, ou menos, de 21 anos”. Para melhor compreender a forma como a estatística nos pode auxiliar neste tipo de situações é importante conhecer as várias metodologias estatísticas que foram empregues ao longo dos anos para a estimativa da idade à morte e também de que forma podemos criar medidas de precisão na avaliação feita pelo perito.

Timucin ‘O’ foi julgado como adulto, declarado culpado e condenado a prisão perpétua.

2.2 – Evolução Matemática dos Métodos Estatísticos

Como foi referido no tópico dedicado às fontes de erro na estimativa da idade à morte, a idade cronológica de um indivíduo – idade que se pretende estimar – não pode ser obtida directamente quando a identidade é desconhecida. A solução encontrada para este problema passa por identificar indicadores ósseos e/ou dentários que tenham uma boa correlação com a idade cronológica, medir as alterações que vão ocorrendo “com o tempo” nesses indicadores utilizando uma escala predefinida e, por fim, utilizando uma população com idade conhecida, calibrar esses indicadores, isto é, atribuir-lhes uma escala que nos informe sobre a idade cronológica. No primeiro passo, a única restrição que existe prende-se com a relação que o indicador tem com a idade; este tem que manter uma evolução unidireccional ao longo do tempo. Quer isto dizer que, utilizando o exemplo de uma estimativa feita através do crescimento de um osso longo, o crescimento do osso só pode ser utilizado para estimar a idade porque se mantém sempre em crescimento ao longo do tempo, em momento algum o osso vai diminuir de tamanho. A figura 2.1 ilustra uma regressão quadrática a uma hipotética população em que esta relação biunívoca entre indicador e idade não existe. Neste caso seria possível atribuir duas idades completamente distintas ao indivíduo que apresentasse um valor 3 para o hipotético indicador, o que é inaceitável dada a natureza do nosso problema.

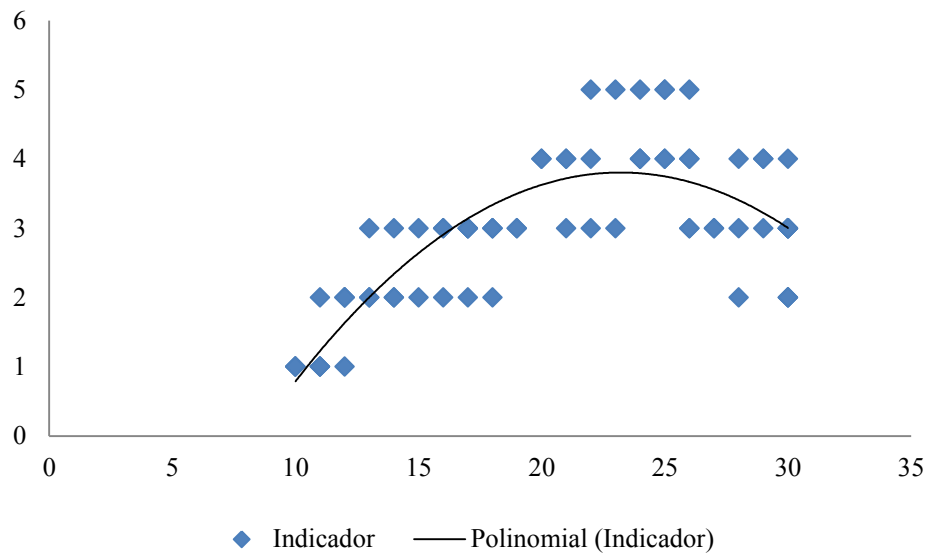


Figura 2.1 – Regressão quadrática: exemplo de uma relação inaceitável.

Neste tópico, iremos fazer uma descrição de algumas das metodologias mais utilizadas pelos antropólogos para abordar o problema da estimativa da idade à morte, nomeadamente a regressão linear, a calibração clássica e metodologias não paramétricas com abordagem Bayesiana, tanto na sua versão discreta e contínua. Iremos também introduzir uma nova forma de tratar os dados neste tipo de problemas chamada “Regularização local sem e com distribuição marginal conhecida”, baseada em Jacob e Oliveira (2011;2012). Sempre que seja necessário, para uma melhor compreensão da exposição teórica referente aos métodos, serão apresentados gráficos ou tabelas que são elaboradas com base nos dados por nós recolhidos. Outras metodologias estatísticas podiam ser aqui mencionadas, como a regressão múltipla (Maples, 1978), o vizinho mais próximo (Ferembach et al., 1980) ou a máxima verosimilhança (Konigsberg & Frankenberg, 1992). No entanto optamos por aprofundar mais as metodologias acima mencionadas uma vez que é através delas que iremos criar os vários modelos recorrendo aos dados por nós recolhidos, sendo o objectivo uma comparação final da regressão linear, da calibração clássica e da regularização local sem e com distribuição marginal conhecida.

2.2.1 – Regressão Linear

Após o período em que a estimativa da idade à morte era feita através da comparação de determinados indicadores ósseos com descrições morfológicas de várias categorias que representavam o processo de desenvolvimento ou maturação do indicador em causa (Dwight, 1878; Parsons & Box, 1905), a regressão linear tornou-se no método favorito para os antropólogos neste tipo de problema.

A regressão linear é uma técnica com um campo de utilização bastante amplo, muito possivelmente pelo facto de a ideia de adaptar uma linha a um conjunto de dados para através de uma variável (exploratória) conseguir prever o valor da outra (variável resposta), ser facilmente compreensível. Genericamente, podemos dizer que a análise por regressão é uma forma de estabelecer uma relação matemática, através de uma equação, entre variáveis que estarão de alguma forma relacionadas. A forma mais simples que a regressão pode tomar é quando em análise estão apenas duas variáveis e a relação entre elas é linear (Regressão Linear Simples). Aplicando ao nosso problema, a relação entre a variável exploratória (indicador ósseo) e variável resposta (idade), é estabelecida pela equação:

$$y = a + bx + \varepsilon \quad (2.1)$$

onde,

x é qualquer valor correspondente à variável x (indicador)

y é o valor correspondente à variável y (idade)

a é o valor tomado por y quando $x = 0$

b é a taxa de crescimento de y com x (declive da recta de regressão)

ε é o erro aleatório tomado inteiramente na direcção de y (diferença entre idade real e estimada)

Esta equação traduz-se graficamente (Figura 2.2) numa recta com a seguinte forma (hipotética):

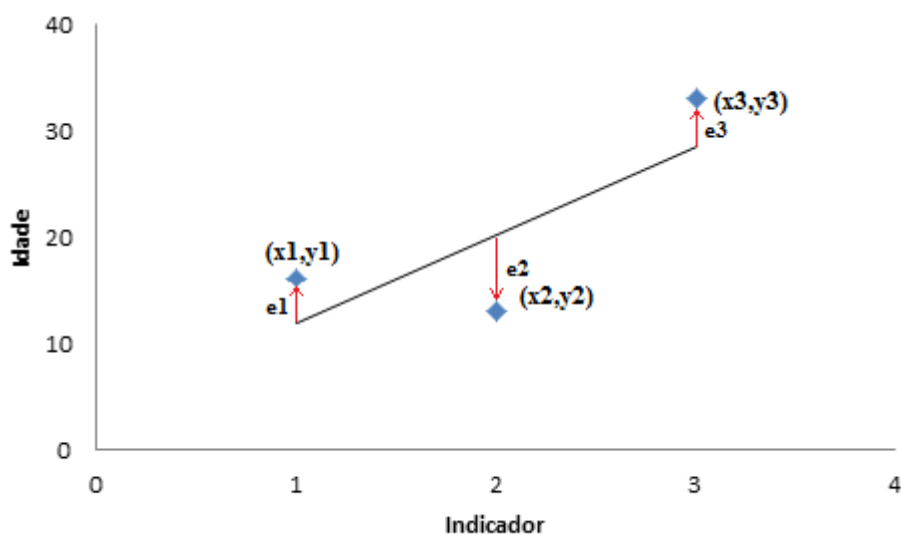


Figura 2.2 – Função de regressão com 3 hipotéticos pontos e respectivos erros assinalados

O que acontece agora é que em vez de termos apenas uma variável para estimar (y - a idade) temos também 2 parâmetros que não são conhecidos (a e b), o que faz parecer com que em vez de facilitarmos a nossa vida, ainda a dificultamos mais. Obviamente isso não é verdade. De facto, os valores de a e b vão ser facilmente estimados através dos nossos dados por um método chamado *mínimos quadrados*. Na figura 2.2 marcámos 3 hipotéticos pontos (x_1, y_1) , (x_2, y_2) e (x_3, y_3) que correspondem a 3 indivíduos com idade (y) e indicador (x) de valor conhecido. Na figura estão também assinalados os erros (e_1 , e_2 e e_3), que são a distância vertical entre o valor real (a azul) e o estimado (respectivo ponto na recta), ou seja, por erro entende-se a diferença entre a idade real do indivíduo (y_1 , y_2 e y_3) e o valor da idade estimada (valor de y na recta com igual abcissa). O objectivo da regressão é então o de encontrar uma recta $y = a + bx$ para a qual a soma dos quadrados dos erros seja minimizada.

Os erros têm que ser elevados ao quadrado devido ao facto de termos distâncias positivas (e_1 e e_3) e outras negativas (e_2) que se iriam anular quando feita a média total.¹

¹ Suponhamos que os valores das distâncias e_1 , e_2 e e_3 são respectivamente 1, -2 e 2. Se apenas somássemos os valores dos erros o resultado seria $E = 1 + (-2) + 2 = 0$. Esta situação daria a ideia de que a nossa recta foi ajustada sem qualquer tipo de erro, o que é falso. Elevando ao quadrado todas as distâncias eliminamos o sinal das mesmas, uma vez que um número negativo ao quadrado é sempre positivo. Daqui: $E = (1)^2 + (-2)^2 + (2)^2 = 1 + 4 + 4 = 9$, que é a distância total entre os pontos reais e os estimados da recta com igual abcissa.

Matematicamente podemos descrever a minimização da soma dos quadrados dos erros da seguinte forma:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.2)$$

Assim, o que nós pretendemos, é encontrar os valores de a e b que minimizem a média dos quadrados das distâncias entre os pontos da distribuição dos nossos dados e os correspondentes na recta com igual abcissa. Os parâmetros estimados \hat{a} e \hat{b} têm a seguinte forma:

$$\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} \quad (2.3)$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad (2.4)$$

Daqui, vem que o valor estimado da idade correspondente a um valor observado de x do indicador, vai ser:

$$\hat{y} = \hat{a} + \hat{b}\bar{x} \quad (2.5)$$

À medida que os estudos de estimativa da idade à morte com base em equações de regressão linear proliferavam, um padrão começou a surgir. O que Masset (1989) apelidou de “a atracção pelo centro” foi um erro que vários outros autores reportaram (Bedford et al., 1993; Albrecht, 1980; Saunders, 1992). O que estava a acontecer era que os indivíduos mais novos apareciam sistematicamente com a idade sobrestimada e os indivíduos mais velhos com a idade subestimada. Masset tentou resolver o problema utilizando uma amostra de referência com uma estrutura de idades mais variada, uma vez que ele acreditava que o facto de se utilizarem amostras enviesadas era o que causava este erro sistemático. Mas tal abordagem não resolveu o problema. De facto, a existência desse viés sistemático é facilmente observável através da comparação dos valores obtidos pela nossa equação com os valores da idade real (Tabela B2 – Apêndice B).

Como foi por nós apontado anteriormente, não existe nenhum indicador que tenha uma correlação perfeita com a idade cronológica, e é aqui que reside a razão deste erro sistemático. A correlação entre um indicador e a idade cronológica é medida pelo coeficiente de correlação de Pearson, que é representado pela letra minúscula “ r ”. Vejamos então de que forma, este coeficiente e o viés sistemático encontrado neste tipo de abordagem estão relacionados.

Sendo o erro que está subjacente à estimativa de cada valor da idade, a diferença entre a idade verdadeira e a estimada temos:

$$e_i = y_i - \hat{y}_i \quad (2.6)$$

de (2.1) e (2.3) temos que:

$$\begin{aligned} e_i &= y_i - (\bar{y} - \hat{b}\bar{x} + \hat{b}x_i) \\ &= y_i - \bar{y} - \hat{b}(x_i - \bar{x}), \end{aligned}$$

e de (2.4)

$$e_i = y_i - \bar{y} - \frac{s_{xy}}{s_{xx}} (x_i - \bar{x})$$

Ao efectuarmos uma regressão dos erros obtidos sobre a idade vamos conseguir ver a relação entre o erro e a idade. Sendo $\hat{e} = \hat{c} + \hat{d}y_i$ a equação da recta da regressão mencionada, o parâmetro que aqui nos interessa é o \hat{d} que é o declive da recta. Resolvendo \hat{d} a partir de:

$$\hat{d} = \frac{\sum(y_i - \bar{y}) - (e_i - \bar{e})}{\sum(y_i - \bar{y})^2} \quad (2.7)$$

vamos obter

$$\hat{d} = 1 - r_{xy}^2 \quad (2.8)$$

onde r_{xy} é o coeficiente de correlação de Pearson entre as variáveis x (indicador) e y (idade). Esta é a relação existente entre a correlação e as variáveis indicadores e idade, mas falta ainda

mostrar de que forma é que esta relação faz com que indivíduos mais novos sejam sistematicamente sobrestimados e os mais velhos subestimados.

Relembrando o facto de que o coeficiente de correlação de Pearson varia sempre entre -1 e 1, uma análise atenta da equação (2.7) facilmente permite compreender que o declive da recta de regressão entre o erro e a idade real vai sempre ser positivo, ou seja, varia apenas entre 0 e 1. Isto porque se a correlação for perfeita $r^2=1$, a recta da regressão dos erros vai ter um declive nulo. Por outro lado, se a correlação for nula, o declive vai ser 1. A relação que aqui se estabelece é que quanto menor for a correlação entre as variáveis maior vai ser o declive e *vice-versa*. Assim sendo, podemos retirar daqui que o declive (também por esta razão chamado de viés) pode ser uma forma de avaliar a precisão de um modelo de regressão, uma vez que nos permite perceber se estamos a estimar ou não as idades por excesso ou por defeito. É esta a razão pela qual os indivíduos mais novos são estimados com idades superiores à verdadeira e os mais velhos com idades inferiores à real. A Figura 2.3 mostra a regressão dos erros sobre a idade verdadeira com equação:

$$\text{Erro} = 0,3059 \times \text{Idade Real} - 6,607$$

Nela é possível confirmar o que foi anteriormente dito. Uma vez que o erro é a idade verdadeira menos a estimada, caso este seja negativo significa que a idade real é inferior à estimada (sobrestimação). No caso do erro ter um valor positivo significa que estamos perante uma idade real superior à estimada (subestimação). A análise da figura permite observar que a sobrestimação ocorre nos indivíduos mais novos (<20 anos, aproximadamente) e a subestimação nos indivíduos mais velhos (>20 anos, aproximadamente).

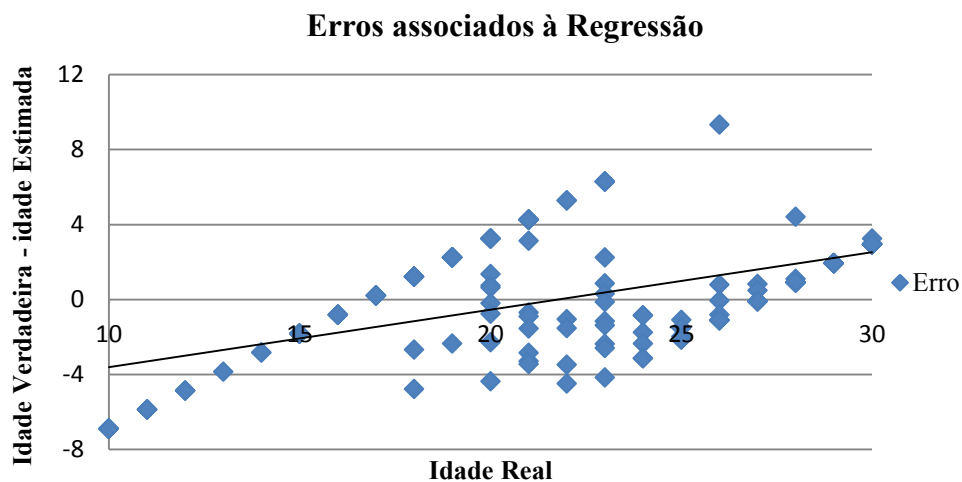


Figura 2.3. – Regressão dos Erros sobre a Idade Real

Este erro é uma clara desvantagem deste método na resposta ao problema da estimativa da idade à morte. Para além disto acresce o facto de a regressão linear ter sido aplicada a dados para os quais não é de todo a metodologia mais indicada. A regressão linear tem um número de pressupostos inerentes ao tipo de dados a que pode ser aplicado (Lucy et al., 1996):

1. As variáveis têm que dar informação independente sobre a idade, isto é, é esperado que elas estejam correlacionadas entre si uma vez que variam com a idade mas não podem estar fisicamente dependentes uma(s) da(s) outra(s);
2. As variáveis têm que variar continuamente com a idade. Uma variável continua pode adoptar virtualmente um número infinito de valores, contrariamente a uma variável discreta que apenas adopta um número finito de valores, normalmente designados fases ou estádios;
3. A distribuição dos erros em torno da média tem que estar normalmente distribuída. Quer isto dizer que qualquer estimativa do erro de uma idade desconhecida vai seguir uma distribuição normal acerca do seu valor preditivo. Este pressuposto pode ser um sério entrave à utilização desta abordagem estatística uma vez que durante a análise dos dados dificilmente vai ser cumprido;
4. As variáveis analisadas têm que se alterar de forma linear com a idade;
5. Quando são utilizados somatórios de vários indicadores é assumido que todos os indicadores contribuem com o mesmo peso informativo acerca da idade.

Uma breve revisão da literatura permite observar que nem sempre estes pressupostos foram cumpridos em métodos que aplicaram regressões lineares. Johanson (1971) utilizou sete fases no seu modelo de estimativa de idade à morte por indicadores dentários que não podem ser considerados contínuos, mas sim ordinais. O método de Gustafson (1950) não cumpre o 5º pressuposto, uma vez que as variáveis por ele analisadas não contribuem todas com o mesmo peso informativo acerca da idade. A aplicação desta metodologia a dados categoriais pode também ser uma fonte dos erros encontrados pelos autores nos métodos que desenvolvem, como é o caso de Gustafson (1950) que utilizava uma escala ordinal de 0 a 3 para avaliar as alterações verificadas nos 6 indicadores dentários que analisou.

Dados os problemas associados com a regressão linear, foram sugeridas alternativas à metodologia. Konigsberg et al. (1994) e Lucy & Pollard (1995) recomendaram uma variação

da regressão linear chamada Calibração Clássica como uma técnica mais adequada à estimativa da idade à morte através de indicadores ósseos e/ou dentários.

As fórmulas (2.3) e (2.7) têm o seu desenvolvimento explicado no Apêndice C.

2.2.2 – Calibração Clássica

Este método é praticamente igual ao da regressão linear, com excepção de um único detalhe. A estimativa é feita para x e não para y como na regressão linear, passando o y a ser a variável observada (indicador) e o x a variável que pretendemos determinar (idade). Assim, a equação apresentada em (2.1) passa a ler-se da seguinte forma:

$$\text{Indicador } (y) = a + b \cdot \text{Idade } (x)$$

Uma vez que na prática a variável que continuamos a observar é o Indicador, o que temos que fazer é resolver a equação anterior em ordem à Idade (x). Ficando:

$$\hat{x}_i = \frac{y_i - \hat{a}}{\hat{b}} \quad (2.8)$$

Esta abordagem faz sentido não só de um ponto de vista estatístico mas também de um ponto de vista biológico. Como referiu Schmitt, (2001):

“Biologicamente esta forma de processar a estimação da idade é mais lógica. A idade deve ser considerada como a variável fixa e a observação das mudanças devidas à idade como a variável dependente. É a idade que induz mudanças biológicas e não o contrário”

Vamos então ver de que forma esta abordagem nos ajuda na resolução do problema do viés criado pela regressão linear. Após utilizarmos a equação (2.8) para estimar a idade da nossa amostra e depois de calcularmos os respectivos erros associados com cada estimativa (Tabela B3 – Apêndice B), efectuamos uma regressão dos resíduos sobre a idade (Figura 2.4)

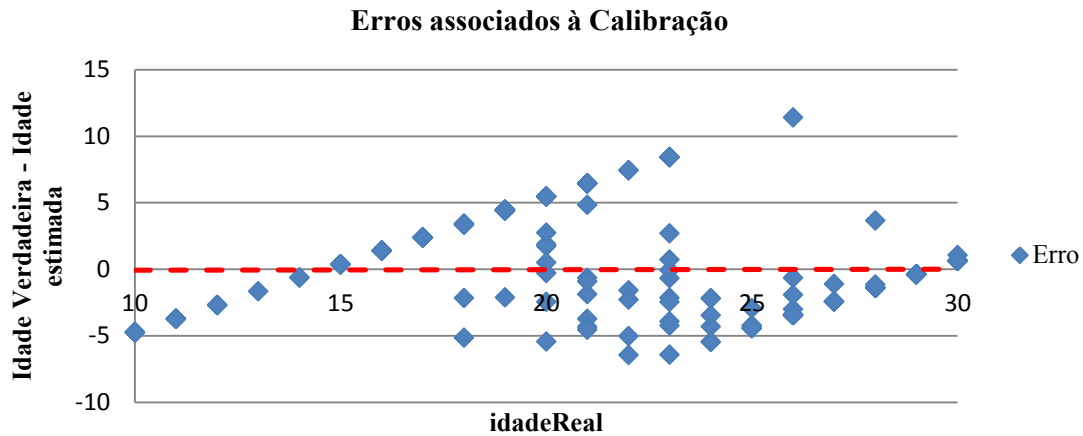


Figura 2.4 – Regressão dos Erros sobre a Idade Real

Como é facilmente observável, o viés foi completamente eliminado, ficando a recta da regressão (a vermelho) coincidente com o eixo das abcissas. Isto quer dizer que o erro está distribuído simetricamente em torno da idade real, ou seja, não existe sobrestimação dos indivíduos mais novos nem subestimação dos indivíduos mais velhos. Como foi referido, na regressão linear o erro ficava todo na direcção da idade, quer isto dizer que o método dos mínimos quadrados ajusta sempre a equação do sentido da idade, assumindo que as medições do indicador são feitas sem qualquer erro. Mas esta abordagem também tem algumas desvantagens. Nomeadamente, existe uma redução na eficiência do método (Konigsberg et al.,1994) que leva a que a variabilidade seja maior na calibração clássica que na regressão linear, ou seja, a diferença entre as idades estimadas e as idades reais é maior nesta abordagem, aumentando assim o erro global da estimação (Lucy, 1997).

Devido aos notórios problemas que as abordagens anteriormente referidas acarretam, formas alternativas de estimar a idade à morte que cortassem por completo com o que se vinha fazendo até então eram necessárias. É assim que surge a abordagem Bayesiana (Konigsberg & Frankenberg, 1992; Lucy et al., 1996). Esta abordagem não é completamente nova, tendo já sido utilizada na literatura sobre pescas com o nome de “*age length key*”, mas começa a ganhar grande notoriedade e melhorias significativas principalmente devido a um grande aumento da capacidade de processamento dos computadores necessários ao cálculo das distribuições *a posteriori* (Martins et al., 2012)

2.2.3 – Abordagem Bayesiana

Esta abordagem vem trazer à literatura antropológica conceitos como probabilidade *a priori*, verosimilhança e probabilidade *a posteriori*, até aí estranhos à mesma, e que importa definir.

A probabilidade *a priori* é a probabilidade atribuída a uma hipótese antes de efectuado qualquer teste ou experiência; esta advém apenas do conhecimento que temos por a mesma experiência ter sido efectuada em ocasiões anteriores. Na estimativa da idade à morte é a probabilidade de um indivíduo ter uma determinada idade sem haver qualquer informação referente a qualquer indicador ósseo e/ou dentário (Martins et al., 2012; Lucy et al., 1996). Em notação matemática esta probabilidade é representada por $P(A_i)$, ou seja, a probabilidade de um indivíduo ter uma idade A que pertence ao grupo etário i .

A verosimilhança é a probabilidade condicional de determinada hipótese ser verdadeira condicionada à amostra observada, ou seja, a probabilidade de o indivíduo ter determinado valor no indicador observado sabendo que tem uma determinada idade, baseado na distribuição da idade da amostra para esse valor do indicador (Martins, 2005). Matematicamente representa-se por $P(I|A_i)$ e lê-se a probabilidade de apresentar o indicador I dado que tem a idade A_i .

A probabilidade *a posteriori* é a probabilidade condicional da hipótese ser verdadeira dado o valor da observação realizada; escreve-se $P(A_i|I)$. No nosso caso, é a probabilidade do indivíduo pertencer ao grupo etário A_i dado que tem um valor I para o indicador observado.

O teorema de Bayes diz-nos que a probabilidade *a posteriori* é proporcional à probabilidade *a priori* multiplicado pela verosimilhança, ou seja, é a probabilidade de um indivíduo pertencer a um determinado grupo etário tendo em conta a probabilidade *a priori* e a verosimilhança. Isto traduz-se na expressão para a versão discreta do teorema:

$$P(A_i|I^j) = \frac{P(I^j|A_i)P(A_i)}{P(I^j)} \quad (2.9)$$

onde,

$$P(I^j) = \sum_{i=1}^n P(I^j|A_i)P(A_i) \quad (2.10)$$

e,

$$P(I^j|A_i) = \frac{n(I^j, A_i)}{n(A_i)} \quad (2.11)$$

sendo,

$$P(A_i) = \frac{n(A_i)}{N} \quad (2.12)$$

Na notação utilizada, a equação (2.9) dá-nos a probabilidade *a posteriori*, a (2.11) a verosimilhança e a (2.12) a probabilidade *a priori*; para todas as equações j refere-se ao conjunto de todos os i 's sendo i cada um dos grupos etários. As várias probabilidades necessárias ao cálculo de (2.9) podem facilmente ser encontradas através da tabela com a distribuição do número de indivíduos por indicador e idade.

A tabela 2.2 é a tabela da distribuição dos nossos dados, através dela iremos calcular uma estimativa para exemplificar como se procede ao cálculo das várias probabilidades acima expostas. Vamos supor que temos para estudo um indivíduo para o qual queremos estimar a idade. Após análise da epífise medial da clavícula, verificamos que esta tem uma ossificação que ocupa entre 50% e 75% da área total da epífise. Sabendo isto, queremos saber qual a idade do indivíduo em causa, ou seja, qual a $P(A|I)$, sabendo que $I = \text{Fase 3 de união}$. Assim, (2.12) vai ser: $P(A_1) = 34/157$; $P(A_2) = 62/157$; $P(A_3) = 30/157$; $P(A_4) = 31/157$, ou seja, a probabilidade de um indivíduo pertencer a um determinado grupo etário é igual ao número de indivíduos nesse grupo a dividir pelo total da amostra. Sendo que a equação (2.11) se obtém pela divisão do número de indivíduos em cada grupo etário dentro da fase 3 pelo total de indivíduos do respectivo grupo etário, ou seja, $P(I|A_1) = 0/34$; $P(I|A_2) = 17/62$; $P(I|A_3) = 5/30$; $P(I|A_4) = 1/31$; torna-se simples obter o numerador para a equação (2.9), basta para isso multiplicar o resultado obtido de (2.12) com o de (2.11). Por exemplo, se quisermos saber qual a probabilidade do indivíduo em causa ter entre 18 e 23 anos, basta multiplicar $P(I|A_2) \times P(A_2) = (17/62) \times (62/157) = 0,108$. De seguida repetimos este cálculo para todos os grupos etários e efectuamos o somatório dos resultados, o que originará a equação (2.10) e que neste caso tem como resultado 0,146497. Para obter a probabilidade do indivíduo ter entre 18 e 23 anos sabendo que se encontra na fase de ossificação 3, dividimos $P(I|A_2) \times P(A_2)$ por $P(I)$, ou seja, $P(A_2|I) = (0,108/0,146497) \times 100\% = 74\%$. E está calculada a equação (2.9) e a pretendida probabilidade *a posteriori*. A tabela com todas as probabilidades

a posteriori possíveis por grupo etário dada a fase encontra-se nos resultados, no tópico referente aos resultados do modelo Bayesiano discreto.

Tabela 2.2

Distribuição dos indivíduos da amostra pelas Fases de União epifisial e Grupos Etários

Grupo Etário	Fase				Ni
	F1	F2	F3	F4	
	0%	>0%;<50%	>50%;<75%	>75%	
G1 [10;17]	34	0	0	0	34
G2 [18;23]	32	11	17	2	62
G3 [24;27]	0	0	5	25	30
G4 [28;30]	0	0	1	30	31
Nf	66	11	23	57	157

Nota. Ni - Total de indivíduos por classe etária; Nf - Total de indivíduos por fase

Apesar de ser extremamente simples de compreender e de utilizar, este modelo tem, pelo menos, duas desvantagens. Primeiro o facto de as probabilidades *a posteriori* serem muito dependentes dos intervalos etários criados; basta aumentarmos ou diminuirmos ligeiramente alguns ou todos os intervalos, para que a distribuição do número de indivíduos por fase e grupo etário se altere; consequentemente, todas as probabilidades calculadas, com as equações de 2.9 a 2.12, seriam diferentes. Segundo, ser bastante sensível à existência de zeros nas observações, isto é, para permitir estimativas de confiança necessita de amostras muito grandes.

Para exemplificar o que está a ser referido imaginemos que os nossos grupos etários em vez dos expostos na Tabela 2.2 eram: G1 – [11;18]; G2 – [19;24]; G3 – [25;28] e G4 – [29;30]. A diferença em relação ao actual seria de apenas 1 ano, uma vez que a idade 10 anos não era incluída. No entanto, vejamos a diferença no cálculo da probabilidade *a posteriori* para o exemplo acima referido. Para a Fase 3 a distribuição do número de indivíduos por grupo etário seria G1 – 1; G2 – 19; G3 – 3; G4 – 0, e para o grupo etário total seria G1 – 35; G2 – 61; G3 – 33; G4 – 22. Desta forma, e repetindo todos os cálculos acima já explicados, iríamos obter uma $P(A_2|I) = 83\%$ que é, como podemos verificar, quase 10% superior à

obtida anteriormente e apenas eliminamos uma idade da nossa amostra. Interessa por isto, encontrar uma forma de calcular a probabilidade *a posteriori* que não esteja tão dependente dos intervalos criados. Tal solução pode passar por tratarmos a idade como uma variável contínua (que é) em vez de a agruparmos em categorias discretas.

Ao efectuarmos este procedimento, temos que adaptar a forma do teorema de Bayes para os chamados casos mistos, isto é, em que uma variável é contínua (idade) e a outra discreta (indicador). Assim, seja A uma variável aleatória contínua (idade) e I uma variável aleatória discreta (indicador) para um indivíduo escolhido aleatoriamente, a forma correcta do Teorema de Bayes será:

$$f(A|I) = \frac{f(I|A)f(A)}{f(I)} \quad (2.13)$$

onde temos que $f(I|A)$ é a verosimilhança dos indicadores dada a idade, $f(I)$ é a função de probabilidade de I , $f(A|I)$ é a função de probabilidade condicionada de A dado I , a que chamaremos função densidade de probabilidade *a posteriori* e $f(A)$ é a função densidade de probabilidade de A a que chamaremos função de densidade de probabilidade *a priori*.

Comparando a equação (2.13) com a (2.9) verificamos que apenas têm uma diferença; enquanto a (2.9) apenas implica o conhecimento e cálculo de algumas probabilidades conhecidas através da tabela de distribuição dos dados, a (2.13) implica conhecer as várias funções de probabilidade e de densidade que aparecem na equação. A forma como se procede a esse tipo de cálculos e as vantagens que esta abordagem nos assegura será exposta no tópico seguinte.

2.2.3.1 – Estimação pelo método dos núcleos

A estimação da função densidade de probabilidade de um conjunto de dados é um dos mais vulgares problemas para quem lida frequentemente com estatística (Ryan, 1996). Por função de densidade de probabilidade entende-se uma função algébrica que, dados determinados valores para os parâmetros, indica frequências relativas para todos os eventos x no espaço amostral (Lynch, 2007), isto é, são funções que nos informam acerca da localização dos eventos mais (e menos) prováveis num espaço amostral. Quando traduzidas

para um gráfico, estas funções adoptam formas e escalas que são definidas pelos parâmetros nelas contidas. Um exemplo conhecido de uma função de densidade de probabilidade é a da distribuição Gaussiana (também conhecida como distribuição Normal) cuja distribuição dos dados é dada pela função de densidade de probabilidade (2.14) que resulta no conhecido gráfico com a “forma de sino” (Figura 2.5). Neste caso, apenas temos que estimar os parâmetros $\theta = (\mu, \sigma)$, respectivamente a média e a variância.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (2.14)$$

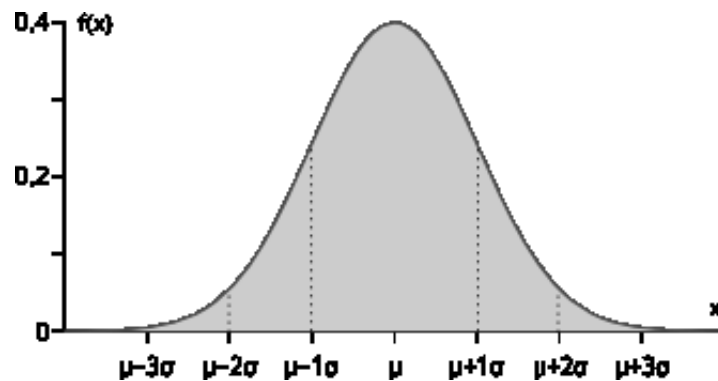


Figura 2.5 – Exemplo de um gráfico de uma Distribuição Normal

Para a estimação de funções de densidade existem duas abordagens possíveis: a paramétrica e a não paramétrica. No caso da estimação paramétrica, mais comumente utilizada, é assumido que a distribuição dos dados pertence a uma família de densidade conhecida sendo apenas necessário estimar os parâmetros correspondentes. Na abordagem não paramétrica o objectivo passa por obter uma boa estimativa para a função de densidade $f(\cdot)$ sem assumir à partida qualquer tipo de hipótese sobre a sua forma.

Um exemplo conhecido e simples de um estimador não paramétrico da densidade é o histograma. O histograma é uma representação gráfica da distribuição de frequências dos dados recolhidos da amostra; quanto maior a frequência de medições numa determinada classe, maior vai ser a barra correspondente. Este estimador é simples de criar e de ler e útil na sumarização de informação recolhida da amostra estudada. No entanto, o posterior cálculo

de probabilidades através dele pode apresentar um grande enviesamento devido ao facto de este método estar muito dependente das classes criadas, como verificamos pela abordagem discreta do Teorema de Bayes. As classes dos histogramas são os intervalos representados no eixo dos xx's e pelas quais os dados se vão distribuir, sendo no nosso caso os intervalos etários. Posto isto, como foi já referido no final do tópico anterior era necessário encontrar outro estimador não paramétrico que não estivesse tão dependente dos intervalos criados; o estimador que foi por nós utilizado para encontrar as funções de distribuição de probabilidades necessários ao cálculo da probabilidade *a posteriori* chama-se “*estimador do núcleo com função núcleo K*” (daqui para a frente será chamado estimador do núcleo).

Este tipo de estimadores apresenta pelo menos duas vantagens que nos trazem melhorias significativas aos modelos de estimativa da idade criados. Estas vantagens advêm do facto de serem aplicados aos dados, factores de regularização (traduzido do Inglês: *smoothing*) que levam a que observações adjacentes contribuam para a construção de uma estimativa (Wand & Jones, 1995). Desta forma, existe total independência das classes que eram utilizadas nos histogramas e consegue-se também informar pontos da nossa distribuição acerca dos quais não existe qualquer informação disponível. Quer esta última vantagem dizer que, este método poderá ser uma boa alternativa quando dispomos de amostras de pequenas dimensões, o que é relativamente frequente em antropologia.

O tipo de estimador que vamos utilizar para o cálculo da função de densidade da probabilidade *a priori* é baseado no proposto em 1956 por Rosenblatt e tem a forma:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K\left(\frac{x - X_i}{h_n}\right) \quad (2.15)$$

onde, K é uma função de densidade de probabilidade e h é o já referido parâmetro de regularidade.

Wand e Jones (1995) deram uma forma de pensarmos na estimação da densidade de acordo com este método. Se tivermos n pontos X_1, X_2, \dots, X_n na linha das abcissas, obtemos a densidade colocando curvas normais centradas em cada X_i e depois somamos a altura de cada uma dessas curvas assentes em cada ponto do eixo dos xx's. A forma da curva vai ser

definida pelo núcleo K e a largura, tal como a altura, dependerão do parâmetro h . Pode-se mostrar que a escolha da função K não tem grande preponderância no resultado final, uma vez que as várias funções que existem disponíveis apresentam todas resultados ótimos ou quase ótimos. Já a escolha do parâmetro de regularidade implica um maior conhecimento matemático, pelo que será apenas descrito muito resumidamente no capítulo da metodologia como é que este foi decidido.

Vai ser este tipo de abordagem que nos vai permitir criar os três modelos em relação aos quais existem maior expectativa neste estudo. Nos dois primeiros modelos, baseados em Jacob e Oliveira (2012), simplificamos a mais usual abordagem Bayesiana empregue por Konigsberg e Frankenberg (1992) e Lucy et al. (2002) ao aplicarmos o método de Regularização local sem e com distribuição marginal conhecida. No terceiro modelo pretendemos saber se é possível esta última abordagem ajudar-nos na criação de modelos de estimativa da idade à morte quando estão disponíveis amostras muito pequenas. Desta forma, o tipo de abordagem que aqui introduzimos é uma simplificação da mais comum abordagem Bayesiana para casos mistos não só no cálculo das várias funções mas também na posterior aplicação prática dos modelos. Enquanto nas metodologias propostas por Konigsberg e Frankenberg (1992) e Lucy et al. (2002) as equações referentes às funções de verosimilhança e *a priori* são calculadas independentemente, no método proposto por Jacob e Oliveira (2011;2012) estas são calculadas de forma conjunta para o modelo que inclui informação marginal e a função *a priori* nem sequer é necessária para o modelo que não inclui a informação marginal. Para além disto, uma leitura dos estudos dos autores referidos primeiramente permite perceber que para a aplicação prática dos métodos propostos é necessário um bom conhecimento de programação em “R” ou algum *software* Matemático (MatLab, por exemplo) assim como conhecimentos de estatística algo avançados. Nos modelos por nós propostos baseados em Jacob e Oliveira (2011;2012) isto também foi simplificado de forma a tornar a aplicação do método o mais simples possível e livre de qualquer cálculo estatístico. Pretendemos assim criar o que consideramos ser o compromisso ideal entre a estatística e a antropologia.

Capítulo 3 – Amostra

A amostra utilizada é composta por indivíduos provenientes de duas colecções de esqueletos identificados: A Colecção de Esqueletos Identificados do Museu Antropológico da Universidade de Coimbra (CEIMA) e a Colecção de Esqueletos Identificados do Museu Bocage, Lisboa.

A CEIMA é uma colecção composta por 505 esqueletos (266 masculinos e 239 femininos) provenientes do cemitério da Conchada (Coimbra) de indivíduos que nasceram entre os anos de 1817 e 1924 e morreram entre os anos de 1904 e 1938. Os indivíduos são todos de nacionalidade Portuguesa excepção feita a 9, provenientes de África (6), Espanha (2) e Brasil (1). A idade à morte varia entre os 7 e os 96 anos (Cunha & Wasterlain, 2007). Todos os indivíduos têm uma ficha de identificação da qual fazem parte dados como nome, sexo, idade à morte, causa da morte e local de residência. A maioria dos indivíduos presentes nesta colecção são caracterizados por pertencerem a um estatuto socioeconómico considerado baixo.

A Colecção de esqueletos identificados do Museu Bocage é constituída por indivíduos provenientes de vários cemitérios que foram inumados entre 1903 e 1975, tendo a maioria morrido entre 1930 e 1960; os anos de nascimento variam entre 1887 e 1960 (Cardoso, 2008)

No total foram utilizados 158 indivíduos e a sua distribuição por colecção, sexo e idade encontra-se na tabela 3.1, apesar de posteriormente ter sido retirado 1 indivíduo por levantar dúvidas quanto à Fase de união, pelo que os modelos foram criados com uma amostra de 157 indivíduos. Na tabela 3.2 é possível ver a amplitude dos intervalos de morte dos indivíduos por colecção, factor que, caso apresenta-se uma variação significativa poderia levar à existência de diferenças significativas entre ambas as colecções no que diz respeito ao indicador analisado neste estudo por efeitos de tendência secular. Tal diferença não se verifica, tendo sido realizado um teste t para amostras independentes a fim de verificar a hipótese da existência de diferenças estatisticamente significativas entre as médias do indicador analisado em ambas as colecções. Os resultados do teste t estão apresentados na tabela 3.3.

Tabela 3.1

Distribuição dos indivíduos da amostra por Coleção, Sexo e Idade à morte

Idade	Coleção/Sexo				Total i
	Coimbra		Lisboa		
	F	M	F	M	
10	5	1	0	0	6
11	2	1	2	0	5
12	3	0	0	0	3
13	0	0	0	1	1
14	1	0	1	0	2
15	0	1	2	2	5
16	0	3	1	1	5
17	4	1	0	2	7
18	1	0	5	1	7
19	5	3	2	2	12
20	1	3	3	4	11
21	7	3	2	1	13
22	3	1	2	0	6
23	3	5	2	3	13
24	3	2	1	0	6
25	3	1	2	1	7
26	4	7	0	0	11
27	2	3	1	1	7
28	3	4	1	1	9
29	4	4	1	0	9
30	3	5	0	5	13
10	5	1	0	0	6
11	2	1	2	0	5
12	3	0	0	0	3
13	0	0	0	1	1
14	1	0	1	0	2
Total s	57	48	28	25	158
Total c	105		53		

Nota. Total i: Total de indivíduos com a respectiva idade; Total s: Total de indivíduos do respectivo sexo; Total c: Total de indivíduos da respectiva coleção; F: feminino; M: masculino

Tabela 3.2

Distribuição dos indivíduos da amostra por Ano de morte Mínimo, Máximo e Médio e por Coleção

Coleção	N	Mínimo	Máximo	Média
Coimbra	105	1904	1936	1925
Lisboa	53	1903	1958	1939

Tabela 3.3

Teste t para as diferenças das médias dos Rácios por Coleção

Coleção	N	Média	DP	<i>t</i>	df	sig. (2 - tailed)	Diferença média
Coimbra	105	.51	.46				
Lisboa	53	.43	.43	.096	156	.256	.080

Capítulo 4 – Metodologia

Sendo um dos principais objectivos desta tese o desenvolvimento de uma nova abordagem à criação de metodologias de estimativa da idade à morte através do processo de união da extremidade medial da clavícula, este capítulo reveste-se de particular importância

Aqui serão descritos de forma exhaustiva os procedimentos de recolha dos dados e a descrição das metodologias estatísticas utilizadas na criação dos modelos. Dentro do tópico da descrição das metodologias serão explicadas as formas através das quais se procedeu a várias decisões ao longo do estudo (como a criação das fases de união e dos grupos etários, por exemplo) e também de que forma avaliamos os métodos finais e a sua capacidade de realizarem uma correcta estimativa da idade à morte a indivíduos de identidade desconhecida.

4.1 – Procedimentos de Recolha dos Dados

Os procedimentos utilizados para a recolha dos dados foram criados tendo em mente a preocupação de nos permitirem uma compreensão rigorosa da evolução do processo de união da epífise à diáfise assim como o de permitir uma quantificação deste processo, com o intuito de diminuir a subjectividade trazida das análises puramente descritivas. Todos os dados foram recolhidos de clavículas esquerdas por uma questão de consistência interna do estudo e uma vez que existem estudos que indicam não existir diferenças no tempo de união epifisial entre o lado esquerdo e o direito (Schmeling et al., 2004).

Na amostra seleccionada todas as clavículas foram fotografadas na sua extremidade medial. Para a realização das fotografias foi utilizada uma Nikon D40X com recurso a um tripé, o que permitiu um aumento da exposição da imagem sempre que existia pouca luz. Este aumento da exposição possibilitou-nos assim que não fossem utilizadas luzes artificiais (“*flash*”) para iluminar a área a ser fotografada, o que ia criar falsos contrastes ou fazer perder profundidade que seria útil na posterior análise das imagens. Desta forma, as fotografias tiradas captam a real imagem da clavícula, o que reduz os possíveis erros de análise da imagem.

Posteriormente estas fotografias foram utilizadas para efectuar duas medições: 1) Área da extremidade medial da clavícula (Ae) e 2) Área da ossificação (Ao).

As medições foram realizadas recorrendo ao *software* ImageJ. O ImageJ é um *software open source* que pode ser encontrado em <http://rsbweb.nih.gov/ij/> e que é distribuído pelo *National Institutes of Health – USA*. Este *software* permite efectuar operações de fotometria e cujos algoritmos utilizados para o cálculo de áreas irregulares foram já utilizados e revistos em vários estudos de morfometria (Macaluso Jr, 2011). As Figuras 4.1 e 4.2 dão um exemplo das medições das áreas Ae e Ao, respectivamente.

Para evitar o enviesamento nas medições que poderia advir do facto de as fotografias não serem todas tiradas precisamente à mesma distância das clavículas, foi incluída em todas as fotografias uma escala que permitiu que, antes de serem efectuadas as medições das áreas, a escala de medida utilizada pelo ImageJ fosse calibrada para a imagem em questão, fazendo também de imediato a conversão de Pixéis para Centímetros. Assim, todos os resultados referentes a áreas encontram-se nesta unidade de medida.



Figura 4.1 – Exemplo da colocação dos pontos para determinação da área da extremidade da clavícula (Ae) no *software* ImageJ



Figura 4.2 – Exemplo da colocação dos pontos para determinação da área de ossificação da clavícula (A_o) no *software* ImageJ

Efectuadas as medições ao total da amostra, as variáveis A_e e A_o foram utilizadas para a criação de uma terceira que corresponde ao Rácio (R) entre a área de ossificação e a área da extremidade, isto é:

$$R = \frac{A_o}{A_e} \quad (4.1)$$

Como já foi referido, um dos nossos objectivos passa por conseguir quantificar de forma rigorosa o processo de desenvolvimento da união da epífise. Isto apenas podia ser conseguido eliminando factores que pudessem de alguma forma influenciar o tamanho e robustez das clavículas e que não estivessem relacionados com a idade. São exemplos dos factores referidos, a robustez física, sexo, factores de desenvolvimento, essencialmente todo e qualquer factor responsável pela variabilidade inter pessoal que caracteriza os humanos. Através da utilização do Rácio entre as duas áreas a distribuição fica uniformizada no intervalo $[0;1]$, eliminando assim a maioria dos factores enviesadores que possam estar presentes uma vez que todas as clavículas vão ser caracterizadas obedecendo a uma mesma

proporção. Para melhor compreensão do que foi dito, vamos dar um exemplo hipotético. É perfeitamente possível que os indivíduos A e B tenham a mesma idade e sexo, mas que, por razões de variabilidade que não nos interessa avaliar, o indivíduo A tenha uma A_e de 2cm^2 e o B tenha uma A_e de 3cm^2 (valores hipotéticos). Sendo a área de ossificação proporcional à área da extremidade é de esperar que a ossificação do indivíduo B também seja maior que a do A, por exemplo A_o do indivíduo A = 1 cm^2 e a do B = $1,5\text{ cm}^2$. Esta diferença no modelo final ia ser atribuída à idade e o indivíduo B ia ser estimado com uma idade superior à do A se apenas tomássemos em consideração os valores das áreas. Aplicando a fórmula (4.1) a ambos os indivíduos vamos obter um rácio de 0,5 para o A e para o B, o que nos vai permitir inferir correctamente que ambos têm a mesma idade.

O valor de R será no final transformado numa percentagem de forma a possibilitar uma melhor leitura dos dados, bastando para isso que se multiplique o resultado de (4.1) por 100%. Assim, o resultado de $R \times 100\%$ deverá ler-se: “O Rácio indica que X% da extremidade medial da clavícula em análise já se encontra ossificada”.

Este Rácio vai ser a variável que nos vai permitir medir a evolução do nosso indicador. Como referimos no tópico 2.2 – “Evolução Matemática dos métodos Estatísticos” o passo seguinte ao da medição do nosso indicador é calibrá-lo, isto é, atribuir-lhe uma escala que nos permita inferir a idade cronológica. Para este procedimento, foi realizada uma análise de Clusters, que será explicada de seguida.

4.2 – Análise de Clusters

De uma forma geral podemos dizer que o objectivo de uma análise de Clusters passa por interpretar e reconhecer de uma forma plausível uma estrutura existente num conjunto de dados criando grupos (clusters) que reúnam, em cada um, os indivíduos que têm uma menor distância entre eles na medida do indicador utilizado (Helmut, 1980). Aplicado ao nosso problema isto quer dizer que a análise de clusters nos vai informar acerca dos grupos ideais a criar no que respeita à divisão dos rácios obtidos em Fases de união e da distribuição das idades em grupos etários. Esta análise foi realizada recorrendo ao software estatístico SPSS 20.0.

A análise foi repetida de forma a conseguirmos verificar o comportamento dos rácios quando agrupados em 3, 4 e 5 fases e dos grupos etários em 4 e 5 intervalos. De acordo com essa análise, os grupos óptimos (que apresentam melhores resultados) são os que resultam da divisão das idades em 4 grupos etários e dos rácios em 4 fases. Os intervalos obtidos, assim como o número de indivíduos que corresponde a cada um deles, está apresentado na tabela 4.1. No Apêndice A podemos encontrar imagens exemplificativas de cada uma das fases de união; para cada fase existem duas imagens, correspondendo a primeira ao limite inferior da fase e a segunda ao limite superior de união.

Tabela 4.1

Distribuição dos indivíduos da amostra de acordo com a Fase de união epifisial e Grupo Etário criados

Grupo Etário	Fase				Ni
	F1	F2	F3	F4	
	0%	>0%;<50%	>50%;<75%	>75%	
G1 [10;17]	34	0	0	0	34
G2 [18;23]	32	11	17	2	62
G3 [24;27]	0	0	5	25	30
G4 [28;30]	0	0	1	30	31
Nf	66	11	23	57	157

Nota. Ni - Total de indivíduos por classe etária; Nf - Total de indivíduos por fase

Desta forma, devemos considerar que os indivíduos pertencentes à primeira fase não apresentam qualquer sinal de união da epífise à diáfise. Os indivíduos pertencentes à segunda fase apresentam uma união que se caracteriza por ocupar uma área inferior a 50% da extremidade medial da clavícula, isto é, mais de metade da extremidade ainda se encontra por ossificar. A fase 3 caracteriza-se por ter área de ossificação compreendida entre os 50% e os 75% e a fase 4 inclui os indivíduos que têm a extremidade medial com uma taxa de ossificação superior a 75%. Assim, esta tabela lê-se da seguinte forma: “Existem 34 indivíduos com idades compreendidas entre os 10 e os 17 que têm a extremidade medial da clavícula sem qualquer sinal de ossificação da epífise”.

Através deste tipo de análise, conseguimos contornar a usual abordagem de criar as fases de evolução dos indicadores baseados apenas na sua descrição. A quantificação do processo evolutivo e a divisão informada pela análise de clusters vem trazer maior rigor quanto ao rácio ideal para se proceder à transição de uma fase para a outra.

Da divisão feita relativamente às fases talvez salte à vista o facto de a última não corresponder ao estágio de união completa da epífise à diáfise. De facto, esta era a hipótese inicialmente levantada pelos autores. O que acontece é que caso fosse criada uma 5ª Fase que contemplasse apenas os indivíduos com união epifisial completa, a fase que englobava os indivíduos com um rácio superior a 50% e inferior a 99%, ficava com um número extremamente reduzido de indivíduos, uma vez que a actual Fase 4 é composta maioritariamente (75%) por casos que têm um rácio de união igual a 100%. Isto ia levar a que as inferências realizadas acerca da Fase 4 de 5 fossem muito possivelmente enviesadas. Por esta razão, optamos por manter apenas 4 fases.

O valor do rácio obtido por cada indivíduo assim como a sua idade real, sexo, colecção e ano da morte, encontram-se na Tabela B1 (Apêndice B) em apêndice.

É de salientar que, dada a crítica realizada à utilização da regressão linear com dados categoriais, esta divisão em fases apenas será aplicada nas metodologias Bayesiana e de Regularização local, sendo que para a estimativa pela regressão linear e calibração clássica serão utilizados os valores do rácio sem qualquer discretização.

4.3 – Medidas de Avaliação dos Erros e Qualidade dos Modelos

Como foi referido no capítulo da introdução, é essencial que todos os modelos apresentem medições do erro do método que nos indiquem o comportamento deste nos testes realizados assim como a sua capacidade de ser reproduzido por outros autores. Estes dois pontos são essencialmente aqueles que necessitam de ser avaliados.

Para avaliar o comportamento dos modelos criados nos testes efectuados utilizamos um conjunto de medidas que nos vão permitir obter uma correcta avaliação dos vários modelos. As medidas utilizadas assim como uma breve descrição de cada uma delas é de seguida apresentada:

1. **Erro** – É a diferença entre a idade real do indivíduo e a estimada pelo modelo por nós criado (apresentados no Capítulo 5 – Resultados). Quanto maior o erro menos precisa se vai revelar a metodologia empregue;
2. **Desvio Médio Absoluto (DMA)** – É uma média dos valores absolutos dos erros. Esta média é portanto calculada eliminando o facto de certos valores dos erros poderem ser negativos. Num caso em que a idade real seja inferior à idade estimada, o erro vai ser negativo. No cálculo do DMA o valor negativo é tomado como positivo, conseguindo desta forma obter uma verdadeira noção da média do erro na amostra total;
3. **Intervalo de Confiança a 95% (IC) e Amplitude Média do Intervalo de Confiança (AMIC)** – A forma como os IC 95% foram calculados encontra-se descrito nos tópicos 4.5.1 e 4.5.2. ;
4. **Taxa de Cobertura** – Percentagem de intervalos de confiança 95% que incluem a idade real;
5. **Viés** – declive da recta de regressão linear dos erros sobre a idade real que vai, como referido anteriormente, indicar-nos se a idade está ou não a ser sobre ou subestimada.

Tendo em mente as medidas anteriores, iremos considerar um bom modelo aquele que apresente um DMA baixo, um viés próximo de 0 e uma AMIC curta com uma taxa de cobertura o mais próximo possível de 100%. Em termos práticos, a Figura 4.3 ilustra de uma forma simples o que algumas das medidas estão a avaliar.

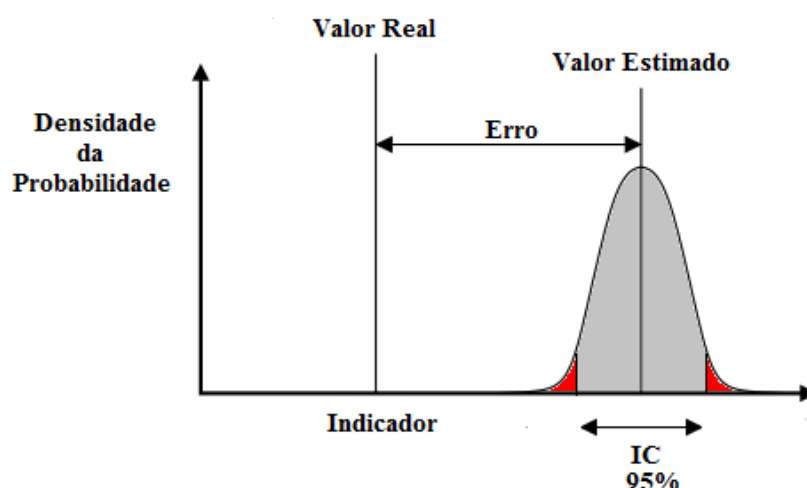


Figura 4.3 – Ilustração das medidas de avaliação do erro “Intervalo de Confiança 95%” e “Erro”

Importa salientar que, no caso dos modelos Bayesiano e de Regularização local, o valor da nossa estimativa surge associado a uma probabilidade *a posteriori* que já nos informa sobre a distribuição dos dados em torno dessa idade, não necessitando assim de qualquer medida de erro associado à estimativa. No entanto, estas medidas são-lhes aqui aplicadas com o intuito de possibilitar uma comparação final de todos os modelos por nós criados. Na sua utilização prática, apenas nos modelos de regressão linear e de calibração clássica se utilizarão as medidas do IC juntamente com a idade estimada pela equação.

Encontradas as medidas que nos vão dar uma ideia acerca da capacidade do modelo realizar boas estimativas, importa testar a sua capacidade de poder ser aplicado por outros investigadores. Para isso, serão medidos os erros intra e inter observador.

As medições das áreas Ae e Ao foram realizadas pelo primeiro autor deste trabalho em dois momentos distintos, separados por aproximadamente um mês. A amostra utilizada para realizar a segunda medição coincide em número e indivíduos com a amostra utilizada na primeira, ou seja, para o cálculo do erro intra observador foi utilizada a totalidade da amostra analisada (N=157). Para verificar o comportamento da forma de medição das áreas quando realizada por um investigador independente, foram escolhidas aleatoriamente 30 imagens do total de 157 (20% da amostra total) e as medidas de Ae e Ao foram efectuadas por um investigador externo ao presente estudo sem qualquer experiência em processos de fotometria ou algum tipo de treino prévio à realização das medições. A escolha por um investigador sem treino e experiência justifica-se pela garantia de melhor se captar as verdadeiras dificuldades de qualquer perito que possa vir a efectuar estas medições.

Para medir os referidos erros, foi utilizada a metodologia avançada por Utermohle e colaboradores (1983) que nos indicam a existência de três características do erro inerentes às medições que devem ser tidas em conta. São elas a frequência, a magnitude e a direcção.

Para a obtenção da informação referente à frequência do erro, foi calculada a Percentagem de concordância (%C), isto é, a percentagem de pares de valores que se mantiveram iguais ou idênticos entre os 2 momentos de medição pelo mesmo observador ou nas medições realizadas por ambos os investigadores. A fórmula utilizada para o cálculo desta percentagem é:

$$\%C = \frac{n - n'}{n} \times 100\%$$

Como forma de medir a magnitude do erro utilizamos uma forma adaptada do Desvio Médio Absoluto a que chamamos Desvio Médio Absoluto para a magnitude do Erro (DMAME), calculado de acordo com a fórmula:

$$DMAME = \frac{\sum_{i=1}^n |x_{i2} - x_{i1}|}{n}$$

Onde X_1 e X_2 correspondem ao valor obtido no primeiro e segundo momento de medição, respectivamente e n corresponde ao valor total de indivíduos replicados.

Tendo em mente que o que nos interessa captar com a medição do erro inter e intra observador é a percentagem do erro que é devida ao erro de medição e não à variação natural entre indivíduos, a Percentagem de Erro de Medição (PEM) é das medidas mais uteis para este efeito uma vez que elimina a variabilidade proveniente das diferenças inter individuais (Cardoso, 2000). A fórmula para o cálculo do PEM é:

$$PEM = \frac{DMAME}{\%C} \times 100\%$$

A fórmula anterior é avançada por Droessler (1981 *in* Utermohle et al., 1983) e, de acordo com ela, a percentagem de erro de medição para o erro intra observador para a amostra por nós estudada é de 0,14%, ou seja, a medição das áreas de acordo com a metodologia já exposta apresenta um erro intra observador praticamente nulo. Já o erro inter observador é considerado elevado, tendo sido obtido um valor significativamente diferente em 10,87% dos casos.

Outro passo importante num estudo desta natureza é a aplicação do método criado a uma amostra de teste. No entanto, este tipo de abordagem implica uma redução significativa na amostra disponível, uma vez que parte da amostra nunca contribui para a criação do modelo para depois poder ser utilizada para testar o modelo criado. Com o intuito de

aproveitar toda a amostra disponível, o teste aos modelos criados foi realizado recorrendo à técnica de Jackknife, que será seguidamente apresentada.

4.4 – Reamostragem por Jackknife

A reamostragem por Jackknife é uma técnica que nos permite, no problema em causa, maximizar o aproveitamento da amostra disponível uma vez que a mesma amostra pode ser utilizada como amostra de referência e amostra de teste. Devido à sua forma de funcionamento esta técnica é também por vezes apelidada de *leave one out*.

Tomando o exemplo da regressão linear, o Jackknife vai testar o modelo utilizando a amostra de referência calculando uma equação de regressão com base em todos os indivíduos da amostra menos o primeiro. Depois disto, a equação encontrada é aplicada ao primeiro indivíduo (que não contribuiu para a criação desta equação) e o erro é calculado para ele. Após o cálculo do erro, o primeiro indivíduo é novamente introduzido na nossa amostra e é retirado o segundo. O procedimento acima referido é sucessivamente repetido para todos os indivíduos da amostra. Como a estimativa da idade à morte é feita para cada indivíduo com uma equação criada sem a contribuição do próprio, o erro que é calculado é o verdadeiro erro que vai ser encontrado num caso prático de identidade desconhecida. Após termos encontrado o erro para todos os indivíduos, é calculado o Desvio Médio Absoluto para os erros. É este valor do DMA que nos vai indicar o desvio médio entre as idades reais e as idades estimadas, ou seja, ficamos a saber o quão distante está a estimativa feita pelo nosso modelo da idade real. Com a aplicação desta metodologia temos a obtenção do DMA que nos avalia o método da mesma forma que teríamos com a utilização de uma amostra de teste separada, com a vantagem de termos aproveitado toda a amostra disponível como referência para a criação dos modelos, o que os torna mais robustos. Caso se verifique que o modelo testado apresenta valores de erro indicativos de um bom método, será criada uma equação com base na amostra total, e é essa equação final que vai posteriormente ser utilizada num caso de estimativa da idade à morte de um indivíduo de identidade desconhecida.

A aplicação do Jackknife para a regressão linear e para a Calibração Clássica foi efectuada com recurso ao *software* MS Excel 2010, tendo sido criadas equações diferentes para cada indivíduo que ia sendo eliminado da amostra, ou seja, as equações que foram criadas para estimar a idade nos testes de cariz regressivo foram criadas com uma amostra de

156 indivíduos (N-1). Para aplicar a técnica e testar assim os modelos criados com Regularização local foi criado pelos autores uma rotina para ser corrida em “R”. A rotina criada encontra-se no Apêndice D. O “R” é um *software opensource* para estatística computacional que utiliza por base uma linguagem de programação com o mesmo nome (Core Team, 2003a; 2003b; 2003c) e pode ser encontrado em <http://www.r-project.org/>.

Todas as medidas de erro deste estudo são apresentadas de acordo com a aplicação desta técnica. Estes resultados assim como os modelos finais estão descritos no capítulo dos resultados.

4.5 – Descrição da Metodologia Estatística Utilizada para a Criação dos Modelos

Como já deve ter ficado patente a esta altura, esta tese visa a criação de vários modelos distintos que tornem possível a estimativa da idade à morte pela observação do processo de união da epífise medial à diáfise da clavícula.

O que torna distintos estes modelos não é a forma de observação da clavícula, mas sim a metodologia com que os dados são tratados e, por conseguinte a forma como depois de feita a observação, a idade vai ser estimada.

No final deste trabalho ficarão disponíveis 5 modelos distintos cuja comparação será feita no capítulo 6 - Discussão e a forma como cada um deverá ser utilizado ficará explícita no capítulo 8 - Aplicação prática dos modelos.

Os princípios teóricos subjacentes à criação destes modelos foram já discutidos, tanto do ponto de vista antropológico como do ponto de vista estatístico. Resta então conhecer alguns detalhes finais referentes à forma como os vários testes estatísticos foram utilizados.

4.5.1 – Regressão Linear e Calibração Clássica

As equações de regressão linear e a da Calibração Clássica foram criadas através do software Excel. O teste a este modelo foi feito de acordo com a técnica de Jackknife, também no mesmo software.

As medidas de erro utilizadas para estes modelos são as 5 descritas no tópico 4.3 referentes à avaliação dos erros dos modelos. Os intervalos de confiança a 95% foram criados de acordo com a forma:

$$\hat{Y} \pm t_{N-2} \left(\frac{\gamma}{2} \right) \cdot S \cdot \sqrt{1 + \frac{1}{N} + \frac{(X - \bar{X})^2}{S_{xx}}}$$

4.5.2 – Versão discreta do Teorema de Bayes

Todas as probabilidades referentes a este modelo foram calculadas no Excel de acordo com o que foi explicado no tópico 2.2.3 – Abordagem Bayesiana.

As medidas de erro para este tipo de modelo e para os seguintes servem apenas o propósito de possibilitar a comparação destes com os restantes. Isto porque, enquanto na regressão linear a equação nos dá uma idade estimada, por exemplo $Y=20$ anos e nós temos que lhe associar o intervalo de confiança a 95% para obtermos uma confiança de 95% de que a idade real está dentro do intervalo reportado, tal não acontece nos modelos que nos dão como resultado uma probabilidade, uma vez que a idade já se encontra em intervalos etários e a estimativa é-nos dada com uma probabilidade de a idade real se encontrar nesse grupo etário.

Para este modelo e para os que serão apresentados a seguir os intervalos de confiança a 95% foram calculados assumindo que a idade estimada era a mediana e o intervalo era considerado o intervalo entre os percentis 0,025 e 0,975. As fórmulas de cálculo para a determinação das medianas e dos percentis foram criadas pelos autores e apresentam variações de acordo com o respectivo modelo.

A forma como os resultados dos vários modelos se devem ler e reportar encontra-se no capítulo 8 – Aplicação prática dos modelos.

4.5.3 – Modelos criados com o método de Regularização local sem e com distribuição marginal conhecida

Foram criados dois tipos de modelos distintos baseados nestes métodos.

O primeiro, a que chamaremos *Modelo 1*, reporta a estimativa da idade na forma de uma probabilidade condicional sem incluir para o seu cálculo algum *a priori*, isto é, o cálculo da probabilidade de o indivíduo pertencer a um determinado grupo etário dado que apresenta uma determinada fase de união, é feito sem qualquer informação da distribuição *a priori* dos dados referentes às idades. Matematicamente esta distribuição *a priori* refere-se à distribuição marginal indicada no nome da metodologia empregue, isto é, de uma forma simplista podemos dizer que estas duas abordagens diferem por uma ter em consideração a distribuição marginal (*a priori*) e a outra não. O Modelo 1 resulta de uma rotina criada em “R” (Apêndice D) para o efeito e que permite o cálculo da função de probabilidade utilizada para a criação da tabela apresentada no respectivo tópico do Capítulo 5 – Resultados. Para chegar aos resultados apresentados na referida tabela, várias combinações entre graus de polinómios para a função baseada em (2.15) e factor de regularização foram testados, tendo sido cada um deles sujeito a reamostragem por Jackknife e os respectivos erros avaliados. Este processo permitiu-nos determinar qual o melhor modelo para o nosso problema. De todas as combinações entre polinómios e factores de regularização testados apenas o melhor modelo é apresentado uma vez que, no total, foram criados e testados 36 modelos só com base na rotina Modelo 1. Somando os 72 modelos criados e testados com o modelo a seguir apresentado (Modelo 2) seria totalmente incompatível com a boa apresentação deste trabalho que os mesmos fossem incluídos, pelo que se optou, por uma questão de transparência, por incluir em anexo todos os dados obtidos assim como as rotinas criadas, resultados obtidos e erros inerentes a cada modelo, para além de explicação detalhada de todos os passos referentes aos testes realizados.

O segundo modelo criado com base no método de Regularização local é o que mais se assemelha a um estimador Bayesiano como nos é apresentado por Lucy et al. (2002), na medida em que a probabilidade reportada (*a posteriori*) é obtida num processo mais próximo com o formulado na equação (2.13). Assim, para este cálculo já foi introduzida a informação *a priori* que nos é dada pela amostra recolhida. Existem várias formas de obter esta informação *a priori*, e a escolha nem sempre é consensual entre os autores (Lucy et al., 2002). No presente estudo, optamos por criar o nosso *a priori* com base nos dados, isto é, a

nossa função de densidade de probabilidades $f(A)$ é construída com base nos dados por nós recolhidos da amostra seleccionada. A inclusão desta função para o cálculo da probabilidade final foi realizada através de uma rotina também criada em R (Apêndice D) para este efeito a que chamaremos *Modelo 2*. Desta forma a probabilidade *a posteriori* é proporcional ao produto da distribuição conjunta (bidimensional e tratada pelo Modelo 1) multiplicada pela distribuição marginal (tratada pelo Modelo 2).

Para além destes dois modelos, foi ainda calculado um terceiro cujo objectivo era verificar o comportamento destas metodologias em amostras reduzidas. Para tal, foram eliminados 138 indivíduos da amostra inicial de forma aleatória. A distribuição dos indivíduos pelas fases e grupos etários depois desta redução ter sido levada a cabo encontra-se na tabela 4.2. A esta amostra foi aplicada a mesma rotina do Modelo 2.

Tabela 4.2

Distribuição dos indivíduos da amostra reduzida por Fase de união epifisial e Grupo Etário

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	5	0	0	0
G2 [18;23]	0	3	3	0
G3 [24;27]	0	0	2	2
G4 [28;30]	0	0	0	4

Capítulo 5 – Resultados

De seguida, serão apresentados os principais resultados deste estudo. Os resultados encontram-se descritos em maior detalhe no Apêndice B na mesma ordem pela qual se apresentam aqui resumidos. As tabelas B1 a B7 (Apêndice B) referem-se aos resultados obtidos para cada um dos indivíduos da amostra quando submetidos aos respectivos modelos pelo método de Jackknife, o que nos dá uma verdadeira percepção do comportamento dos modelos numa situação prática, em que a idade do indivíduo em análise não é conhecida.

Dado o tamanho da amostra, caso os resultados fossem apresentados para todos os indivíduos estudados em cada um dos modelos, as tabelas teriam tamanhos incompatíveis com uma boa leitura das mesmas, pelo que se optou por apresentar apenas um indivíduo representativo de cada idade dada a fase de ossificação (nos modelos Bayesiano e com Regularização local) ou o Rácio (no Modelo de Regressão e de Calibração Clássica), isto porque diferentes indivíduos com a mesma Fase ou Rácio e a mesma idade vão apresentar valores iguais entre si nas diferentes estimativas e medidas de erro. Assim sendo, será apenas de esperar encontrar nas referidas tabelas apenas um indivíduo com 10 anos e fase ossificação “1” (ou Rácio “0%”) (por exemplo) ainda que exista, no total da amostra, mais do que um sujeito com essas características, mas cujos valores calculados são iguais ao indivíduo apresentado.

Antes de termos procedido à criação dos modelos havia duas análises que se impunham. Saber a correlação da variável R com a idade e se existiam diferenças entre as médias obtidas para o rácio nos indivíduos de sexo masculino e os de sexo feminino. Para tal, efectuamos um teste t para amostras independentes o qual demonstrou não existirem diferenças estatisticamente significativas entre os sexos ($t = 1,215$; $df = 155$; $p = 0,226$). De facto, como é possível verificar na tabela 5.1, podemos afirmar com uma probabilidade de erro de 5% que a diferença entre as médias destes dois grupos não é significativamente diferente de $0,1\text{cm}^2$.

Tabela 5.1

Teste t para as diferenças entre as médias dos Rácios obtidos para ambos os sexos

Sexo	N	Média	DP	t	df	Sig. (2-tailed)	Diferença Média
Masculino	73	.56	.058	1.215	155	.226	.097
Feminino	84	.46	.055				

Para saber a correlação entre a variável R e a idade efectuamos uma correlação de Pearson que nos indicou existir uma correlação de $r=0,84$, que demonstra que o processo evolutivo do indicador medido pela variável por nós criada está muito bem correlacionado com a idade cronológica dos indivíduos da amostra, sendo à partida um bom indicador para estimar a idade à morte.

5.1 – Modelo de Regressão Linear

Para a regressão linear a equação obtida através da distribuição dos dados foi:

$$Idade = 10,301 \times Rácio + 16,796$$

O coeficiente de determinação é de $R^2=0,6982$, ou seja, 69,82% da variação da idade é explicada pela relação linear a partir da variação do Rácio.

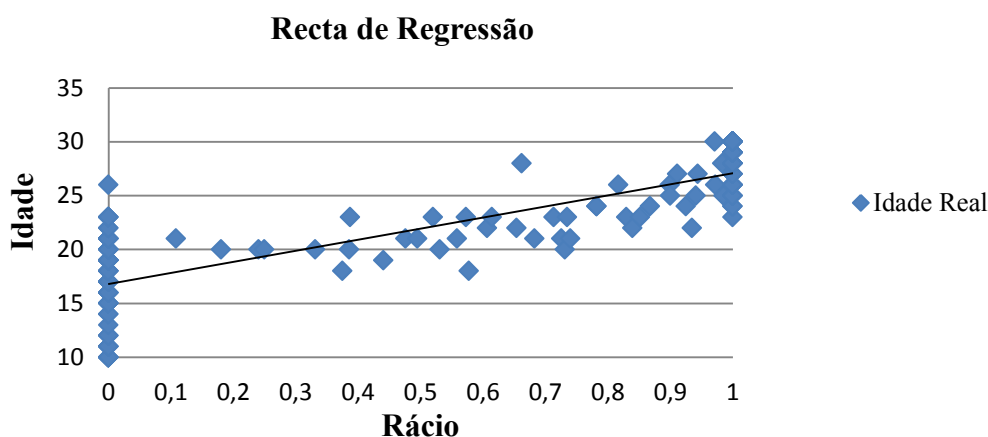


Figura 5.1 – Regressão da Idade sobre o indicador Rácio

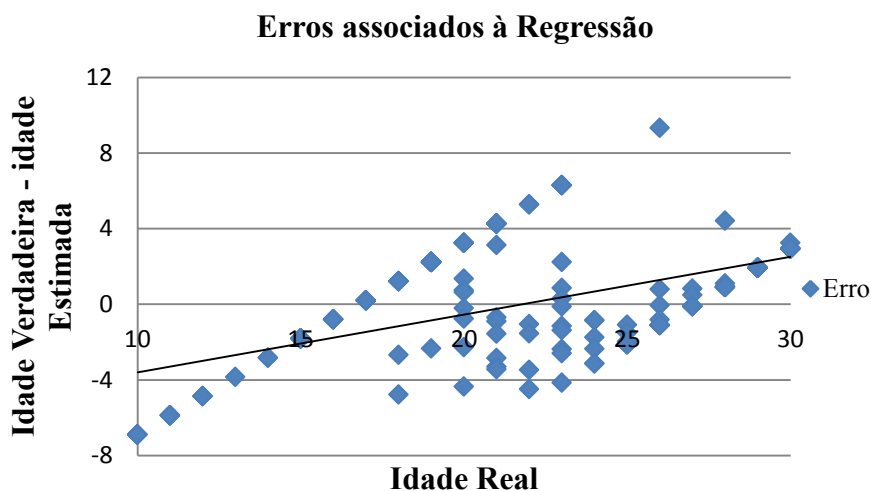


Figura 5.2 – Regressão dos Erros sobre a Idade Real

A aplicação desta recta à amostra em estudo (Figura 5.1) permite-nos observar de imediato a existência de um viés sistemático. Este modelo apresenta um Desvio Médio Absoluto (DMA) = 2,4350 e uma amplitude média dos intervalos de confiança a 95% de 12 anos, o que quer dizer que existe uma confiança de 95% que a idade estimada esteja dentro de um limite médio de aproximadamente 12 anos. A taxa de cobertura do modelo é de 97%, o que indica que 97% dos intervalos de confiança calculados possuem mesmo a idade verdadeira.

Efectuando uma regressão dos erros encontrados sobre a idade verdadeira, é facilmente observado (Figura 5.2) o declive positivo da recta de regressão que tem como consequência a já referida sobrestimação das idades em indivíduos mais novos e a subestimação das idades em indivíduos mais velhos. A recta da regressão dos erros sobre a idade tem a forma:

$$\text{Erro} = 0,3059 \times \text{Idade Real} - 6,607$$

5.2 – Modelo de Calibração Clássica

Para a calibração clássica foi, como já referido, estimada a recta de regressão do Rácio sobre a Idade (Figura 5.3), passando assim a Idade a ser tratada como a variável independente. A equação obtida para esta abordagem foi:

$$Rácio = 0,0678 \times Idade - 0,9925$$

Sendo que a variável observada é o Rácio, temos que resolver a equação anterior em ordem à Idade. Daqui vem:

$$Idade = \frac{(Rácio + 0,9925)}{0,0678}$$

Pela aplicação da equação anterior, obtemos a recta dos Erros associados à Regressão (Figura 5.4), que nos permite visualizar o que foi já referido; o viés, ao coincidir com o eixo das abcissas, é completamente eliminado, o que se traduz na eliminação da usual sobrestimação das idades dos indivíduos mais novos e na subestimação da idade dos indivíduos mais velhos. Os erros ficam assim distribuídos de forma simétrica em relação ao eixo das abcissas e com um DMA = 2,9803. O Coeficiente de determinação deste modelo é igual ao Regressão Linear ($R^2=0,6982$)

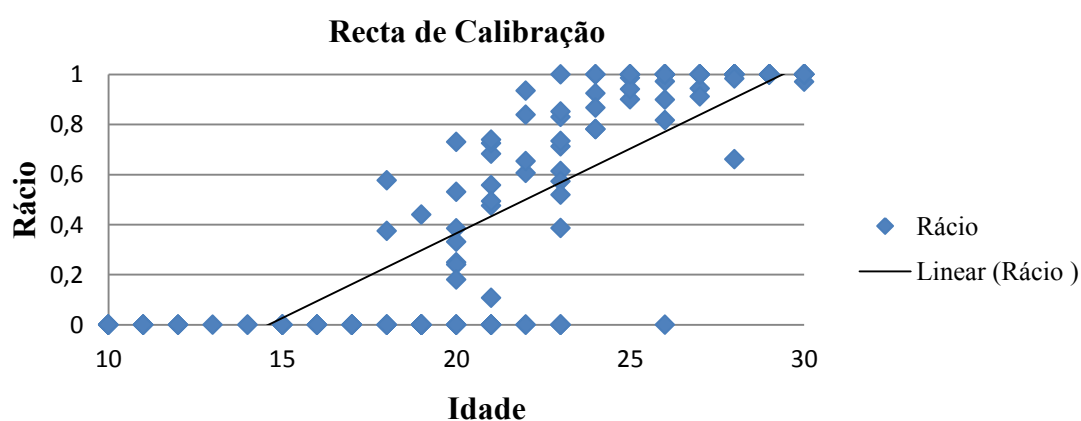


Figura 5.3 – Gráfico da regressão do Rácio sobre a Idade

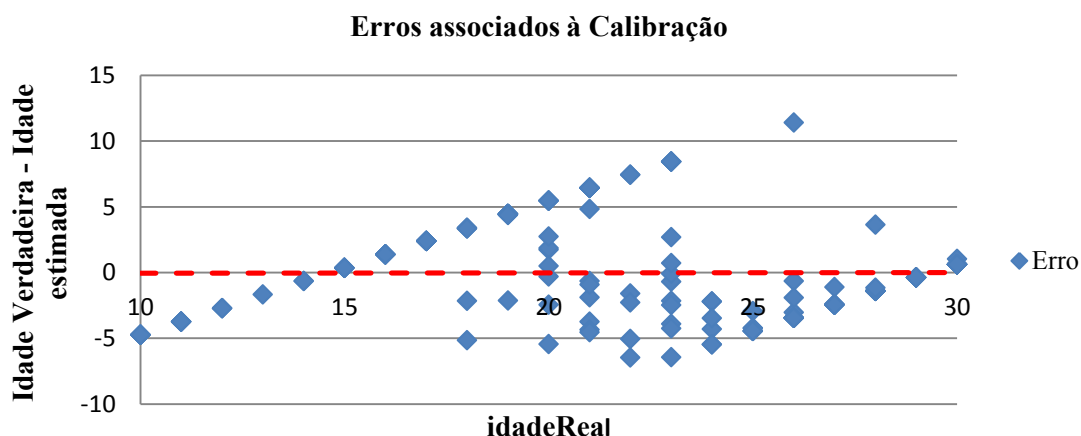


Figura 5.4 – Recta dos Erros provenientes da Calibração. Recta de Regressão assinalada a tracejado vermelho

Este modelo apresenta uma amplitude média dos intervalos de confiança (AMIC) significativamente inferior ao anterior. Para um intervalo de confiança de 95%, a AMIC é de apenas 1 ano. No entanto, esta redução na amplitude dos intervalos de confiança não deve ser vista como algo de notoriamente positivo, uma vez que a cobertura do modelo da Calibração Clássica cai drasticamente face ao da Regressão Linear para 18,18%, ou seja, em apenas 18,18% dos casos a idade real se encontra de facto no intervalo de confiança médio de 1 ano.

5.3 – Versão Discreta do Modelo Bayesiano

As probabilidades *a posteriori* deste modelo encontram-se descritas na tabela 5.2.

Os valores referentes às medidas da avaliação do erro encontram-se na Tabela sumária 5.6, uma vez que para este modelo têm apenas interesse para a comparação deste com os restantes modelos.

Tabela 5.2

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o modelo Bayesiano discreto

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	52%	0%	0%	0%
G2 [18;23]	48%	100%	74%	4%
G3 [24;27]	0%	0%	22%	44%
G4 [28;30]	0%	0%	4%	53%

5.4 – Modelo 1 (Regularização local sem distribuição marginal conhecida)

Tabela 5.3

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 1

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	49%	31%	0%	0%
G2 [18;23]	51%	63%	41%	12%
G3 [24;27]	0%	5%	31%	42%
G4 [28;30]	0%	1%	28%	46%

5.5 – Modelo 2 (Regularização local com distribuição marginal conhecida)

Tabela 5.4

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 2

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	52%	37%	6%	4%
G2 [18;23]	44%	50%	27%	4%
G3 [24;27]	4%	12%	39%	46%
G4 [28;30]	0%	1%	28%	46%

5.6 – Modelo com Amostra Reduzida (Regularização local com distribuição marginal conhecida)

Tabela 5.5

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 2 para a amostra reduzida

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	84%	34%	1%	1%
G2 [18;23]	13%	50%	45%	12%
G3 [24;27]	1%	14%	35%	41%
G4 [28;30]	1%	2%	19%	46%

5.7 – Resumo e Comparação de todos os Modelos e respectivas Medidas de Erro

Tabela 5.6

Valores das medidas de avaliação dos modelos

Modelo	Medidas de erro			
	DMA	AMIC 95% (anos)	Taxa de Cobertura	Viés
Regressão Linear	2.205	12	97%	0,33
Calibração Clássica	2.67	1	18,18%	0
Bayes - Discreto	2.206	9	95,45%	0,34
Modelo 1	2.55	13	100%	0,34
Modelo 2 (N=157)	2.6	15	100%	0,31
Modelo 2 (N=19)	1.9	14	89%	0,17

Capítulo 6 – Discussão

Apresentados os principais resultados do estudo, importa agora discuti-los de forma a verificarmos como é que as metodologias aqui propostas nos podem ajudar numa melhoria aos métodos de estimativa da idade à morte.

Este estudo apresenta dois tipos de resultados que necessitam de uma comparação. Os primeiros referem-se aos resultados provenientes da análise dos dados em bruto, isto é, sem a aplicação de qualquer teste estatístico, apenas de formas de sumarização de dados. Os segundos dizem respeito à comparação dos vários modelos apresentados no capítulo anterior de forma a podermos saber qual o melhor a ser utilizado num contexto prático.

Da análise dos dados em bruto importa clarificar como resolvemos uma questão levantada na revisão da literatura. A idade mínima a que ocorre a união completa da epífise à diáfise foi aqui considerada ser a idade em que o primeiro indivíduo apresenta ossificação completa da extremidade medial. Esta escolha justifica-se pelo facto de, uma vez que esta tese tem como *background* teórico o pensamento da antropologia forense, nos parecer uma escolha mais segura na medida em que engloba assim uma maior variabilidade inter pessoal existente, isto é, reportando o valor do primeiro indivíduo estamos a precaver o facto de poderem vir a aparecer casos em que a união da epífise à diáfise se dá mais cedo do que seria normal. Sendo que poderemos considerar a idade normal para união aquela em que mais de 50% da amostra já apresenta união completa. Assim, no nosso estudo, o primeiro indivíduo a ter a união total da epífise à diáfise tem uma idade 23 anos e é do sexo feminino. O primeiro indivíduo do sexo masculino com união completa tem 25 anos. A idade em que todos os indivíduos apresentam união total é os 28 anos. Comparando os nossos resultados com os de outros autores que também analisaram este indicador ósseo, podemos concluir que a idade mínima para que a união completa ocorra não é significativamente diferente dos estudos de Mckern e Stewart (1957), Todd e D'Errico (1928) e o de MacLaughlin (1990). A diferença maior face a estudos que têm por base métodos anatómicos é encontrada quando os resultados são comparados com os de Webb e Suchey (1985). No entanto, o facto de estes autores terem reportado uma idade mínima de união completa inferior à nossa (21 anos) pode dever-se à amostra por eles utilizada. A amostra de Webb e Suchey é composta por indivíduos com anos de morte compreendidos entre 1977 e 1979, quase 50 anos após a média

de anos de morte da nossa amostra. Shirley (2009) num estudo que efectuou com uma amostra significativa de clavículas reportou a existência de efeitos de tendência secular neste indicador, que tinham como resultado a idade de união da epífise à diáfise começar a acontecer cada vez mais cedo. Este facto pode ser a explicação para a diferença encontrada entre o estudo de Webb e Suchey e o nosso.

Em relação aos estudos que têm por base uma análise de dados imagiológicos, estes reportam sistematicamente idades de união mais novas que a nossa. Em relação a este ponto, conclusões definitivas não podem ser tecidas uma vez que a variabilidade entre os métodos é demasiada para que conclusões se possam tirar de uma comparação. Para além de se utilizarem amostras de diferentes contextos (arqueológico e forense) o método de análise é também distinto e não existem evidências que seja possível aplicar métodos anatómicos a material que tem por base uma análise imagiológica (Schmelting et al., 2004). De facto esta pode ser uma das fontes das diferenças encontradas nos resultados reportados pelos vários autores. O facto de se aplicarem metodologias distintas leva muito provavelmente a que os resultados obtidos para a idade mínima em que se atinge a última fase do processo de ossificação seja diferente. Esta questão poderá ser resolvida com a aplicação de um método anatómico a imagens radiológicas (convencionais ou digitais), ou vice-versa, e pela comparação dos resultados obtidos. Enquanto este tipo de estudos não for feito não podemos afirmar se a diferença existente entre os vários estudos é devida a diferenças populacionais, às metodologias empregues ou a ambas. Desta forma, não conseguimos inferir se as diferenças encontradas são devidas a efeitos de tendência secular, às diferentes metodologias aplicadas, às diferentes populações em que os métodos foram testados ou a outra causa.

A impossibilidade desta comparação serve, uma vez mais, para realçar a importância da criação de *standards* de criação e aplicação de métodos de estimativa da idade. Cremos que seria imprescindível a criação de grupos de estudo à semelhança do que já existe na Alemanha – *Study Group of Forensic Age Diagnostics* – para que estas e outras limitações possam ser discutidas e, acima de tudo, para que protocolos tanto para a criação como para a aplicação de alguns métodos e da sua avaliação possam ser desenvolvidos, pois só assim será possível uma comparação de dados quer intra quer inter países e investigadores que nos permita compreender os pontos que necessitam de maior estudo nesta temática, de que forma o ambiente socioeconómico e cultural influencia a idade biológica de um indivíduo, que métodos necessitam de mais ou menos experiência para a sua aplicação por parte dos

antropólogos forenses e, como consequência inevitável desta uniformização, quais os melhores métodos a serem aplicados nas diversas situações.

São também da maior importância para este efeito a criação de colecções de esqueletos identificados de cariz forense, à imagem da que existe no Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra (Curate, 2010) – Colecção de Esqueletos Identificados do Século XXI – que sendo a maior colecção deste género em Portugal e por se encontrar em expansão adquire grande importância no meio da investigação em antropologia forense em Portugal.

No que diz respeito aos modelos por nós criados importa realçar antes de mais o comportamento do Modelo 2 quando aplicado à amostra de reduzidas dimensões. Apesar de reportar uma taxa de cobertura baixa (89%) em comparação com o mesmo modelo com a amostra total, os baixos valores do viés e do DMA, assim como um AMIC não muito diferente das restantes abordagens, deixa antever que este modelo de criação das funções de densidade pode ser de grande utilidade em antropologia forense, dada a sua capacidade de extrair informação de um pequeno número de dados.

Consideramos também que os restantes modelos baseados no método de Regularização local apresentam-se como melhores estimadores da idade uma vez que, apesar de o valor do viés, do DMA e do AMIC ser ligeiramente mais alto que o da regressão linear a taxa de cobertura é de 100%, ou seja, podemos ter 100% de confiança que o intervalo que é dado para a idade estimada está correcto, o que num processo de carácter judicial se reveste de enorme importância.

Outra clara vantagem na escolha dos métodos baseados no Modelo 1 e Modelo 2 é a facilidade de utilização face ao da regressão linear. Enquanto para a aplicação da regressão linear temos que tirar a fotografia da extremidade da clavícula, medir as áreas e calcular o rácio para obter a estimativa, com os referidos modelos apenas temos que identificar a fase de união e verificar qual a distribuição de probabilidades na tabela reportada para o respectivo modelo. Temos ainda que acrescentar que, dado o erro inter observador encontrado, qualquer perito deveria receber algum tipo de treino antes de efectuar as medições nas fotografias.

Todo o processo de criação desta metodologia foi pensado e realizado tendo em mente as propostas indicadas em Cunha et al. (2009) acerca dos requerimentos das metodologias de estimativa da idade à morte e, do ponto de vista estatístico, as principais influências foram

Konigsberg e colaboradores (1994), Lucy et al. (2002), Martins e colaboradores (2011). Contudo, teve este estudo a preocupação de permitir que os modelos criados fossem de simples utilização, não sendo necessário na aplicação dos modelos finais efectuar qualquer teste estatístico ou outro tipo de cálculo para se chegar à estimativa da idade. Isto não acontece com as abordagens feitas pelos autores referidos que impõem que para se aplicar os métodos por eles propostos se realizem alguns cálculos matemáticos e processos computacionais que tornam o método pouco prático e intuitivo. Esta característica é assim uma melhoria face aos estudos acima citados que implicavam algum trabalho matemático por parte do perito para conseguir chegar à estimativa.

Outra clara vantagem desta abordagem para a antropologia forense é o facto de, caso seja necessário dar uma resposta com maior precisão na estimativa, como o caso de Timucin ‘O’ por exemplo, podemos sempre aproveitar as rotinas desenvolvidas em *R* para este estudo e alterar a forma de reportar os resultados. Assim, é possível obtermos uma probabilidade *a posteriori* não para um intervalo etário mas sim para uma idade específica, o que com um método baseado em modelos de regressão é impossível.

Os modelos baseados em estimadores do núcleo deixam claramente uma grande expectativa quanto aos resultados que nos podem dar. Os autores do presente estudo visam melhorar os resultados do método baseado nos modelos Modelo 1 e Modelo 2 através de uma conjugação entre estes e da informação que é fornecida pela versão discreta do teorema de Bayes. O facto de a distribuição das probabilidades provenientes da versão discreta do teorema poder informar os parâmetros de regularização utilizados pelo Modelo 1 e Modelo 2 irá aumentar as probabilidades *a posteriori* para cada grupo etário, tornando-as mais uteis e mais próximas das verdadeiras probabilidades, isto é, com um erro mais reduzido. Este procedimento de conjugação das duas abordagens será realizado num estudo posterior uma vez que a ideia surgiu apenas após os resultados estarem concluídos e este procedimento necessitar de algumas alterações na computação das rotinas de cada modelo que necessitam de mais tempo do que aquele que este estudo nos permite.

Capítulo 7 – Considerações finais

O presente estudo analisou o processo de união da epífise medial à diáfise da clavícula. Pretendíamos com as metodologias criadas para este estudo e posteriormente seguidas conseguir quantificar o processo analisado, diminuindo desta forma a subjectividade inerente às análises baseadas em descrições morfológicas e criar uma análise estatística para os modelos que fossem robustas em termos de erros e úteis para a prática forense, cumprindo os pressupostos para a criação de metodologias avançadas em Cunha et al. (2009) e as guidelines saídas do processo Daubert (1993).

O primeiro passo deste estudo foi a análise do erro inerente à realização da variável por nós criada como indicador da idade. Após verificação do erro intra e inter observador e do facto de este ser praticamente inexistente e contornável com treino assumimos que esta medida tem uma boa reprodutibilidade, o que nos deixou com segurança para prosseguir o estudo. Importava também saber se a medida criada era um bom indicador do processo de união em estudo, e para tal foi calculado o coeficiente de correlação de Pearson, cujo valor de 0,89 nos indica que a variável é um óptimo indicador para estimar a idade á morte.

Foram de seguida analisadas as possíveis diferenças existentes entre os sexos no que diz respeito à variável por nós criada. A realização de um teste t para amostras independentes permitiu concluir que não existem diferenças estatisticamente significativas na média do rácio de ossificação entre masculinos e femininos.

A análise do processo de união da nossa amostra revela que a idade mínima para o processo de união estar completo é 23 anos, o que está de acordo com o observado por outros investigadores que utilizam métodos anatómicos em amostras diferentes da nossa que indicam que este processo varia entre os 22 e os 24 anos, excepção feita para o estudo de Webb e Suchey (1985) que reporta uma idade de 21 anos. Esta diferença pode ser atribuída a efeitos de tendência secular entre as amostras utilizadas pelos diversos autores e a amostra utilizada por Webb e Suchey.

Foram também encontradas diferenças face aos métodos que têm por base de análise a imagiologia. No entanto, não nos é possível avançar com uma explicação definitiva para estas

diferenças devido ao facto de existirem demasiadas diferenças metodológicas entre os estudos imagiológicos e o nosso.

Os 5 modelos por nós avançados reportaram valores de erro satisfatórios, o que nos leva a crer que a amostra utilizada é robusta à criação da metodologia, que a variável utilizada é um bom indicador da idade à morte e que estes métodos estatísticos devem ser os preferidos pelos autores quando tratam este tipo de dados. De facto, com excepção do modelo de calibração, todos os modelos mostram um bom comportamento na estimação da idade pretendida. O modelo de calibração, apesar de eliminar completamente o viés, apresenta uma taxa de cobertura que leva a que não seja um modelo muito fiável, isto é, a estimativa dada pela equação do modelo tem uma alta probabilidade de não estar dentro do intervalo de confiança reportado.

O Modelo 2 mostra uma diminuição no valor do viés que nos indica que a inclusão da distribuição marginal no cálculo da probabilidade de uma idade à morte dado um indicador é uma mais-valia para este tipo de abordagem. Este facto e a simplicidade associada à sua aplicação faz com que este seja considerado pelos autores deste estudo o melhor modelo encontrado, ressalvando no entanto que poderá ainda ser melhorado.

Findo este estudo importa realçar que num futuro próximo é da maior importância que sejam criados grupos de estudo de estimativa da idade à morte que nos tragam uma base sob a qual devemos criar e aplicar métodos para este tipo de análise. É importante realçar também a importância de colecções de esqueletos identificados de cariz forense, pois só desta forma podemos criar verdadeiros métodos que possam ser aplicados em antropologia forense de forma segura e rotineira. Os modelos aqui apresentados foram criados tendo por base uma amostra identificada de cariz arqueológico e, dado a existência de um comprovado efeito da tendência secular para o indicador em causa, a utilização destes modelos a casos forenses pode não ser muito segura. No entanto, toda a metodologia que levou aos modelos finais pode ser aplicada à criação de um modelo com aplicação forense, basta que para isso se utilize uma amostra proveniente de cadáveres identificados. Para este trabalho isso não foi possível uma vez que não ia ser conseguida uma amostra significativa em tempo útil. Tal facto é apoiado pelos números do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. que reportam a realização de apenas 14 autópsias a indivíduos com idades entre os 10 e os 30 anos entre o dia 1 de Setembro de 2011 e 31 de Março de 2012 na delegação do centro do INMLCF,IP.

Fica assim a vontade de esta metodologia ser aplicada a uma amostra de cariz forense para além de ser posteriormente reproduzida em populações distintas da Portuguesa com o intuito de se analisar o seu comportamento e podermos fazer inferências em relação às diferenças encontradas.

Capítulo 8 – Aplicação prática dos Modelos

Para uma melhor compreensão da forma como se aplicam os métodos acima referidos num verdadeiro caso de antropologia forense, serão de seguida explicados os passos para aplicar cada um dos modelos e a forma como se devem reportar as respectivas respostas aos pedidos feitos pelas entidades judiciais no que à estimativa da idade à morte diz respeito.

Todos os modelos serão aplicados a um hipotético caso de um indivíduo de identidade desconhecida cuja clavícula esquerda apresenta características morfológicas compatíveis com a Fase de ossificação 2, tendo uma área $A_e = 4,171$ e $A_o = 1,562$ o que perfaz um Rácio = 0,374490

Modelo de Regressão Linear

Para a aplicação deste modelo temos que começar por fotografar a extremidade medial da clavícula com uma escala incluída para a imagem. Após digitalização da fotografia tirada com recurso a um *software* que permita a realização de fotometria iremos calcular as áreas A_e e A_o que nos vão permitir calcular o Rácio de ossificação da clavícula chegando assim à Fase em que esta se encontra.

No referido caso o $R = 0,374490$ vai ser substituído na equação linear encontrada através da nossa amostra:

$$Idade = 10,301 \times Rácio + 16,796$$

ou seja,

$$Idade = 10,301 \times 0,374490 + 16,796$$

$$\Leftrightarrow Idade = 20,65$$

De acordo com este método, o indivíduo em análise tem uma idade média de 21 anos, compreendida no intervalo de confiança a 95% entre os 15 e os 27. A taxa de certeza desta estimativa é de 97%

Modelo de Calibração Clássica

O procedimento inicial para a aplicação deste modelo é em tudo idêntico ao anteriormente descrito.

Após encontrado o Rácio, este é substituído na equação:

$$Idade = \frac{(Rácio + 0,9925)}{0,0678}$$

ou seja,

$$Idade = \frac{(0,374490 + 0,9925)}{0,0678}$$

$$\Leftrightarrow Idade = 20,16$$

De acordo com este método, o indivíduo em análise tem uma idade média de 20 anos, compreendida no intervalo de confiança a 95% entre os 19 e os 21. A taxa de certeza desta estimativa é de 18%.

Modelo Bayesiano e Modelos 1 e 2

Para a aplicação destes 3 modelos, apenas necessitamos de identificar a fase de ossificação em que a clavícula se encontra através da análise morfológica da mesma e, sabendo isso, ler o valor das respectivas probabilidades *a posteriori* referentes a essa mesma fase.

Sendo as tabelas para o modelo Bayesiano, Modelo 1 e Modelo 2 as seguintes:

Tabela 8.1

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o modelo Bayesiano discreto

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	52%	0%	0%	0%
G2 [18;23]	48%	100%	74%	4%
G3 [24;27]	0%	0%	22%	44%
G4 [28;30]	0%	0%	4%	53%

Tabela 8.2

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 1

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	49%	31%	0%	0%
G2 [18;23]	51%	63%	41%	12%
G3 [24;27]	0%	5%	31%	42%
G4 [28;30]	0%	1%	28%	46%

Tabela 8.3

Probabilidades a posteriori consoante a Fase de união epifisial e Grupo etário de acordo com o Modelo 2

	Fase			
	F1	F2	F3	F4
Grupo Etário	0%	>0%;<50%	>50%;<75%	>75%
G1 [10;17]	52%	37%	6%	4%
G2 [18;23]	44%	50%	27%	4%
G3 [24;27]	4%	12%	39%	46%
G4 [28;30]	0%	1%	28%	46%

Para o caso em análise, os resultados são reportados da seguinte forma (tendo em conta o Modelo1, por exemplo):

Estando o indivíduo em análise na Fase 2 do processo de união da epífise medial da clavícula à diáfise de acordo com o método proposto por Vicente e colaboradores (2012), podemos afirmar com 100% de certeza que existe uma probabilidade de 31% de ele ter entre 10 e 17 anos; 63% de ele ter entre 18 e 23 anos; 5% de ele ter entre 24 e 27 anos e 1% de ele ter entre 28 e 30 anos.

Bibliografia

- Ahlqvist J, Damsten O. 1969. A modification of Kerley's method for the microscopic determination of age in human bone. *Journal of Forensic Science* **14**:205-212.
- Albrecht G. 1980. Multivariate analysis and the study of form, with special reference to canonical variate analysis. *American Zoologist* **20**:679-93.
- American Board of Forensic Anthropology. 2006. *Official website* [www.theabfa.org]. Acedido em: 25 de Maio de 2012.
- Aykroyd RG, Lucy D, Pollard AM. 1996. *Statistical methods for the estimation of human age at death*. Research report n° Stat-96/08 [não publicado]. University of Leeds.
- Baccino E, Schmitt A. 2006. Determination of Adult Age at Death in the Forensic Context. In *Forensic Anthropology and Medicine: Complementary Sciences from recovery to cause of death*, Schmitt A, Cunha E, Pinheiro J (eds). Totowa, NJ: Humana Press; 259-280.
- Baccino E, Ubelaker DH, Hayek L-AC, Zerilli A. 1999. Evaluation of seven methods of estimating age at death from mature human skeletal remains. *Journal of Forensic Science* **44(5)**:931-936.
- Baccino E, Zerilli A. 1997. The Two step strategy (TSS) or the right way to combine a dental (Lamendin) and an anthropological (Suchey-Brooks system) method for age determination (abstract). *Proceedings of American Academy of Forensic Sciences* **150**.
- Bass WM. 1969. Recent developments in the identification of human skeleton material. *American Journal of Physical Anthropology* **30**:459-461.
- Bass WM. 1979. Developments in the identification of human skeletal material. *American Journal of Physical Anthropology* **51**:555-562.
- Bassed RB, Drummer OH, Briggs C, Valenzuela A. 2010. Age estimation and the medial clavicular epiphysis: analysis of the age of majority in an Australian

- population using computed tomography. *Forensic Science and Medical Pathology* **7**(2):148-154.
- Bedford ME, Russel KF, Lovejoy CO, Meindl RS, Simpson SW, Stuart-Macadam PL. 1993. Test of the multifactorial aging method using skeletons with known ages-at-death from the Grant collection. *American Journal of Physical Anthropology* **91**:287-297.
- Bouvier M, Ubelaker DH. 1977. A comparison of two methods for the microscopic determination of age at death. *American Journal of Physical Anthropology* **46**:391-394.
- Boyes R. 2003. Murderer says he is too young to grow old in jail. *The Times* 8th July.
- Brooks ST, Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Ascádi-Nemeskeri and Suchey-Brooks methods. *Human Evolution* **5**:227-238.
- Buckberry JL, Chamberlain AT. 2002. Age estimation from the auricular surface of the ilium: A revised method. *American Journal of Physical Anthropology* **119**:231-239.
- Buikstra JE, King JL, Nystrom KC. 2003. Forensic anthropology and bioarchaeology in the American Anthropologist: rare but exquisite gems. *American Anthropologist* **105**:38-52.
- Cardoso HFV. 2008. Epiphyseal Union at the Innominate and Lower Limb in a Modern Portuguese Skeletal Sample, and Age Estimation in Adolescent and Young Adult Male and Female Skeletons. *American Journal of Physical Anthropology* **135**:161-170.
- Core Team 2003c. *An Introduction to R*.
- Core Team. 2003a. *R Language Definition*.
- Core Team. 2003b. *Writing R extensions*.

- Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, Lynnerup N, Cattaneo C. 2009. The problem of aging human remains and living individuals: a review. *Forensic Science International* **193**:1-13.
- Cunha E, Wasterlain S. 2007. The Coimbra identified osteological collections. In *Skeletal series and their socio-economic context*, Grupe G, Peters J (eds). Rahden/Westf: Documenta Archaeobiologiae; 23-33.
- Curate F. 2010. *O perímetro do declínio: osteoporose e fracturas de fragilidade em três amostras osteológicas identificadas portuguesas : séculos XIX & XX*. Universidade de Coimbra.
- Daubert. 1993. Daubert et ux., individually and as guardians and litem for Daubert, et al. v. Merrel Dow Pharmaceuticals, INC. Supreme Court of the United States of America.
- Dirkmaat DC, Symes SA, Vey E, Smith OC. 2002. Recognizing child abuse in the thoracic region through a multidisciplinary approach. *Proceedings of the 54th Annual Meeting of the American Academy of Forensic Sciences* **8**:248-249.
- Dwight T. 1878. *The Identification of the Human Skeleton. A Medico-Legal Study*. Boston.
- Fazekas I, Kosa F. 1978. *Forensic Fetal Osteology*. Budapest: Akademiai Kiado.
- Ferembach D, Schwidetzky I, Stloukal M. 1980. Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution* **9**:517-549.
- Flecker H. 1932. Roentgenographic observations of the times appearance of epiphyses and their fusion with the diaphysis. *Journal of anatomy* **67**:118-164.
- Flecker H. 1942. Time of appearance and fusion of ossification centers as observed by roentogenographic methods. *American Journal of Roentgenology* **47**:95-159.
- Galstaun G. 1937. A study of ossification as observed in Indian subjects. *Indian Journal of Medical Research* **25**:267-324.
- Gustafson G. 1950. Age determination on teeth. *Journal of the American Dental Association* **41**:45-54.

- Helmut S. 1980. *Cluster analysis algorithms for data reduction and classification of objects*. New York: Halsted Press.
- Hrdlička A. 1920. *Anthropometry*. Philadelphia: Wistar Institute of Anatomy and Biology.
- Iordanidis P. 1961. Détermination du sexe par les os du squelette. *Annales de Médecine Légale* **41**:280-291.
- Iscan MY, Loth SR, Wright RK. 1984. Age estimation from the rib by phase analysis: white males. *Journal of Forensic Science*; **29**: 1094-1104.
- Jackes M. 2000. Building the bases for paleodemographic analysis: adult age estimation. In *Biological Anthropology of the Human Skeleton*, Katzenberg MA, Saunders SR (eds). New York: Wiley-Liss; 417-466.
- Jacob P, Oliveira PE. 2011. Relative smoothing of discrete distributions with sparse observations. *Journal of Statistical Computation and Simulation* 81(1):109-121.
- Jacob P, Oliveira PE. 2012. Local smoothing with given marginal. *Journal of Statistical Computation and Simulation* 82(6):915-926.
- Jit I, Singh S. 1956. Estimation of stature from clavicles. *Indian Journal of Medical Research* **44**:137-155.
- Jit I, Kulkarni M. 1976. Times of appearance and fusion of epiphyses at the medial end of the clavicle. *Indian Journal of Medical Research* **64**: 773-782.
- Johanson G. 1971. Age determination from human teeth. *Odontologisk Revy*. **22(2)**.
- Johnston FE. 1961. Sequence of epiphyseal union in a prehistoric Kentucky population from Indian Knoll. *Human Biology* **33**:66-81.
- Kellinghaus M, Schulz R, Vieth V, Schmidt S, Schmelting A. 2010. Forensic age estimation in living subjects based on the ossification status of the medial clavicular epiphysis as revealed by thin-slice multidetector computed tomography. *International Journal of Legal Medicine* **124**:149-154.

- Kerley E. 1965. The microscopic determination of age un human bones. *American Journal of Physical Anthropology* **23**:149-165.
- Kerley ER, Ubelaker DH. 1978. Revisions in the microscopic method of estimating age at death in human cortical bone. *American Journal of Physical Anthropology* **49**:545-546.
- Klepinger L. 2006. *Fundamentals of Forensic Anthropology*. Hoboken, NJ: Wiley.
- Komar DA, Buikstra JE. *Forensic Anthropology: Contemporary Theory and Practice*. Oxford: Oxford University Press.
- Konigsberg LW, Frankenberg SR, Walker RB. 1994. Regress what on what?: Paleodemographic age estimation as a calibration problem. In *Integrating archaeological demography: Multidisciplinary approaches to prehistoric population*, Paine RR (ed). Carbondale: Southern Illinois University; 64-88.
- Konigsberg LW, Frankenberg SR. 1992. Estimation of age structure in anthropological demography. *American Journal of Physical Anthropology* **89**:235-256.
- Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. 2008. Estimation and Evidence in Forensic Anthropology. *Journal of Forensic Science* **53(3)**:541-557.
- Kreitner KF, Schweden FJ, Riepert T, Nafe B, Thelen M. 1998. Bone age determination based on the study of the medial extremity of the clavicle. *European Radiology* **8**:1116-1122.
- Krogman A. 2010. *The Human Skeleton in Forensic Medicine*. Springfield, IL: Charles C. Thomas.
- Lamendin H, Baccino E, Humbert JF, Taverier JC, Nossintchouk RM, Zerili A. 1992. A simple technique for age estimation in adult corpses: the two-criteria dental method. *Journal of Forensic Sciences* **37**:1373-1379.

- Lovejoy CO, Meindl RS, Pryzbeck TR, Mensforth RP. 1985. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology* **68**:15-28.
- Lucy D, Aykroyd RG, Pollard AM, Solheim T. 1996. A Bayesian approach to adult human age estimation from dental observation by Johanson's age changes. *Journal of Forensic Science* **41(2)**:189-194.
- Lucy D, Aykroyd RG, Pollard AM. 2002. Nonparametric age estimation. *Applied Statistics* **52(2)**:185-196.
- Lucy D, Pollard AM. 1995. Further comments on the estimation of error associated with the Gustafson dental age estimation method. *Journal of Forensic Sciences* **40(2)**:222-227.
- Lucy D. 1997. *Human age estimation from skeletal and dental evidence*. Tese de Doutoramento [não publicado]. University of Bradford.
- Lucy D. 2010. The presentation of the results and statistics for legal purposes. In *Age estimation in the living*, Black S, Aggrawal A, Payne-James J (eds). Wiley-Blackwell
- Lynch SM. 2007. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer.
- Macaluso Jr PJ. 2011. Sex discrimination from the glenoid cavity in black South Africans: morphometric analysis of digital photographs. *International Journal of Legal Medicine* **125(6)**:773-778.
- MacLaughlin SM. 1990. Epiphyseal fusion at the sternal end of the clavicle in a modern Portuguese skeletal sample. *Antropologia Portuguesa* **8**:59-68.
- Maples WR. 1978. An improved technique using dental histology for the estimation of adult age. *Journal of Forensic Science* **23**:764-770.
- Martins R, Oliveira PE, Schmitt A. 2012. Estimation of age at death from the pubic symphysis and the auricular surface of the ilium using a smoothing procedure. *Forensic Science International* **219(1)**:287-293.

- Martins R. 2005. *Calibração não paramétrica na estimação da idade*. Tese de mestrado [não publicado]. Departamento de Matemática – Universidade de Coimbra.
- Masset C. 1989. Age estimation on the basis of cranial sutures. In *Age markers in the human skeleton*, Iscan MY (ed). Springfield, IL: Charles C Thomas; 71-103.
- McKern TW, Stewart TD. 1957. *Age changes in young American males. Analyzed from the standpoint of age identification*. Environmental Protection Research Division, Technical Report EP-45. Natick, MA: Quartermaster Research and Development Center, US Army.
- Nawrocki SP. 2010. The Nature and Sources of Error in the Estimation of Age at Death from the Skeleton. In *Age Estimation of the Human Skeleton*, Latham KE, Finnegan M (eds). Springfield, IL: Charles C. Thomas; 79-101.
- Ogden JA. 1979. The development and growth of the musculo-skeletal system. In *Scientific Basis of Orthopaedics*, Albright JA, Brands R (eds). New York: Appleton-Century-Crofts; 41-103.
- Orfila MJB. 1823. *Leçons de Médecine Légale*. Paris: Béchet Jeune.
- Parsons FG, Box CR. 1905. The relation of the sutures to age. *J. Anthropol. Inst* **35**:30-38.
- Pearson K. 1899. Mathematical contributions to the theory of evolution: on the reconstruction of the stature of prehistoric races. *Philos. Trans. R. Soc.* **192**:169-244.
- Pfeiffer S. 1992. Cortical bone age estimates from historically known adults. *Z. Morphol. Anthropol.* **79**:1-10.
- Quirnbach F, Ramsthaler F, Verhoff MA. 2009. Evaluation of the ossification of the medial clavicular epiphyses with a digital ultrasonic system to determine the age threshold of 21 years. *International Journal of Legal Medicine* **123**:241-245.

- Reichs KJ. 1995. A professional profile of diplomates of the American Board of Forensic Anthropology:1984-1992. *Journal of Forensic Sciences* **40(2)**:176-182.
- Ritz-Timme S, Cattaneo C, Collins MJ, Waite ER, Schütz HW, Kaatsch HJ, Borrman HIM. 2000. Age estimation: The state of the art in relation to the specific demands of forensic practice. *International Journal of Legal Medicine* **113**:129-136.
- Ruff CB. 2008. Biomechanical analysis of archaeological human skeletons. In *Biological Anthropology of the Human Skeleton*, Katzenberg MA, Saunders SR (eds). New York: Wiley-Liss; 183-206.
- Ryan TP. 1996. *Modern Regression Methods*. New York: Wiley Series on Probability and Statistics.
- Sanders I, Woesner ME, Ferguson RA, Noguchi TT. 1972. A new application of forensic radiology: identification of deceased from a single clavicle. *American Journal of Roentgenology* **115**: 619-622.
- Santos AL. 1995. *Certezas e incertezas sobre a idade à morte*. Provas de aptidão pedagógica e capacidade científica, Departamento de Antropologia, Universidade de Coimbra.
- Saunders SR, Fitzgerald C, Rogers T, Dudar C, McKillop H. 1992. A test of several methods of skeletal age estimation using a documented archaeological sample. *Canadian Society of Forensic Science* **25**:97-118.
- Schaefer MC. 2008. A summary of epiphyseal union timings in Bosnian males . *International Journal of Osteoarchaeology*. **18**: 536–545.
- Scheuer L, Black S. 2000. *Developmental Juvenile Osteology*. San Diego, California: Elsevier.
- Schmeling A, Geserick G, Reisinger W, Olze A. 2007. Age estimation. *Forensic Science International* **165**:171-181.

- Schmeling A, Schluz R, Reisinger W, Müller M, Wernecke KD, Geserick G. 2004. Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography. *International Journal of Legal Medicine* **118**:5-8.
- Schmitt A. 2001. Variabilité de la sénescence du squelette humain Réflexions sur les indicateurs de l'âge au décès: à la recherche d'un Outil performant. Dissertação [não publicado]. Université Bordeaux.
- Schulz R, Müller M, Reisinger W, Schmidt S, Schmeling A. 2008. Radiographic staging of ossification of the medial clavicular epiphysis. *International Journal of Legal Medicine* **122**:55-58.
- Sherer DM, Sokolovski M, Dalloul M, Khoury-Collado F, Osho JA, Lamarque MD, Abulafia O. 2006. Fetal clavicle length throughout gestation: a nomogram. *Ultrasound in Obstetrics & Gynecology* **27**:306-310.
- Shirley NR. 2009. *Age and sex estimation from the human clavicle: an investigation of traditional and novel methods*. Tese de Doutorado [não publicada]. University of Tennessee.
- Singh IJ, Gunberg DL. 1970. Estimation of age at death in human males from quantitative histology of bone fragments. *American Journal of Physical Anthropology* **33**:373-381.
- Snow CC. 1982. Forensic anthropology. *Annual Review of Anthropology* **11**:97-131.
- Spencer F. 1997. *History of Physical Anthropology, Vol 1*. New York, NY: Garland.
- Steel FLD. 1966. Further observations on the osteometric discriminant function: the human clavicle. *American Journal of Physical Anthropology* **25**:319-322.
- Steele J, Mays S. 1995. Handedness and directional asymmetry in the long bones of the human upper limb. *International Journal of Osteoarchaeology* **5**:39-49.
- Stephan CN, Winburn AP, Christensen AF, Tyrrel AJ. 2011. Skeletal identification by radiographic comparison: Blind tests of a Morphoscopic method using antemortem chest radiographs. *Journal of Forensic Science* **56(2)**:320-332.

- Stevenson PH. 1924. Age order of epiphyseal union in man. *American Journal of Physical Anthropology* 7:53-93.
- Stewart TD. 1979. *Essentials of Forensic Anthropology, Especially as Developed in the United States*. Springfield, IL: Charles C. Thomas.
- Stewart TD. 1979b. Forensic Anthropology. In *The Use of Anthropology*, Goldschmidt W (ed). Washington, DC: American Anthropological Association; 169-183.
- Stout SD, Gehlert SJ. 1980. The relative accuracy and reliability of histological aging methods. *Forensic Science International* 15:181-190.
- Thieme FP, Schull WJ. 1957. Sex determination from the skeleton. *Human Biology* 29: 242-273.
- Thompson DD. 1979. The core technique in the determination of age at death of skeletons. *Journal of Forensic Science* 24:902-915.
- Todd T. 1920. Age Changes in the Pubic Bones: The White Male Pubis. *American Journal of Physical Anthropology* 3: 285-334.
- Todd TW, D'Errico J. 1928. The Clavicular epiphyses. *American Journal of Anatomy* 41:25-50.
- Ubelaker DH, Parra RC. 2011. Radiocarbon analysis of dental enamel and bone to evaluate date of birth and death: perspective from the southern hemisphere. *Forensic Science International* 208:103-107.
- Ubelaker DH. 1986. Estimation of age at death from histology of human bone. In *Dating and Age Determination of Biological Materials*, Zimmerman MR, Angel JL (eds). London: Croom Helm; 240-247.
- Ubelaker DH. 1990. Positive identification of American Indian skeletal remains from radiographic comparison. *Journal of Forensic Sciences* 35(2):466-472.
- Ubelaker DH. 1996. Skeleton testify: anthropology in forensic science. *Yearbook of physical anthropology* 39:229-244.

- Ubelaker DH. 1997. Forensic anthropology. In *History of Physical anthropology*, Vol 1 A-L, Spencer F (ed). New York, NY: Garland Publishing; 392-395.
- Ubelaker DH. 2000a. Methodological considerations in the forensic applications of human skeletal biology. In *Biological Anthropology of the Human Skeleton*, Katzenberg MA, Saunders SR (eds). New York: Wiley-Liss; 41-67.
- Ubelaker DH. 2000b. The Forensic anthropology legacy of T. Dale Stewart (1901-1998). *Journal of Forensic Sciences* **45(2)**:245-252.
- Ubelaker DH. 2000c. T. Dale Stewart's perspective on his career as a forensic anthropologist at the Smithsonian. *Journal of Forensic Sciences* **45(2)**:269-278.
- Ubelaker DH. 2006. Introduction to Forensic Anthropology. In *Forensic Anthropology and Medicine: Complementary Sciences from recovery to cause of death*, Schmitt A, Cunha E, Pinheiro J (eds). Totowa, NJ: Humana Press; 3-12.
- Ubelaker DH. 2010. Issues in forensic anthropology. In *A Companion to Biological Anthropology*, Larsen CS (ed). West Sussex: Wiley-Blackwell; 412-426.
- Utermohle CJ, Zegura SL, Heathcote GM. 1983. Multiple Observers, Humidity, and choice of precision statistics: Factors influencing craniometric data quality. *American Journal of Physical Anthropology* **61**:85-95.
- Walker RA, Lovejoy CO. 1985. Radiographic changes in the clavicle and proximal femur and their use in the determination of skeletal age at death. *American Journal of Physical Anthropology* **68**:67-78.
- Wand MP, Jones MC. 1995. *Kernel Smoothing*. Chapman & Hall.
- Webb PAO, Suchey JM. 1985. Epiphyseal Union of the Anterior Iliac Crest and Medial Clavicle in a Modern Multiracial Sample of American Males and Females. *American Journal of Physical Anthropology* **68**:457-466.

Apêndice A

Imagens exemplificativas das fases de união

Imagem A.1 – Clavícula pertencente à Fase 1



Imagem A.2 – Clavícula pertencente à Fase 2/Imagem A.2.1 – Clavícula pertencente à Fase 2



Imagem A.3 – Clavícula pertencente à Fase 3/Imagem A.3.1 – Clavícula pertencente à Fase 3



Imagem A.4 – Clavícula pertencente à Fase 4/Imagem A.4.1 – Clavícula pertencente à Fase 4



Apêndice B
Resultados

Tabela B1

Valores do Rácio obtidos por indivíduo

Nº Identificação	Ano	Sexo	Idade de Morte	Colecção	Rácio 1
12	1910	F	10	Coimbra	0
218	1926	F	10	Coimbra	0
265	1927	F	10	Coimbra	0
295	1928	F	10	Coimbra	0
346	1929	F	10	Coimbra	0
453	1932	M	10	Coimbra	0
50	1911	F	11	Coimbra	0
100A	1926	M	11	Coimbra	0
195	1926	F	11	Coimbra	0
365	1947	F	11	Lisboa	0
516	1919	F	11	Lisboa	0
101	1926	F	12	Coimbra	0
278	1927	F	12	Coimbra	0
411	1931	F	12	Coimbra	0
291	1934	M	13	Lisboa	0
267	1927	F	14	Coimbra	0
83	1931	F	14	Lisboa	0
193	1926	M	15	Coimbra	0
204	1947	F	15	Lisboa	0
336	1945	M	15	Lisboa	0
380	1955	M	15	Lisboa	0
452	1903	F	15	Lisboa	0
6	1904	M	16	Coimbra	0
111	1926	M	16	Coimbra	0
412	1931	M	16	Coimbra	0
258	1923	M	16	Lisboa	0
1566	1933	F	16	Lisboa	0
63	1912	F	17	Coimbra	0
226	1926	M	17	Coimbra	0
271	1927	F	17	Coimbra	0
284	1927	F	17	Coimbra	0
472	1933	F	17	Coimbra	0
1568	1944	M	17	Lisboa	0
1570	1946	M	17	Lisboa	0
269	1927	F	18	Coimbra	0
8	1931	F	18	Lisboa	0
632	1941	F	18	Lisboa	0
1403	1949	F	18	Lisboa	,3745
1418	1936	F	18	Lisboa	,5773
1587	1944	M	18	Lisboa	0
1588	1942	F	18	Lisboa	0
51	1911	F	19	Coimbra	0
117	1926	F	19	Coimbra	0
152	1926	F	19	Coimbra	0
234	1927	F	19	Coimbra	0
349	1929	F	19	Coimbra	0

405	1931	M	19	Coimbra	0
467	1933	M	19	Coimbra	0
498	1936	M	19	Coimbra	0
39	1924	F	19	Lisboa	0
136	1949	M	19	Lisboa	0
394	1941	M	19	Lisboa	,4404
1565	1929	F	19	Lisboa	0
97	1926	M	20	Coimbra	0
124	1926	M	20	Coimbra	,2493
189	1926	F	20	Coimbra	0
363	1930	M	20	Coimbra	,2405
221	1941	F	20	Lisboa	0
301	1952	M	20	Lisboa	,1804
314	1940	F	20	Lisboa	,7309
391	1945	M	20	Lisboa	,5306
1417	1958	F	20	Lisboa	,3854
1569	1955	M	20	Lisboa	,3310
1626	1940	M	20	Lisboa	0
45	1911	M	21	Coimbra	0
67	1912	F	21	Coimbra	,4757
73	1914	F	21	Coimbra	,5582
84	1914	M	21	Coimbra	0
138	1926	M	21	Coimbra	0
160	1926	F	21	Coimbra	,6826
169	1926	F	21	Coimbra	,7257
361	1930	F	21	Coimbra	,4945
429	1932	F	21	Coimbra	0
465	1933	F	21	Coimbra	0
361	1948	F	21	Lisboa	0
422	1944	M	21	Lisboa	,7397
748	1942	F	21	Lisboa	,1078
105	1926	F	22	Coimbra	,6066
252	1927	F	22	Coimbra	0
299	1928	M	22	Coimbra	,6535
364	1930	F	22	Coimbra	,8395
496	1925	F	22	Lisboa	0
1571	1930	F	22	Lisboa	,9349
47	1911	M	23	Coimbra	0
205	1926	M	23	Coimbra	0
222	1926	F	23	Coimbra	0
340	1929	M	23	Coimbra	,3868
371	1930	F	23	Coimbra	0
440	1932	M	23	Coimbra	,6142
486	1934	F	23	Coimbra	,5728
500	1936	M	23	Coimbra	,8297
54	1934	M	23	Lisboa	,5200
177	1925	F	23	Lisboa	1
474	1928	M	23	Lisboa	,7129
501	1934	M	23	Lisboa	,8517
1598	1948	F	23	Lisboa	,7345

98	1926	M	24	Coimbra	,9249
315	1928	F	24	Coimbra	,7818
382	1930	M	24	Coimbra	,7818
445	1932	F	24	Coimbra	1
461	1932	F	24	Coimbra	1
725	1949	F	24	Lisboa	,8677
42	1911	F	25	Coimbra	1
78	1915	M	25	Coimbra	1
142	1926	F	25	Coimbra	,9000
273	1927	F	25	Coimbra	1
468	1944	F	25	Lisboa	,9854
495	1943	M	25	Lisboa	,9409
1596	1944	F	25	Lisboa	1
131	1920	F	26	Coimbra	,8170
177	1926	F	26	Coimbra	1
179	1926	M	26	Coimbra	,9719
246	1927	M	26	Coimbra	1
313	1928	F	26	Coimbra	1
375	1930	M	26	Coimbra	1
403	1931	M	26	Coimbra	1
428	1932	M	26	Coimbra	0
459	1932	M	26	Coimbra	,8995
488	1934	M	26	Coimbra	1
493	1936	F	26	Coimbra	1
287	1927	M	27	Coimbra	1
328	1928	F	27	Coimbra	1
387	1930	F	27	Coimbra	1
487	1934	M	27	Coimbra	,9440
501	1932	M	27	Coimbra	1
276	1945	F	27	Lisboa	,9113
344	1937	M	27	Lisboa	1
128	1920	M	28	Coimbra	,6618
231	1921	M	28	Coimbra	1
251	1927	F	28	Coimbra	1
342	1929	F	28	Coimbra	1
359	1930	M	28	Coimbra	1
385	1930	F	28	Coimbra	1
395	1931	M	28	Coimbra	1
115	1929	F	28	Lisboa	1
472	1933	M	28	Lisboa	,9832
99	1926	F	29	Coimbra	1
144	1926	F	29	Coimbra	1
146	1926	M	29	Coimbra	1
199	1926	M	29	Coimbra	1
200	1926	F	29	Coimbra	1
240	1927	M	29	Coimbra	1
347	1929	M	29	Coimbra	1
434	1932	F	29	Coimbra	1
247	1940	F	29	Lisboa	1
2	1910	F	30	Coimbra	1

14	1910	M	30	Coimbra	1
69	1912	F	30	Coimbra	1
86	1926	M	30	Coimbra	1
139	1926	M	30	Coimbra	1
221	1926	M	30	Coimbra	1
286	1927	M	30	Coimbra	1
454	1932	F	30	Coimbra	1
294	1954	M	30	Lisboa	1
305	1937	M	30	Lisboa	,9712
309	1942	M	30	Lisboa	1
152	1934	M	30	Lisboa	1
1617	1944	M	30	Lisboa	1

Nota. F: Feminino; M: Masculino

Tabela B2

Resultados obtidos para Regressão Linear

Rácio	IdadeR	IdadeE	Erro	95%				
				LB	UB	AIC	Cobertura	AMIC
0,0000	10	16,89	-6,889	11	23	12,0	97,01%	12,149
0,0000	11	16,88	-5,876	11	23	12,0		
0,0000	12	16,86	-4,862	11	23	12,0		
0,0000	13	16,85	-3,848	11	23	12,0		
0,0000	14	16,83	-2,834	11	23	12,0		
0,0000	15	16,82	-1,821	11	23	12,0		
0,0000	16	16,81	-0,807	11	23	12,0		
0,0000	17	16,79	0,207	11	23	12,0		
0,0000	18	16,78	1,221	11	23	12,0		
0,0000	19	16,77	2,235	11	23	12,0		
0,0000	20	16,75	3,248	11	23	12,0		
0,0000	21	16,74	4,262	11	23	12,0		
0,0000	22	16,72	5,276	11	23	12,0		
0,0000	23	16,71	6,290	11	23	12,0		
0,0000	26	16,67	9,331	11	23	12,0		
0,1078	21	17,87	3,127	12	24	12,0		
0,1804	20	18,64	1,359	12	25	13,0		
0,2405	20	19,27	0,733	13	25	12,0		
0,2493	20	19,36	0,642	13	25	12,0		
0,3310	20	20,21	-0,208	14	26	12,0		
0,3745	18	20,67	-2,671	15	27	12,0		
0,3854	20	20,77	-0,771	15	27	12,0		
0,3868	23	20,77	2,235	15	27	12,0		
0,4404	19	21,35	-2,347	15	27	12,0		
0,4757	21	21,70	-0,701	16	28	12,0		
0,4945	21	21,90	-0,895	16	28	12,0		
0,5200	23	22,15	0,853	16	28	12,0		
0,5306	20	22,28	-2,277	16	28	12,0		
0,5582	21	22,56	-1,556	16	29	13,0		
0,5728	23	22,69	0,306	17	29	12,0		
0,5773	18	22,77	-4,774	17	29	12,0		
0,6066	22	23,05	-1,053	17	29	12,0		
0,6142	23	23,12	-0,124	17	29	12,0		
0,6535	22	23,54	-1,539	17	30	13,0		
0,6618	28	23,58	4,419	17	30	13,0		
0,6826	21	23,85	-2,849	18	30	12,0		
0,7129	23	24,15	-1,149	18	30	12,0		
0,7257	21	24,30	-3,298	18	30	12,0		
0,7309	20	24,36	-4,361	18	30	12,0		
0,7345	23	24,37	-1,374	18	30	12,0		
0,7397	21	24,44	-3,445	18	31	13,0		
0,7818	24	24,86	-0,856	19	31	12,0		
0,7818	24	24,86	-0,856	19	31	12,0		
0,8170	26	25,20	0,796	19	31	12,0		
0,8297	23	25,37	-2,366	19	31	12,0		

0,8395	22	25,48	-3,480	19	32	13,0
0,8517	23	25,60	-2,597	19	32	13,0
0,8677	24	25,75	-1,753	20	32	12,0
0,8995	26	26,06	-0,062	20	32	12,0
0,9000	25	26,08	-1,080	20	32	12,0
0,9113	27	26,17	0,827	20	32	12,0
0,9249	24	26,35	-2,352	20	32	12,0
0,9349	22	26,48	-4,483	20	33	13,0
0,9409	25	26,51	-1,508	20	33	13,0
0,9440	27	26,51	0,486	20	33	13,0
0,9712	30	26,76	3,245	21	33	12,0
0,9719	26	26,82	-0,819	21	33	12,0
0,9832	28	26,91	1,092	21	33	12,0
0,9854	25	26,97	-1,974	21	33	12,0
1,0000	23	27,16	-4,158	21	33	12,0
1,0000	24	27,14	-3,143	21	33	12,0
1,0000	25	27,13	-2,128	21	33	12,0
1,0000	26	27,11	-1,113	21	33	12,0
1,0000	27	27,10	-0,099	21	33	12,0
1,0000	28	27,08	0,916	21	33	12,0
1,0000	29	27,07	1,931	21	33	12,0
1,0000	30	27,05	2,946	21	33	12,0

Nota. IdadeR = Idade real; IdadeE = Idade Estimada; LI = Limite Inferior do Intervalo de confiança; LS = Limite Superior do Intervalo de Confiança; AIC = Amplitude do Intervalo de Confiança; AMIC = Média da Amplitude do Intervalo de Confiança.

Tabela B3

Tabela de resultados para a Calibração Clássica

Rácio	IdadeR	IdadeE	Erro	95%				
				LI	LS	AIC	Cobertura	AMIC
0,0000	10	14,75	-4,746	14,00	15,00	1,000	18,18%	0,985
0,0000	11	14,73	-3,734	14,00	15,00	1,000		
0,0000	12	14,71	-2,708	14,00	15,00	1,000		
0,0000	13	14,67	-1,671	14,00	15,00	1,000		
0,0000	14	14,64	-0,644	14,00	15,00	1,000		
0,0000	15	14,65	0,353	14,00	15,00	1,000		
0,0000	16	14,62	1,384	14,00	15,00	1,000		
0,0000	17	14,62	2,385	14,00	15,00	1,000		
0,0000	18	14,60	3,398	14,00	15,00	1,000		
0,3745	18	20,16	-2,164	20,00	21,00	1,000		
0,5773	18	23,16	-5,161	23,00	24,00	1,000		
0,0000	19	14,56	4,438	14,00	15,00	1,000		
0,4404	19	21,13	-2,127	21,00	22,00	1,000		
0,0000	20	14,55	5,455	14,00	15,00	1,000		
0,2493	20	18,27	1,734	18,00	19,00	1,000		
0,2405	20	18,13	1,867	18,00	19,00	1,000		
0,1804	20	17,26	2,742	17,00	18,00	1,000		
0,7309	20	25,46	-5,456	25,00	26,00	1,000		
0,5306	20	22,46	-2,455	22,00	23,00	1,000		
0,3854	20	20,31	-0,308	20,00	21,00	1,000		
0,3310	20	19,49	0,507	19,00	20,00	1,000		
0,0000	21	14,56	6,440	14,00	15,00	1,000		
0,4757	21	21,65	-0,647	21,00	22,00	1,000		
0,5582	21	22,88	-1,878	22,00	23,00	1,000		
0,6826	21	24,74	-3,739	24,00	25,00	1,000		
0,7257	21	25,34	-4,344	25,00	26,00	1,000		
0,4945	21	21,93	-0,927	21,00	22,00	1,000		
0,7397	21	25,55	-4,553	25,00	26,00	1,000		
0,1078	21	16,17	4,826	16,00	17,00	1,000		
0,6066	22	23,59	-1,585	23,00	24,00	1,000		
0,6535	22	24,28	-2,283	24,00	25,00	1,000		
0,8395	22	27,05	-5,049	27,00	28,00	1,000		
0,9349	22	28,47	-6,467	28,00	29,00	1,000		
0,0000	23	14,57	8,427	14,00	15,00	1,000		
0,3868	23	20,32	2,678	20,00	21,00	1,000		
0,6142	23	23,69	-0,689	23,00	24,00	1,000		
0,5728	23	23,08	-0,077	23,00	24,00	1,000		
0,8297	23	26,92	-3,920	26,00	27,00	1,000		
0,5200	23	22,30	0,705	22,00	23,00	1,000		
1,0000	23	29,45	-6,446	29,00	30,00	1,000		
0,7129	23	25,15	-2,151	25,00	26,00	1,000		
0,8517	23	27,25	-4,246	27,00	28,00	1,000		
0,7345	23	25,47	-2,471	25,00	26,00	1,000		
0,9249	24	28,31	-4,312	28,00	29,00	1,000		
0,7818	24	26,20	-2,198	26,00	27,00	1,000		

0,7818	24	26,20	-2,198	26,00	27,00	1,000
1,0000	24	29,46	-5,464	29,00	30,00	1,000
0,8677	24	27,47	-3,466	27,00	28,00	1,000
1,0000	25	29,45	-4,448	29,00	30,00	1,000
0,9000	25	27,94	-2,935	27,00	28,00	1,000
0,9854	25	29,23	-4,233	29,00	30,00	1,000
0,9409	25	29,24	-4,238	29,00	30,00	1,000
1,0000	26	29,44	-3,441	29,00	30,00	1,000
0,8170	26	26,65	-0,646	26,00	27,00	1,000
0,9719	26	29,03	-3,028	29,00	30,00	1,000
0,0000	26	14,59	11,406	15,00	15,00	0,000
0,8995	26	27,93	-1,926	27,00	28,00	1,000
1,0000	27	29,44	-2,442	29,00	30,00	1,000
0,9440	27	29,41	-2,412	29,00	30,00	1,000
0,9113	27	28,11	-1,106	28,00	29,00	1,000
0,6618	28	24,36	3,638	24,00	25,00	1,000
1,0000	28	29,41	-1,409	29,00	30,00	1,000
0,9832	28	29,17	-1,166	29,00	30,00	1,000
1,0000	29	29,38	-0,385	29,00	30,00	1,000
1,0000	30	29,37	0,629	29,00	30,00	1,000
0,9712	30	28,96	1,042	28,00	29,00	1,000

Nota. IdadeR = Idade real; IdadeE = Idade Estimada; LI = Limite Inferior do Intervalo de confiança; LS = Limite Superior do Intervalo de Confiança; AIC = Amplitude do Intervalo de Confiança; AMIC = Média da Amplitude do Intervalo de Confiança

Tabela B4

Tabela de resultados para o Modelo 1

Rácio	IdadeR	Fase	IdadeE	Erro	95%				
					LI	LS	AIC	Cobertura	AMIC
0,000	10	1	17,24	-7,24	9	23	14	100,00%	13,26
0,000	11	1	17,24	-6,24	9	23	14		
0,000	12	1	17,24	-5,24	9	23	14		
0,000	13	1	17,24	-4,24	9	23	14		
0,000	14	1	17,24	-3,24	9	23	14		
0,000	15	1	17,24	-2,24	9	23	14		
0,000	16	1	17,24	-1,24	9	23	14		
0,000	17	1	17,24	-0,24	9	23	14		
0,000	18	1	17,07	0,93	9	23	14		
0,000	19	1	17,07	1,93	9	23	14		
0,440	19	2	18,76	0,24	10	26	16		
0,000	20	1	17,07	2,93	9	23	14		
0,249	20	2	18,76	1,24	10	26	16		
0,731	20	3	24,37	-4,37	17	30	13		
0,000	21	1	17,07	3,93	9	23	14		
0,476	21	2	18,76	2,24	10	26	16		
0,558	21	3	24,37	-3,37	17	30	13		
0,607	22	3	24,37	-2,37	17	30	13		
0,000	22	1	17,07	4,93	9	23	14		
0,935	22	4	26,75	-4,75	18	30	12		
0,000	23	1	17,07	5,93	9	23	14		
0,387	23	2	18,76	4,24	10	26	16		
0,614	23	3	24,37	-1,37	17	30	13		
1,000	23	4	26,75	-3,75	19	30	11		
0,925	24	4	26,74	-2,74	18	30	12		
0,782	24	3	24,22	-0,22	17	30	13		
1,000	25	4	26,74	-1,74	18	30	12		
0,817	26	3	24,22	1,78	17	30	13		
1,000	26	4	26,74	-0,74	18	30	12		
1,000	27	4	26,74	0,26	18	30	12		
0,662	28	3	24,15	3,85	17	30	13		
1,000	28	4	26,60	1,40	18	30	12		
1,000	29	4	26,60	2,40	18	30	12		
1,000	30	4	26,60	3,40	18	30	12		

Nota. IdadeR = Idade real; IdadeE = Idade Estimada; LI = Limite Inferior do Intervalo de confiança; LS = Limite Superior do Intervalo de Confiança; AIC = Amplitude do Intervalo de Confiança; AMIC = Média da Amplitude do Intervalo de Confiança

Tabela B5

Tabela de resultados para o Modelo 2

Rácio	IdadeR	Fase	IdadeE	Erro	95%				
					LI	LS	AIC	Cobertura	AMIC
0,000	10	1	16,75	-6,75	9,00	24,00	15,00		
0,000	11	1	16,75	-5,75	9,00	24,00	15,00		
0,000	12	1	16,75	-4,75	9,00	24,00	15,00		
0,000	13	1	16,75	-3,75	9,00	24,00	15,00		
0,000	14	1	16,75	-2,75	9,00	24,00	15,00		
0,000	15	1	16,75	-1,75	9,00	24,00	15,00		
0,000	16	1	16,75	-0,75	9,00	24,00	15,00		
0,000	17	1	16,75	0,25	9,00	24,00	15,00		
0,000	18	1	16,57	1,43	9,00	24,00	15,00		
0,374	18	2	18,47	-0,47	10,00	26,00	16,00		
0,577	18	3	24,84	-6,84	12,00	30,00	18,00		
0,000	19	1	16,57	2,43	9,00	24,00	15,00		
0,440	19	2	18,47	0,53	10,00	26,00	16,00		
0,000	20	1	16,57	3,43	9,00	24,00	15,00		
0,249	20	2	18,47	1,53	10,00	26,00	16,00		
0,731	20	3	24,84	-4,84	12,00	30,00	18,00		
0,000	21	1	16,57	4,43	9,00	24,00	15,00		
0,476	21	2	18,47	2,53	10,00	26,00	16,00		
0,558	21	3	24,84	-3,84	12,00	30,00	18,00		
0,607	22	3	24,84	-2,84	12,00	30,00	18,00		
0,000	22	1	16,57	5,43	9,00	24,00	15,00		
0,935	22	4	26,76	-4,76	14,00	30,00	16,00		
0,000	23	1	16,57	6,43	9,00	24,00	15,00		
0,387	23	2	18,47	4,53	10,00	26,00	16,00		
0,000	23	1	16,57	6,43	9,00	24,00	15,00		
0,614	23	3	24,84	-1,84	12,00	30,00	18,00		
1,000	23	4	26,76	-3,76	14,00	30,00	16,00		
0,925	24	4	26,74	-2,74	14,00	30,00	16,00		
0,782	24	3	24,76	-0,76	12,00	30,00	18,00		
1,000	25	4	26,74	-1,74	14,00	30,00	16,00		
0,817	26	3	24,76	1,24	12,00	30,00	18,00		
1,000	26	4	26,74	-0,74	14,00	30,00	16,00		
1,000	27	4	26,74	0,26	14,00	30,00	16,00		
0,662	28	3	24,69	3,31	12,00	30,00	18,00		
1,000	28	4	26,64	1,36	14,00	30,00	16,00		
1,000	29	4	26,64	2,36	14,00	30,00	16,00		
1,000	30	4	26,64	3,36	14,00	30,00	16,00		

Nota. IdadeR = Idade real; IdadeE = Idade Estimada; LI = Limite Inferior do Intervalo de confiança; LS = Limite Superior do Intervalo de Confiança; AIC = Amplitude do Intervalo de Confiança; AMIC = Média da Amplitude do Intervalo de Confiança

Tabela B6

Tabela de resultados obtidos para a versão discreta do teorema de Bayes

Rácio	IdadeR	Fase	IdadeE	Erro	95%				
					LI	LS	AIC	Cobertura	AMIC
0,0000	10	1	16,88	-6,879	10	22	12,000	95,45%	8,879
0,0000	11	1	16,88	-5,879	10	22	12,000		
0,0000	12	1	16,88	-4,879	10	22	12,000		
0,0000	13	1	16,88	-3,879	10	22	12,000		
0,0000	14	1	16,88	-2,879	10	22	12,000		
0,0000	15	1	16,88	-1,879	10	22	12,000		
0,0000	16	1	16,88	-0,879	10	22	12,000		
0,0000	17	1	16,88	0,121	10	22	12,000		
0,0000	18	1	16,65	1,353	10	22	12,000		
0,3745	18	2	20,00	-2,000	17	23	6,000		
0,5773	18	3	21,13	-3,125	17	27	10,000		
0,0000	19	1	16,65	2,353	10	22	12,000		
0,4404	19	2	20,00	-1,000	17	23	6,000		
0,0000	20	1	16,65	3,353	10	22	12,000		
0,2493	20	2	20,00	0,000	17	23	6,000		
0,2405	20	2	20,00	0,000	17	23	6,000		
0,1804	20	2	20,00	0,000	17	23	6,000		
0,7309	20	3	21,13	-1,125	17	27	10,000		
0,5306	20	3	21,13	-1,125	17	27	10,000		
0,3854	20	2	20,00	0,000	17	23	6,000		
0,3310	20	2	20,00	0,000	17	23	6,000		
0,0000	21	1	16,65	4,353	10	22	12,000		
0,4757	21	2	20,00	1,000	17	23	6,000		
0,5582	21	3	21,13	-0,125	17	27	10,000		
0,6826	21	3	21,13	-0,125	17	27	10,000		
0,7257	21	3	21,13	-0,125	17	27	10,000		
0,4945	21	2	20,00	1,000	17	23	6,000		
0,7397	21	3	21,13	-0,125	17	27	10,000		
0,1078	21	2	20,00	1,000	17	23	6,000		
0,6066	22	3	21,13	0,875	17	27	10,000		
0,0000	22	1	16,65	5,353	10	22	12,000		
0,6535	22	3	21,13	0,875	17	27	10,000		
0,8395	22	3	21,13	0,875	17	27	10,000		
0,9349	22	4	27,20	-5,200	23	30	7,000		
0,0000	23	1	16,65	6,353	10	22	12,000		
0,3868	23	2	20,00	3,000	17	23	6,000		
0,6142	23	3	21,13	1,875	17	27	10,000		
0,5728	23	3	21,13	1,875	17	27	10,000		
0,8297	23	3	21,13	1,875	17	27	10,000		
0,5200	23	3	21,13	1,875	17	27	10,000		
1,0000	23	4	27,20	-4,200	23	27	4,000		
0,7129	23	3	21,13	1,875	17	27	10,000		
0,8517	23	3	21,13	1,875	17	27	10,000		
0,7345	23	3	21,13	1,875	17	27	10,000		
0,9249	24	4	27,20	-3,200	23	30	7,000		

0,7818	24	3	20,88	3,118	17	27	10,000
0,7818	24	3	20,88	3,118	17	27	10,000
1,0000	24	4	27,20	-3,200	23	30	7,000
0,8677	24	3	20,88	3,118	17	27	10,000
1,0000	25	4	27,20	-2,200	23	30	7,000
0,9000	25	4	27,20	-2,200	23	30	7,000
0,9854	25	4	27,20	-2,200	23	30	7,000
0,9409	25	4	27,20	-2,200	23	30	7,000
0,8170	26	3	20,88	5,118	17	27	10,000
1,0000	26	4	27,20	-1,200	23	30	7,000
0,9719	26	4	27,20	-1,200	23	30	7,000
0,8995	26	3	20,88	5,118	17	27	10,000
1,0000	27	4	27,20	-0,200	23	30	7,000
0,9440	27	4	27,20	-0,200	23	30	7,000
0,9113	27	4	27,20	-0,200	23	30	7,000
0,6618	28	3	20,88	7,118	17	26	9,000
1,0000	28	4	27,10	0,897	23	30	7,000
0,9832	28	4	27,10	0,897	23	30	7,000
1,0000	29	4	27,10	1,897	23	30	7,000
1,0000	30	4	27,10	2,897	23	30	7,000
0,9712	30	4	27,10	2,897	23	30	7,000

Nota. IdadeR = Idade real; IdadeE = Idade Estimada; LI = Limite Inferior do Intervalo de confiança; LS = Limite Superior do Intervalo de Confiança; AIC = Amplitude do Intervalo de Confiança; AMIC = Média da Amplitude do Intervalo de Confiança.

Tabela B7

Resultados obtidos para a amostra reduzida com o Modelo 2

Rácio	IdadeR	Fase	IdadeE	Erro	95%				
					LI	LS	AIC	Cobertura	AMIC
0,000	10	1	13,84	-3,84	9,24	24,11	14,87	89,47%	13,69
0,000	13	1	13,84	-0,84	9,24	24,11	14,87		
0,000	14	1	13,84	0,16	9,24	24,11	14,87		
0,000	15	1	13,84	1,16	9,24	24,11	14,87		
0,000	16	1	13,84	2,16	9,24	24,11	14,87		
0,440	19	2	18,72	0,28	9,55	26,39	16,84		
0,331	20	2	18,72	1,28	9,55	26,39	16,84		
0,108	21	2	18,72	2,28	9,55	26,39	16,84		
0,740	21	3	24,17	-3,17	17,38	29,69	12,31		
0,607	22	3	24,17	-2,17	17,38	29,69	12,31		
0,573	23	3	24,17	-1,17	17,38	29,69	12,31		
0,868	24	3	23,32	0,68	17,13	29,67	12,54		
1,000	24	4	27,41	-3,41	17,70	29,87	12,17		
0,817	26	3	23,32	2,68	17,13	29,67	12,54		
0,972	26	4	27,41	-1,41	17,70	29,87	12,17		
0,983	28	4	26,59	1,41	17,62	29,84	12,22		
1,000	29	4	26,59	2,41	17,62	29,84	12,22		
1,000	30	4	26,59	3,41	17,62	29,84	12,22		
1,000	30	4	26,59	3,41	17,62	29,84	12,22		

Nota. IdadeR = Idade real; IdadeE = Idade Estimada; LI = Limite Inferior do Intervalo de confiança; LS = Limite Superior do Intervalo de Confiança; AIC = Amplitude do Intervalo de Confiança; AMIC = Média da Amplitude do Intervalo de Confiança

Apêndice C

Desenvolvimento de equações

Desenvolvimento da equação (2.3)

Sendo:

$$\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}$$

temos que,

$$s_{xy} = \frac{1}{n} \sum(x_i - \bar{x})(y_i - \bar{y})$$

e,

$$s_x^2 = \frac{1}{n} \sum(x_i - \bar{x})^2$$

Desenvolvimento da equação (2.7)

Sendo:

$$\begin{aligned} \hat{d} &= \frac{\sum(y_i - \bar{y}) - (e_i - \bar{e})}{\sum(y_i - \bar{y})^2} \\ &= \frac{\sum(y_i - \bar{y}) \left\{ y_i - \bar{y} - \frac{s_{xy}}{s_x^2} (x_i - \bar{x}) \right\}}{\sum(y_i - \bar{y})^2} \\ &= \frac{\sum(y_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} - \frac{s_{xy}}{s_x^2} \frac{\sum(y_i - \bar{y}) - (x_i - \bar{x})}{\sum(y_i - \bar{y})^2} \\ &= 1 - \frac{s_{xy}}{s_x^2} \frac{s_{xy}}{s_y^2} \\ &= 1 - r_{xy}^2 \end{aligned}$$

Apêndice D

Rotinas em R

Rotina base utilizada para o cálculo das funções de densidade de probabilidades – Modelo 1

```
dist2d<-function(K,L){
  N<-K*L
  prob<-runif(N)
  soma<-cumsum(prob)[N]
  prob<-prob/soma
  coorx<-seq(1,K,by=1)/K
  coory<-seq(1,L,by=1)/L
  prob1<-matrix(prob,K,L)
  persp(coorx,coory,prob1,phi=45,shade=0.6)
  return(prob1)
}

margdist<-function(prob2d){
  # calcule le distribution marginale selon les lignes
  K<-nrow(prob2d)
  L<-ncol(prob2d)
  prow<-NULL
  for (i in 1:K){
    prow[i]<-cumsum(prob2d[i,])[L]
  }
  return(prow)
}

epanech_disc<-function(m){
  # m <- nombre de points (impair dans nos modèles)
  ker<-NULL
  for (j in 1:m){
    ker[j]<-3/(2*m)-((-1+2*j/m)^3-(-1+2*(j-1)/m)^3)/4
  }
  return(ker)
}

somaquad<-function(obs,ker1,ker2){
  # calcule la somme des éléments de la matrice obs
  # avec les poids ker1, ker2
  # obs à la dimension définie par les supports
  # des fonctions de poids
  # h1, h2 sont les vraies longuers des supports
  # des fonctions de poids
  h1<-length(ker1)
  h2<-length(ker2)
  soma<-0
  for (i in 1:h2){
```



```

for (j in 1:h1){
  if (!is.null(dim(obs))) {soma<-soma+obs[i,j]*ker1[j]*ker2[i]}
  else {soma<-soma+obs[j]*ker1[j]*ker2[i]}
}
}
return(soma)
} # fin de somaquad

```

```

penpolsmooth<-function(p,y,ker1,ker2,marg=NULL){
# régression polynomiale pénalisée
# p <- degré du polynome
# y <- matrice des proportions
# ker1, ker2 <- fonction de poids
# marg <- distribution marginale connue
# K <- nombre de lignes
# L <- nombre de colonnes
K<-nrow(y)
L<-ncol(y)
# h1 <- demi-longuer du support du noyau ker1
# h2 <- demi-longuer du support du noyau ker2
h1<-(length(ker1)-1)/2
h2<-(length(ker2)-1)/2
# yrep <- matrice élargie des proportions
yrep<-enlarge(y)
# calcul des moments des fonctions de poids
if (p>=1)
{sigma1<-0
if (p==2)
{tau1<-0
tau2<-0
}
}
for (i in 1:h1){
sigma1<-sigma1+2*i^2*ker1[h1+1+i]
if (p==2){tau1<-tau1+2*i^4*ker1[h1+1+i]}
}
sigma2<-0
for (i in 1:h2){
sigma2<-sigma2+2*i^2*ker2[h2+1+i]
if (p==2){tau2<-tau2+2*i^4*ker2[h2+1+i]}
}
} # termine le if p>=1
if (p==2){delta<-tau1*tau2-sigma1^2*sigma2^2}
pchap<-array(0,c(K,L))
for (i in 1:K){
sumrow<-0
for (j in 1:L){
# pour le terme correspondant à p=0
# toujours présent
tmp1<- (yrep*yrep)[(K+i-h2):(K+i+h2),(L+j-h1):(L+j+h1)]

```

```

tmp<-yrep[(K+i-h2):(K+i+h2),(L+j-h1):(L+j+h1)]
pchap[i,j]<-somaquad(tmp1,ker1,ker2)
if (p>=1)
  {# ker1p1, ker2p1 fonctions de poids correspondant au terme p=1
  ker1m<-changeweight(1,ker1)
  ker2m<-changeweight(1,ker2)
  pchap[i,j]<-pchap[i,j]-(somaquad(tmp,ker1m,ker2))^2/sigma1-
(somaquad(tmp,ker1,ker2m))^2/sigma2
  }
if (p==2)
  {# il y a quatre termes a àjouter dont le dernier
  # est encore un produit de sommes sur le carré
  # autour du point (i,j)
  # premier terme: modifie seulement ker1 avec p=2
  ker1m<-changeweight(2,ker1)
  pchap[i,j]<-pchap[i,j]-(somaquad(tmp,ker1m,ker2))^2*(tau2/delta)
  # deuxième terme: modifie seulement ker2 avec p=2
  ker2m<-changeweight(2,ker2)
  pchap[i,j]<-pchap[i,j]-(somaquad(tmp,ker1,ker2m))^2*(tau1/delta)
  # troisième terme: modifie les deux fonctions de poids avec p=1
  ker1m<-changeweight(1,ker1)
  ker2m<-changeweight(1,ker2)
  pchap[i,j]<-pchap[i,j]-(somaquad(tmp,ker1m,ker2m))^2/(sigma1*sigma2)
  # quatrième terme: modifie les deux fonctions de poids avec p=2
  ker1m<-changeweight(2,ker1)
  ker2m<-changeweight(2,ker2)
  pchap[i,j]<-pchap[i,j]-
2*sigma1*sigma2*somaquad(tmp,ker1m,ker2)*somaquad(tmp,ker1,ker2m)/delta
  } # fin du if p==2
  pchap[i,j]<-sqrt(abs(pchap[i,j]))
  sumrow<-sumrow+pchap[i,j]
  }
  if(!is.null(marg)){pchap[i,]<-marg[i]*pchap[i,]/sumrow}
  }
if (is.null(marg)) {pchap<-pchap/sum(pchap)}
return(pchap)
} # fin de polsmooth

```

```

polsmooth<-function(p,y,ker1,ker2,marg=NULL){
  # régression polynômiale
  # p <- degré du polynome
  # y <- matrice des proportions
  # ker1, ker2 <- fonction de poids
  # marg <- distribution marginale connue
  # K <- nombre de lignes
  # L <- nombre de colonnes
  K<-nrow(y)
  L<-ncol(y)
  # h1 <- demi-longuer du support du noyau ker1

```

```

# h2 <- demi-longuer du support du noyau ker2
h1<-(length(ker1)-1)/2
h2<-(length(ker2)-1)/2
# yrep <- matrice élargie des proportions
yrep<-enlarge(y)
# calcul des moments des fonctions de poids
if (p>=1)
{sigma1<-0
if (p==2)
{tau1<-0
tau2<-0
}
for (i in 1:h1){
sigma1<-sigma1+2*i^2*ker1[h1+1+i]
if (p==2){tau1<-tau1+2*i^4*ker1[h1+1+i]}
}
sigma2<-0
for (i in 1:h2){
sigma2<-sigma2+2*i^2*ker2[h2+1+i]
if (p==2){tau2<-tau2+2*i^4*ker2[h2+1+i]}
}
} # termine le if p>=1
if (p==2){delta<-tau1*tau2-sigma1^2*sigma2^2}
pchap<-array(0,c(K,L))
for (i in 1:K){
sumrow<-0
for (j in 1:L){
# pour le terme correspondant à p=0 ou p=1
# toujours présent
# tmp1<-(yrep*yrep)[(K+i-h2):(K+i+h2),(L+j-h1):(L+j+h1)]
tmp<-yrep[(K+i-h2):(K+i+h2),(L+j-h1):(L+j+h1)]
pchap[i,j]<-somaquad(tmp,ker1,ker2)
if (p==2)
{# il y a trois termes à ajouter autour du point (i,j)
# premier terme: modifie le cas p=0 en multipliant par U
u<-delta/(((tau1-sigma1^2)*(tau2-sigma2^2))
pchap[i,j]<-pchap[i,j]*u
# deuxième terme: modifie seulement ker1 avec p=2
ker1m<-changeweight(2,ker1)
pchap[i,j]<-pchap[i,j]-somaquad(tmp,ker1m,ker2)*(sigma1^2/(tau1-sigma1^2))
# troisième terme: modifie seulement ker2 avec p=2)
ker2m<-changeweight(2,ker2)
pchap[i,j]<-pchap[i,j]-somaquad(tmp,ker1,ker2m)*(sigma2^2/(tau2-sigma2^2))
} # fin du if p==2
sumrow<-sumrow+pchap[i,j]
}
if (!is.null(marg)){pchap[i,j]<-pchap[i,j]+(marg[i]-sumrow)/L}
}
return(pchap)
} # fin de polsmooth

```

```

projnega<-function(p){
# efectua a projecção de uma distribuição para evitar
# componentes neativas, mantendo a distribuição marginal
K<-nrow(p)
L<-ncol(p)
p1<-p
for (i in 1:K){
  if (min(p[i,])<0){
    tmp<-cumsum(p[i,])[L]
    for (j in 1:L){
      if (p[i,j]>=0){p1[i,j]<-p[i,j]}
      else {p1[i,j]<-0}
    }
    tmp1<-cumsum(p1[i,])[L]
    p1[i,]<-p1[i,]*tmp/tmp1
  }
}
return(p1)
} #fim de projnega

distcond<-function(p){
# Função para calcular as distribuições condicionais
# de acordo com as colunas de uma distribuição bi-dimensional p
# K <- número de linhas
# L <- número de colunas
K<-nrow(p)
L<-ncol(p)
p1<-array(0,c(K,L))
# margcol<-array(0,c(L))
for (j in 1:L){
  margcol<-cumsum(p[,j])[K]
  for (i in 1:K){p1[i,j]<-p[i,j]/margcol}
}
return(p1)
} # fim de distcond

```

Rotina desenvolvida para o cálculo da distribuição marginal e sua inclusão no cálculo da probabilidade *a posteriori* – Modelo 2

```
aerts<-function(p,y,ker){  
  
# p <- grau do polinómio <= 2  
# y <- vector das proporções  
# ker <- função  
# k <- Número de casos  
k<-length(y)  
# yrep <- vector de proporções replicado  
yrep<-NULL  
for (i in 1:k){  
  yrep[k+1-i]<-y[i]  
}  
yrep[(k+1):(2*k)]<-y  
for (i in 1:k){  
  yrep[2*k+i]<-yrep[i]  
}
```

Rotina criada para a aplicação da técnica jacknife e para a criação das tabelas de distribuição das probabilidades *a posteriori* finais

```
aa<-matrix(scan(file='dados1.txt'),byrow=T,ncol=11)
```

```
faztabela<-function(a){  
k<-nrow(a)  
tt<-array(0,c(4,4))  
for (i in 1:k){  
  if(a[i,3]!=-1){  
    ll<-which(a[i,4:7]==1)  
    cc<-which(a[i,8:11]==1)  
    tt[ll,cc]<-tt[ll,cc]+1  
  }  
}  
return(tt)  
}
```

```
ee<-length(which(aa[,3]==-1))  
erro<-array(-1,164-ee)  
tt2<-array(0,4)  
conta<-1  
e<-c(9,17,23,27,30) #pontos finais de cada nível etário  
for (i in 1:164){  
  if (aa[i,3]!=-1){  
    bb<-array(0,c(163,11))  
    if ((i>=2)&&(i<=163)){
```

```

        bb[1:(i-1),]<-aa[1:(i-1),]
        bb[i:163,]<-aa[(i+1):164,]
    }
    if (i==1) bb[1:163,]<-aa[2:164,]
    if (i==164) bb[1:163,]<-aa[1:163,]
    tt1<-faztabela(bb)
    for (j in 1:4) tt2[j]<-sum(tt1[j,])
#   tt2aprox<-aerts(1,tt2/sum(tt2),epanech_disc(5))
    aprox<-polsmooth(0,tt1/sum(tt1),epanech_disc(3),epanech_disc(3))
#   aprox<-polsmooth(0,tt1/sum(tt1),epanech_disc(3),epanech_disc(3),tt2aprox)
    aprox1<-distcond(aprox)
    cc<-which(aa[i,8:11]==1)
    dd<-cumsum(aprox1[,cc])
    tmp<-sum(dd<=0.5)
    if (tmp==0){e_aprox<-e[tmp+1]+(e[tmp+2]-e[tmp+1])/(dd[tmp+1]-0)*(0.5-0)}
    else {e_aprox<-e[tmp+1]+(e[tmp+2]-e[tmp+1])/(dd[tmp+1]-dd[tmp])*(0.5-dd[tmp])}
    erro[conta]<-e_aprox-aa[i,2]
    conta<-conta+1
}
}

resultado_final<-function(){
    tt1<-faztabela(aa)
    for (j in 1:4) tt2[j]<-sum(tt1[j,])
#   tt2aprox<-aerts(1,tt2/sum(tt2),epanech_disc(5))
    aprox<-polsmooth(0,tt1/sum(tt1),epanech_disc(3),epanech_disc(3))
#   aprox<-polsmooth(0,tt1/sum(tt1),epanech_disc(3),epanech_disc(3),tt2aprox)
    aprox1<-distcond(aprox)
    return(aprox1)
}

```