

Metodologia STATIS Dual

Aplicação a dados sobre Infertilidade

por

Ângela Marisa Nordeste Félix de Almeida

Tese de Mestrado em Análise de Dados e Sistemas de Apoio à Decisão

Orientada por:

Prof. Doutora Adelaide Maria Sousa Figueiredo

Prof. Doutora Fernanda Otília Sousa Figueiredo

Ano 2012

Ao meu pai que, sem ter tido oportunidade de o saber, influenciou a maior parte do meu percurso acadêmico.

À minha mãe por ter permitido que este fosse concretizável.

NOTA BIOGRÁFICA

Ângela Marisa Nordeste Félix de Almeida nasceu na cidade de Aveiro no dia 8 de Dezembro de 1979 tendo vivido a sua infância e adolescência em Estarreja onde completou o ensino secundário no agrupamento 3: Sócio-Económico.

Em 1997 entra para a universidade e curso da sua preferência: Licenciatura em Ensino de Matemática na Universidade do Minho a qual conclui em 2003 com a realização de estágio na Escola E.B. 2,3/S de Melgaço.

Em 2004 frequentou, em Guimarães, o curso de Especialização em Técnicas de Apoio à Decisão: Estatística e Investigação Operacional, que lhe despertou o interesse pela área da análise de dados e data mining, e que lhe permitiu o primeiro contato com aquela que viria a ser a sua entidade empregadora.

Em Fevereiro de 2005 assina o seu primeiro contrato de trabalho na Maternidade de Júlio Dinis, no Porto, com a integração num projeto de certificação de qualidade que terminou com sucesso no fim de 2006. Durante esses dois anos foram-lhe atribuídas outras responsabilidades que manteve após termino do projeto, colaborando com o Serviço de Gestão e Planeamento. Em 2007 dá-se a integração de três unidades hospitalares: Hospital de Santo António, Maternidade de Júlio Dinis e Hospital Maria Pia tendo sido constituído o Centro Hospitalar do Porto, EPE. instituição onde desempenha funções na categoria de técnica superior no Serviço de Informação de Gestão até à atualidade.

AGRADECIMENTOS

O Mestrado em Análise de Dados e Sistemas de Apoio à Decisão (MADSAD) engloba uma área de conhecimento bastante interessante e em verdadeira ascensão. O corpo docente é constituído por profissionais exemplares capazes de expor as temáticas focando os aspetos essenciais das mesmas e fomentando a curiosidade dos mestrandos no aprofundar de determinadas questões.

Estes dois anos de trabalho e esforço pessoal intenso foram amenizados com o espírito de camaradagem vivido quer entre os colegas de mestrado quer entre estes e os docentes. Acredito que este conjunto de pessoas foi fundamental para conseguir levar o projeto até ao fim.

Expresso especial agradecimento:

Às orientadoras, Professora Doutora Adelaide Figueiredo e Professora Doutora Fernanda Figueiredo, pelas sugestões, críticas e apoio ao longo de todo o percurso, por se mostrarem sempre disponíveis, pela motivação e profissionalismo com que sempre trataram as minhas questões e, acima de tudo, pela compreensão.

Ao grupo do Centro de Procriação Medicamente Assistida do Centro Hospitalar do Porto - Unidade Maternidade de Júlio Dinis, em particular à Dra. Carla Leal que esteve sempre disponível para me ajudar a compreender conceitos, para me dar uma palavra de incentivo e cujos contributos foram fundamentais para consolidar aspetos teóricos relacionados com Procriação Medicamente Assistida.

À minha família, meu porto de abrigo, à minha mãe, aos meus irmãos e cunhados por compreenderem as sucessivas e frequentes ausências, pela força e por acreditarem em mim. Aos meus sobrinhos porque me conseguem fazer sorrir mesmo quando não estão fisicamente presentes.

Às amigas Cláudia A., Patrícia, Teresa não tenho palavras para descrever a vossa amizade e todo o apoio que me deram desde o início e em particular na fase final. Cláudia M., Cristiana, Daniela, Dulce obrigada por tudo. Tenho realmente muita sorte com as pessoas que fazem parte da minha vida.

Ao Joel o meu sincero agradecimento pelo tempo, que lhe sei tão escasso, pelo apoio em Latex, pela amizade e pelos conselhos que, mesmo à distância, foram preciosos.

Aos colegas de MADSDAD, em especial à Maria e à Vera pelas trocas de ideias e incentivos mútuos. À Ana e à Geraldine pela força e motivação. Ao Luís pela boa disposição.

À Isabel, amiga e colega de trabalho, pela partilha da sua experiência e pela paciência com algum mau humor provocado pelas poucas horas de sono.

Ao Miguel pela amizade, pela força, pelas pausas para lanchar e pela boa música.

A análise multivariada é uma ferramenta estatística que facilita a compreensão de comportamentos complexos através da simplificação da estrutura dos dados e da síntese da informação, quando o número de variáveis envolvidas é grande.

De entre as várias técnicas de análise multivariada existentes, nesta dissertação será explorada a teoria associada à metodologia STATIS Dual com posterior aplicação prática a um conjunto de dados que regista informação sobre ciclos de Procriação Medicamente Assistida, PMA, de casais que recorreram ao Centro de Procriação Medicamente Assistida, CPMA, do Centro Hospitalar do Porto, CHP, entre Setembro de 2010 e Outubro de 2011. Trata-se de casais com diagnóstico de infertilidade, doença que afeta cerca de 15% da população mundial.

O principal objetivo deste trabalho é perceber quais as variáveis que mais contribuem para as diferenças entre quatro grupos de casais: casais em ciclo de Fertilização in Vitro, FIV, que engravidaram, casais em ciclo FIV que não engravidaram, casais em ciclo de Microinjeção Intracitoplasmática de Espermatozóides, ICSI, que engravidaram e casais em ciclo ICSI que não engravidaram.

A análise efetuada revelou uma maior proximidade entre os grupos compostos por casais que não engravidaram e um maior distanciamento entre os grupos de casais que engravidaram. Verificou-se a existência de uma estrutura comum aos quadros de dados que permitiu a construção do objeto compromisso. A percentagem de inércia explicada por cada eixo não se revelou elevada, tendo sido analisados cinco eixos para obter apenas cerca de 53% da variabilidade total. A maioria das variáveis apresenta uma fraca qualidade de representação nos eixos sendo a variável que regista a dose de gonadotrofinas administrada ("Gona") a que mais contribui para os afastamentos entre os quadros.

Palavras chave: Análise Multivariada, Infertilidade, PMA, STATIS Dual.

ABSTRACT

Multivariate analysis is a statistical tool that improves the understanding of complex behaviors by simplifying the structure of data and synthesizing of information, when the number of variables involved is large.

Among the various multivariate analysis techniques, this dissertation will explore the theory associated with STATIS Dual methodology. Afterwards, the usability of this method will be demonstrated in a data set that records information about Assisted Medical Reproduction cycles of couples who resorted to the Assisted Medical Reproduction Center in Oporto Hospital, EPE between September 2010 and October 2011. These couples have infertility diagnosis, disease that affects around 15% of the world population.

The main objective of this work is to understand which variables contribute most to the differences between the four groups of couples: couples undergoing In Vitro Fertilization, IVF, who became pregnant, couples undergoing IVF who did not became pregnant, couples undergoing Intra Cytoplasmic Sperm Injection, ICSI, who became pregnant and couples undergoing ICSI who did not became pregnant.

The analysis performed revealed greater proximity between groups composed of couples who are not pregnant and a greater distance between the groups of couples who become pregnant. It was verified the existence of a common structure which allowed the representation of the object compromise. The percentage of inertia explained by each axis is weak, and were analyzed five axes representing about 53 % of the total variability. Most variables has a poor quality of representation in the axes, being the variable that records the dose of gonadotropins administered ("Gona") that contributes most to the distances between the frames.

Keywords: Assisted Medical Reproduction, Infertility, Multivariate Analysis, STATIS Dual .

Índice

Nota Biográfica	v
Agradecimentos	vii
Resumo	ix
Abstract	xi
Lista de Figuras	xv
Lista de Tabelas	xvii
Abreviaturas	xix
1 INTRODUÇÃO	1
1.1 A Análise de Dados e os Sistemas de Apoio à Decisão	1
1.2 Motivação	2
1.3 Objetivo	3
1.4 Estrutura do documento	4
2 ANÁLISE EM COMPONENTES PRINCIPAIS	7
2.1 Introdução	7
2.2 A matriz de dados	8
2.2.1 Dados Centrados e Reduzidos	9
2.3 Espaço dos Indivíduos	11
2.4 Espaço das Variáveis	12
2.5 Determinação das componentes principais	13
2.5.1 Número de eixos a considerar	15
2.6 Interpretação e qualidade da representação de resultados	17
2.6.1 Correlações entre as componentes principais e as variáveis iniciais	17
2.6.2 Interpretação da posição dos indivíduos	19

3	AS METODOLOGIAS STATIS E STATIS DUAL	21
3.1	Introdução	21
3.2	A Metodologia STATIS Dual	23
3.2.1	Notações	23
3.2.2	Estudo da interestrutura	24
3.2.3	Estudo da intraestrutura	28
3.2.4	Interpretação das trajetórias	31
4	A PROCRIAÇÃO MEDICAMENTE ASSISTIDA	33
4.1	Introdução	33
4.2	Procriação Medicamente Assistida (PMA)	34
4.3	A PMA em Portugal	36
5	ANÁLISE PRELIMINAR DOS DADOS	39
5.1	Apresentação dos dados	39
5.2	Análise Univariada e Bivariada	40
6	ANÁLISE STATIS DUAL	51
6.1	Introdução	51
6.2	Caraterização dos quadros de dados	52
6.2.1	Análise STATIS Dual	55
7	CONCLUSÕES	71
7.1	Conclusão final	71
7.2	Trabalho Futuro	73
	Anexos	75
	Anexos 1	77
	Anexos 2	79
	Anexos 3	83
	Anexos 4	85
	BIBLIOGRAFIA	87

Lista de Figuras

Figura 2.1	Scree plot.	16
Figura 2.2	Círculo de correlações.	18
Figura 3.1	Esquema STATIS Dual.	23
Figura 3.2	Representação de uma estrutura comum.	30
Figura 3.3	Objeto mal representado na estrutura.	30
Figura 3.4	Objetos com normas bastante diferentes.	30
Figura 3.5	Inexistência de uma estrutura.	30
Figura 4.1	Distribuição dos centros de PMA autorizados pelas regiões do país.	38
Figura 5.1	Distribuição da taxa de sucesso por técnica de fertilidade.	45
Figura 5.2	Distribuição das causas de infertilidade por técnica de fertilidade.	49
Figura 6.1	Gráficos Estrela para as variáveis em estudo.	54
Figura 6.2	Estrutura da análise.	56
Figura 6.3	Interestrutura não centrada.	58
Figura 6.4	Interestrutura centrada.	59
Figura 6.5	Gráfico de Cattell.	60
Figura 6.6	Imagem euclidiana do compromisso no 1º e 2º eixos.	62
Figura 6.7	Imagem euclidiana do compromisso no 1º e 3º eixos.	63
Figura 6.8	Imagem euclidiana do compromisso no 1º e 4º eixos.	63
Figura 6.9	Imagem euclidiana do compromisso no 1º e 5º eixos.	64
Figura 6.10	Trajetórias das variáveis "Gona", "Endométrio", "DiasEstim", "EmbTBQ", "LHBasal", "FSHBasal".	66

LISTA DE FIGURAS

Figura 6.11 Trajetória da variável "Estradiol".	67
Figura 6.12 Trajetórias das variáveis "OvFec", "OvInsem" e "E2Basal".	68
Figura 6.13 Trajetórias das variáveis "TxFec", "IM", "%EspermaB", "IdadeF" e "EmbTMQ".	68

Lista de Tabelas

Tabela 5.1	Variáveis disponíveis em cada uma das bases de dados.	40
Tabela 5.2	Estatísticas descritivas da variável "IdadeF".	41
Tabela 5.3	Categorias da variável "IdadeF".	42
Tabela 5.4	Teste Qui-Quadrado: variáveis "Grav" vs "IdadeF" para casais em FIV.	42
Tabela 5.5	Resíduos ajustados: variáveis "Grav" vs "IdadeF" para casais em FIV.	43
Tabela 5.6	Medidas de associação baseadas no Qui-Quadrado.	43
Tabela 5.7	Estatísticas descritivas de algumas variáveis quantitativas.	44
Tabela 5.8	Teste t para a diferença das médias entre os grupos FIV e ICSI relativamente à variável "%EspermaB".	44
Tabela 5.9	Teste t para a diferença das médias entre os grupos FIV e ICSI relativamente à variável "OvInsem".	44
Tabela 5.10	Qui-Quadrado: variáveis "Técnica" vs "Grav".	46
Tabela 5.11	Resíduos ajustados: variáveis "Técnica" vs "Grav".	46
Tabela 5.12	Qui-Quadrado: variáveis "DiasEstim" vs "Grav".	47
Tabela 5.13	Categorias da variável "Causa".	48
Tabela 5.14	Categorias da variável "IMC".	49
Tabela 6.1	Nº de casais por técnica e por resultado do ciclo.	52
Tabela 6.2	Normas dos objetos.	56
Tabela 6.3	Coeficientes RV	57
Tabela 6.4	Distâncias euclidianas.	57

LISTA DE TABELAS

Tabela 6.5	Coeficientes de ponderação β'_k s.	58
Tabela 6.6	Distância dos objetos ao compromisso.	59
Tabela 6.7	Inércia explicada por cada eixo.	60
Tabela 6.8	Coordenadas das variáveis no compromisso.	61
Tabela 6.9	Decomposição da soma das distâncias ao quadrado entre pares de objetos.	65
Tabela 6.10	Decomposição das distâncias entre pares de objetos em contribuições de variáveis para os afastamentos.	65
Tabela I	Descrição das variáveis.	77
Tabela II	Tabela II: Matrizes de Correlação dos quadros de dados FIVGrav e FIVNaoGrav.	80
Tabela III	Matrizes de Correlação dos quadros de dados ICSIGrav e ICSINaoGrav.	81
Tabela IV	Matriz Compromisso.	84
Tabela V	Coordenadas das trajetórias das variáveis no 1º eixo.	85
Tabela VI	Coordenadas das trajetórias das variáveis no 2º eixo.	86
Tabela VII	Coordenadas do compromisso no 1º e 2º eixos.	86

ABREVIATURAS

ACP	Análise em Componentes Principais
APF	Associação Portuguesa de Fertilidade
CHP	Centro Hospitalar do Porto
CNPMA	Centro Nacional de Procriação Medicamente Assistida
CPMA	Centro de Procriação Medicamente Assistida
DCBD	Descoberta de Conhecimento em Bases de Dados
FIV	Fertilização in Vitro
HS	Hilbert-Schmidt
IA	Inseminação Artificial
ICSI	Microinjeção Intracitoplasmática de Espermatozóides
INE	Instituto Nacional de Estatística
KDD	Knowledge Discovery in Database
MJD	Maternidade de Júlio Dinis
OMS	Organização Mundial de Saúde
PMA	Procriação Medicamente Assistida
RV	Coefficiente de correlação vectorial entre objetos
STATIS	Structuration de Tableaux à Trois Indices de la Statistique
WHO	World Health Organization

1.1 A Análise de Dados e os Sistemas de Apoio à Decisão

A tomada de decisão é um processo rotineiro experienciado por todo o Ser Humano ainda que, muitas vezes, de uma forma inconsciente. Questões comuns como: "O que vou vestir hoje? O que vou almoçar? Compro o vermelho ou o azul?" são exemplos básicos de como o ato de decidir está implantado no nosso quotidiano mesmo sem que tenhamos a percepção dele.

Facilmente concordamos que para decisões desta natureza não serão necessários recursos tecnológicos nem algoritmos elaborados que auxiliem no processo. No entanto o grau de complexidade pode ser tão variável e abrangente quanto imaginemos e quanto maior o conhecimento extraído do problema maior o sucesso na sua compreensão e, consequentemente, na apresentação de propostas.

O desenvolvimento tecnológico, nomeadamente ao nível dos sistemas informáticos, proporcionou um crescimento exponencial da informação disponível nas organizações. Os baixos custos associados e a simplicidade do processo são motivos razoáveis para que, mesmo sem que exista um objetivo definido, se armazene a informação na perspetiva de que esta possa vir a ser útil no futuro.

A grande questão deste fenómeno reside no facto da maioria desta informação se encontrar na sua forma bruta - dados - sendo, por isso, necessário ter a capacidade e a sabedoria de transformar os dados em informação e a informação em conhecimento. E é neste contexto que a Análise de Dados e os Sistemas de Apoio à Decisão ganham a sua força.

Knowledge Discovery in Database (KDD), ou seja, a Descoberta de Conhecimento em Bases de Dados (DCBD) é definida como o processo de descoberta de padrões e

tendências através da análise de grandes conjuntos de dados.

Existem diversos métodos que permitem a extração desse conhecimento e que deverão ser selecionados consoante o objetivo que se pretende (previsão, descrição, ...). Segundo Fayyad *et al.* (1996), os mais importantes para a DCBD são: Classificação, Regressão, Análise de Agrupamentos, Sumarização, Modelação de Dependências, Regras de Associação e Séries Temporais. A nova informação, que deverá ser considerada útil e interpretável, será utilizada nos diversos domínios da atividade humana, como por exemplo, na definição de estratégias de marketing direcionadas e assentes no perfil do consumidor, regras para atribuição de crédito a clientes que permitam a diminuição de crédito mal parado, associação entre grupos de clientes e o cabaz de compras; identificar terapias de sucesso no tratamento de determinada doença; adaptação de um portal ao seu utilizador, prever falhas em linhas de produção, entre outras.

1.2 Motivação

A análise multivariada é uma ferramenta estatística que facilita a compreensão de comportamentos complexos através da simplificação da estrutura dos dados e da síntese da informação quando o número de variáveis envolvidas é muito grande.

As análises univariada e bivariada, embora devolvam informações importantes, tornam-se insuficientes para uma análise completa dos dados, principalmente quando o objetivo é estudar com algum detalhe fenómenos que normalmente dependem de um conjunto de fatores, como é o caso da infertilidade.

Grande parte da atividade médica consiste em processar informação: dados sobre pacientes, planeamento de procedimentos e diagnósticos, estratégias de tratamento, interpretação de resultados de laboratório, entre muitos outros. A principal preocupação é o bem-estar do paciente e a conseqüente necessidade de responder positivamente às suas expectativas. E, naturalmente, quanto mais informação o médico possuir sobre o problema em questão melhor resposta dará ao cumprimento deste seu objetivo.

O problema de infertilidade tem vindo a assumir especial enfoque principalmente numa altura em que o contexto económico e social que vivemos se traduz, cada vez mais, numa planificação tardia da primeira gravidez por parte da mulher. Sabe-se que geralmente é a partir dos 35 anos de idade que a fertilidade feminina diminui consideravelmente.

Conforme publicação do INE, "Estatísticas no Feminino" (INE, 2012), enquanto que em 2003 a proporção de parturientes dos grupos etários 25-29 anos e 30-34 anos tinha peso idêntico (cerca de 30%), em 2010 o peso do primeiro destes grupos etários desceu para 26,5% e o do segundo subiu para 34,6%.

Anualmente centenas de casais procuram ajuda médica para verem o sonho de serem pais concretizado. Muitos experimentam longos períodos de ansiedade, esperança e frustração, são submetidos a procedimentos invasivos e vivem alturas de verdadeira pressão psicológica e social, sem nunca chegarem a atingir a gravidez desejada.

Perceber de que forma fatores hormonais e biológicos relacionados com o casal contribuem para o resultado de um ciclo é o objetivo de técnicos-embriologistas e médicos que convivem diariamente com o drama de casais com diagnóstico de infertilidade.

Todas as informações que permitam acrescentar algo de novo e de útil ao complexo problema da infertilidade, padrões e/ou tendências, contribuirão para ajustar o tratamento às características do casal na tentativa de promover maiores taxas de sucesso associadas às técnicas de fertilidade adotadas.

1.3 Objetivo

Neste trabalho será explorada a metodologia STATIS, na sua versão Dual, que tem como objetivo analisar, em simultâneo, vários quadros de dados contendo cada um destes informação sobre p variáveis em n indivíduos, eventualmente diferentes, para T ocasiões ou circunstâncias. Depois de analisada a teoria subjacente a este método de análise de dados, aplicar-se-á o mesmo a uma base de dados real que contém informação sobre os ciclos de Procriação Medicamente Assistida de casais acompanhados no Centro de Procriação Medicamente Assistida do Centro Hospitalar do Porto, EPE.

Estes dados serão divididos em grupos tendo sempre por base a variável "Resultado" que assume uma de duas categorias: "engravidou" ou "não engravidou". Pretende-se distinguir quais os quadros de dados que estão mais próximos uns dos outros e quais os mais afastados, avaliar a possibilidade de resumir a informação contida nos quadros iniciais num único quadro e, finalmente, perceber quais as variáveis que mais contribuem quer para os afastamentos entre pares de quadros de dados, quer para os afastamentos globais.

Espera-se que a análise das variáveis que mais contribuem para as semelhanças e dissemelhanças entre grupos de casais que resultaram em gravidez, e entre estes e os que resultaram em não gravidez, permita compreender melhor o problema da infertilidade, ou pelo menos, acrescentar alguma informação útil ao mesmo.

Antes de proceder à análise STATIS Dual será efetuada uma análise preliminar aos dados que permita descrever com algum detalhe a informação presente nos mesmos. O objetivo desta análise é promover um maior entendimento e conhecimento sobre os dados e relações entre as variáveis, bem como ganhar alguma sensibilidade para o problema da infertilidade.

1.4 Estrutura do documento

O conteúdo da presente dissertação encontra-se dividido em 7 capítulos.

O **capítulo 1**, a Introdução, inicia com uma pequena reflexão sobre a importância da Análise de Dados e dos Sistemas de Apoio à Decisão no contexto real. Apresenta-se de seguida a motivação para a realização deste trabalho, os objetivos propostos e os conteúdos que compõem cada um dos capítulos posteriores.

No **capítulo 2** é apresentada uma abordagem teórica ao método estatístico multivariado que serve de base à metodologia STATIS Dual: Análise em Componentes Principais (ACP). Para analisar um conjunto de p variáveis pode ser útil transformar o conjunto original de variáveis num novo conjunto de variáveis não correlacionadas denominadas por componentes principais. O objetivo desta técnica é analisar se as primeiras componentes principais retêm a maior parte da variação presente nos dados originais permitindo, desta forma, um melhor entendimento sobre os dados e facilitando análises posteriores, uma vez que se diminui o número de variáveis em análise.

Para este capítulo foram utilizadas como referências bibliográficas: Chatfield *et al.* (1980), Dazy *et al.* (1996) e Jolliffe (1986).

No **capítulo 3** é descrita a teoria subjacente à metodologia STATIS Dual, técnica que permite a análise simultânea de vários quadros de dados quantitativos. Cada um dos quadros pode representar um instante temporal ou uma circunstância, referindo-se às mesmas variáveis.

Tem como objetivo destacar quais as variáveis responsáveis pelas semelhanças e dissimelhanças entre os vários quadros de dados.

O suporte bibliográfico utilizado neste capítulo conta com referências como Lavit (1988a), Lavit *et al.* (1994) e Dazy *et al.* (1996).

No **capítulo 4** apresentam-se alguns conceitos associados ao tema PMA e Infertilidade. Inclui uma contextualização do problema no panorama nacional, bem como alguns dados estatísticos sobre a infertilidade. Descrevem-se sucintamente três técnicas de PMA, Fertilização in Vitro (FIV), Microinjeção Intracitoplasmática de Espermatozoides (ICSI) e Inseminação Artificial (IA), sublinhando a importância que estas assumem na vida de casais que se deparam com o problema da infertilidade, definido como doença pela Organização Mundial de Saúde (OMS).

O **capítulo 5** destina-se à apresentação dos resultados das análises univariada e bi-

variada realizadas aos dados disponibilizados pelo CPMA do CHP. São utilizados testes estatísticos do Qui-Quadrado e o teste t para as médias de duas amostras independentes para confirmar a coerência dos dados em análise com os resultados presentes na literatura. O objetivo deste capítulo é conhecer os dados e o tipo de relações existentes entre algumas variáveis.

No **capítulo 6** procede-se à análise das bases de dados à luz da metodologia STATIS Dual. Pretende-se perceber quais as variáveis responsáveis pelas diferenças entre quadros de dados compostos por casais que obtiveram a gravidez no final do ciclo e quadros de dados compostos por casais que não conseguiram atingir o objetivo proposto. Será apresentado um estudo que considera quatro grupos definidos conforme o resultado do ciclo ("gravidez" e "não gravidez") e conforme a técnica de fertilidade a que foram submetidos (FIV e ICSI).

O **capítulo 7** encerra o presente trabalho apresentando as conclusões decorrentes da aplicação da metodologia e uma perspectiva para trabalhos futuros.

2.1 Introdução

As primeiras publicações na área da Análise em Componentes Principais (ACP) devem-se a Pearson (1901) tendo posteriormente sido desenvolvidas por Hotelling (1933). Trata-se de uma técnica de estatística multivariada, essencialmente descritiva, aplicável a variáveis quantitativas (ou que possam ser consideradas como tal) que põe em evidência as relações existentes entre as variáveis (associações e oposições).

O procedimento de ACP propõe transformar um conjunto de dados constituído por p variáveis num novo conjunto de dados composto por um número substancialmente menor de variáveis, q , $q < p$, não correlacionadas que representem a maior parte da informação presente no conjunto de dados original:

$$X_{n \times p} \equiv X = \begin{pmatrix} x_1^1 & \cdots & x_1^p \\ \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^p \end{pmatrix} \xrightarrow{\text{ACP}} Y_{n \times q} = \begin{pmatrix} y_1^1 & \cdots & y_1^q \\ \vdots & \ddots & \vdots \\ y_n^1 & \cdots & y_n^q \end{pmatrix}, \text{ com } q < p.$$

Estas novas variáveis, denominadas de componentes principais, resultam de combinações lineares de todas as variáveis do conjunto inicial sendo, por isso, não correlacionadas. Assim, a ideia subjacente à metodologia assenta no pressuposto de que, se no conjunto de dados original existem variáveis fortemente correlacionadas, então a informação que trazem ao problema é a mesma.

No entanto, ao benefício inerente à diminuição da dimensão do conjunto de dados está associada uma consequência: a perda de alguma da informação inicialmente presente nos dados.

O principal objetivo da ACP é que as primeiras componentes retenham o máximo da informação original, facilitem a compreensão dos dados e sejam úteis para análises subsequentes.

2.2 A matriz de dados

Na ACP dispomos de uma matriz X com n linhas - *indivíduos* e p colunas - *variáveis*:

$$X = \begin{pmatrix} x_1^1 & \cdots & x_1^p \\ \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^p \end{pmatrix},$$

onde o elemento x_i^j representa o valor da variável j para o indivíduo i ($j = 1, \dots, p$ e $i = 1, \dots, n$).

A variável j é representada pelo vector \mathbf{x}^j :

$$\mathbf{x}^j = \begin{pmatrix} x_1^j \\ \vdots \\ x_n^j \end{pmatrix},$$

e o indivíduo i é representado pelo vector $\mathbf{x}_i = (x_i^1, \dots, x_i^p)$.

Para calcular a distância entre duas variáveis é, por vezes, necessário atribuir pesos, p_i , aos indivíduos consoante a importância que lhe queiramos dar no conjunto dos dados.

Seja D a matriz dos pesos:

$$D = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}, \text{ com } 0 < p_i < 1 \text{ e } \sum_{i=1}^n p_i = 1. \quad (2.1)$$

Usualmente atribui-se o mesmo peso a todos os indivíduos e neste caso temos $D = \frac{1}{n}I_n$, sendo I_n a matriz identidade de ordem n .

O valor médio que cada variável toma no conjunto dos indivíduos dá origem ao conceito de *centro de gravidade*.

Definição 2.2.1 (Centro de Gravidade) Designa-se por *Centro de Gravidade* associado à matriz de pesos D , o vector \mathbf{g} definido por:

$$\mathbf{g} = \begin{pmatrix} \bar{x}^1 \\ \vdots \\ \bar{x}^p \end{pmatrix}, \text{ onde } \bar{x}^j = \sum_{i=1}^n p_i x_i^j.$$

O valor \bar{x}^j corresponde à média ponderada dos valores da variável j no conjunto dos n indivíduos. Matricialmente temos,

$$\mathbf{g} = X'D\mathbf{1}_n, \text{ com } \mathbf{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n. \quad (2.2)$$

2.2.1 Dados Centrados e Reduzidos

Numa base de dados é frequente encontrar variáveis medidas em escalas diferentes. Para evitar análises menos corretas influenciadas pelas ordens de grandeza das variáveis, procede-se a uma transformação dos dados que permita eliminar o efeito dessas mesmas diferenças.

Os dados podem ser centrados, reduzidos ou, simultaneamente, centrados e reduzidos. No primeiro caso cada variável apresentará média zero, no segundo caso cada variável apresentará variância 1 e, no último caso, teremos variáveis com média zero e variância igual a 1.

Consideremos as matrizes $Q_{\frac{1}{s}}$ e $Q_{\frac{1}{s^2}}$ tais que:

$$Q_{\frac{1}{s}} = \begin{pmatrix} \frac{1}{s_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_p} \end{pmatrix} \quad \text{e} \quad Q_{\frac{1}{s^2}} = \begin{pmatrix} \frac{1}{s_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{s_p^2} \end{pmatrix}, \quad (2.3)$$

onde $s_j^2 = \sum_{i=1}^n p_i (x_i^j - \bar{x}^j)^2$ representa a variância da variável j .

A covariância entre as variáveis j e j' é dada por:

$$s_{jj'} = \sum_{i=1}^n p_i (x_i^j - \bar{x}^j) (x_i^{j'} - \bar{x}^{j'}), \quad (2.4)$$

e o coeficiente de correlação linear entre as variáveis j e j' por:

$$r_{jj'} = \left(\frac{s_{jj'}}{s_j s_{j'}} \right), \quad (2.5)$$

sendo s_j o desvio padrão da variável j .

Consideremos a matriz centrada Y associada a X . O termo geral de Y é da forma $y_i^j = x_i^j - \bar{x}^j$, ou seja, cada entrada y_i^j é obtida subtraindo ao valor do indivíduo i na variável j a média ponderada da variável j .

Temos:

$$\begin{aligned} Y &= X - \mathbf{1}_n \mathbf{g}' \\ &\stackrel{(2.2)}{=} X - \mathbf{1}_n (\mathbf{X}' \mathbf{D} \mathbf{1}_n)' \\ &= X - \mathbf{1}_n \mathbf{1}_n' \mathbf{D}' \mathbf{X} \\ &= (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n' \mathbf{D}) \mathbf{X} \end{aligned} \quad (2.6)$$

Consideremos a matriz Z centrada e reduzida associada a X . Temos:

$$Z = Y \mathbf{D}_s^{-1} \text{ cujo termo geral é dado por } z_i^j = \frac{x_i^j - \bar{x}^j}{s_j}, \quad (2.7)$$

ou seja, cada entrada da matriz centrada Y é dividida pelo desvio padrão s_j da variável j .

Como já foi referido, esta transformação efetua-se quando as unidades de medida das variáveis são diferentes ou quando tomam valores de ordem de grandeza muito diferentes.

Matriz de Variâncias e Covariâncias e Matriz de Correlações

Definimos a matriz V de variâncias e covariâncias por:

$$V = \mathbf{X}' \mathbf{Q} \mathbf{X} - \mathbf{g} \mathbf{g}' = \mathbf{Y}' \mathbf{Q} \mathbf{Y} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix} \quad (2.8)$$

e a matriz R de correlações por:

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}. \quad (2.9)$$

Temos que,

$$R = Q_{\perp} V Q_{\perp}' \stackrel{(2.8)}{=} Q_{\perp} Y' Q Y Q_{\perp}' \stackrel{(2.7)}{=} Z' Q Z, \quad (2.10)$$

isto é, R coincide com a matriz de variâncias e covariâncias da tabela centrada e reduzida Z definida em (2.7). Esta matriz resume a estrutura das dependências lineares entre as p variáveis.

2.3 Espaço dos Indivíduos

Um indivíduo \mathbf{x}_i é definido pelas p coordenadas que correspondem aos valores das p variáveis que lhe estão associados. Assim, \mathbf{x}_i é um vetor e pode, portanto, ser considerado como um elemento do espaço vectorial $\mathbf{F} \in \mathbb{R}^p$, denominado *espaço dos indivíduos*.

O conjunto dos n indivíduos é uma nuvem de \mathbf{F} onde \mathbf{g} representa o centro de gravidade que será coincidente com a origem dos eixos se os dados estiverem centrados.

Este espaço \mathbf{F} deverá estar munido de uma métrica que permita definir a distância entre dois indivíduos:

Definição 2.3.1 (Métrica sobre o espaço dos indivíduos) *A distância entre dois indivíduos \mathbf{x}_i e \mathbf{x}_j é definida pela métrica Q (matriz simétrica definida positiva de dimensão p) caracterizada pela relação:*

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)' Q (\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_Q^2.$$

Definição 2.3.2 (Produto escalar entre indivíduos) *A matriz dos produtos escalares entre indivíduos é a matriz $W = X Q X'$ de termo geral:*

$$w_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle_Q = \mathbf{x}_i Q \mathbf{x}_j. \quad (2.11)$$

As métricas mais utilizadas em ACP são:

- $Q = I_p$, que corresponde a utilizar o produto escalar euclidiano;
- $Q = Q_{\perp} \frac{1}{s^2}$, definida em (2.3). Esta métrica permite que cada variável tenha uma contribuição análoga na determinação das distâncias.

Resumindo, para dados centrados e reduzidos, a análise da nuvem de pontos em \mathbb{R}^p conduz à translação da origem para o centro de gravidade e à transformação das escalas dos diferentes eixos.

Definição 2.3.3 (Inércia Total) *Define-se inércia total da nuvem de pontos como a média ponderada dos quadrados das distâncias dos pontos ao centro de gravidade da nuvem:*

$$I_g = \sum_{i=1}^n p_i \|\mathbf{x}_i - \mathbf{g}\|_Q^2 = \sum_{i=1}^n p_i (\mathbf{x}_i - \mathbf{g})' Q (\mathbf{x}_i - \mathbf{g}). \quad (2.12)$$

Para dados centrados temos que $\mathbf{g} = \mathbf{0}$.

Teorema 2.3.1 (Teorema de Huyghens) *Se representarmos por*

$$I_a = \sum_{i=1}^n p_i \|\mathbf{x}_i - \mathbf{a}\|_Q^2 = \sum_{i=1}^n p_i (\mathbf{x}_i - \mathbf{a})' Q (\mathbf{x}_i - \mathbf{a})$$

a inércia da nuvem de indivíduos relativamente a um ponto \mathbf{a} , temos que:

$$I_a = I_g + \|\mathbf{g} - \mathbf{a}\|_Q^2. \quad (2.13)$$

Note-se que:

- $I_g = \text{Tr}(QV) = \text{Tr}(VQ)$, onde $\text{Tr}(A)$ representa o traço da matriz A ;
- $I_g = \text{Tr}(WD) = \text{Tr}(DW)$;
- Se $Q = I$ então $I_g = \text{Tr}(QV) = \text{Tr}(V)$, isto é, I_g é a soma das variâncias das p variáveis.
- Se $Q = Q_{\frac{1}{s^2}}$, então $I_g = \text{Tr}(R) = p$.

2.4 Espaço das Variáveis

Cada variável \mathbf{x}_j é considerada um vector de um espaço vectorial $\mathbf{E} \in \mathbb{R}^n$ designado por *espaço das variáveis*. A métrica utilizada neste espaço será a métrica D definida em (2.1) de dimensão $n \times n$, simétrica e definida positiva que, para variáveis centradas, apresenta as seguintes propriedades:

- O produto escalar de duas variáveis centradas é igual à sua covariância, isto é,

$$\langle \mathbf{x}_i, \mathbf{x}_k \rangle_D = (\mathbf{x}_j)' D \mathbf{x}_k = \sum_{i=1}^n p_i x_i^j x_i^k = s_{jk}; \quad (2.14)$$

- O comprimento de uma variável centrada é igual ao seu desvio-padrão, isto é,

$$\|\mathbf{x}_j\|_D = \sqrt{\langle \mathbf{x}_j, \mathbf{x}_j \rangle_D} \stackrel{(2.14)}{=} \sqrt{s_{jj}} = \sqrt{s_j^2} = s_j; \quad (2.15)$$

- O cosseno do ângulo entre duas variáveis centradas é igual ao coeficiente de correlação linear entre elas, isto é,

$$\cos(\mathbf{x}_j, \mathbf{x}_k) = \frac{\langle \mathbf{x}_j, \mathbf{x}_k \rangle_D}{\|\mathbf{x}_j\|_D \|\mathbf{x}_k\|_D} \stackrel{(2.14), (2.15)}{=} \frac{s_{jk}}{s_j s_k} \stackrel{(2.5)}{=} r_{jk}. \quad (2.16)$$

Posteriormente utilizar-se-á a designação **estudo** para referir o tripleto definido por (X, Q, D) , onde:

- X representa a matriz de dados original;
- Q é métrica associada ao espaço dos indivíduos, isto é, aquela que nos permite calcular a distância entre indivíduos;
- D é a métrica associada ao espaço das variáveis, ou seja, a que nos permite calcular as distâncias entre variáveis.

Um estudo pode ser caracterizado, ou representado, por dois objetos W e V diferentes, tais que:

$$W = XQX', \text{ definido em (2.11);}$$

$$V = X'DX, \text{ definido em (2.8) (se os dados estiverem centrados).}$$

2.5 Determinação das componentes principais

A estrutura de interdependência entre as variáveis do conjunto de dados é representada pela matriz V de variâncias e covariâncias ou pela matriz de correlação R .

O objetivo da ACP é transformar a estrutura original representada pelas variáveis $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^p$ numa outra estrutura representada pelas variáveis $\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^p$ não correlacionadas e com variâncias ordenadas para que seja possível comparar indivíduos usando apenas as variáveis $\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^q$, com $q < p$, que apresentem maior variância.

As componentes principais são apresentadas por ordem decrescente de importância, isto é, a primeira explica o máximo possível da variabilidade dos dados originais, a segunda o máximo possível de variabilidade ainda não explicada após o efeito da primeira componente principal, e assim por diante.

Sendo R uma matriz simétrica, temos que R é diagonalizável e portanto o cálculo das raízes da equação $\det[R - \lambda I_p] = |R - \lambda I_p| = 0$ permitirá obter p soluções reais linearmente independentes, $\lambda_1, \dots, \lambda_p$, que são os valores próprios da matriz R .

Convém sublinhar que o resultado encontrado para a análise a partir da matriz V pode ser diferente do resultado encontrado a partir da matriz R . A prática habitual é proceder à padronização dos dados sempre que as unidades de medida das variáveis difiram.

Sejam $\lambda_1, \dots, \lambda_p$ os valores próprios da matriz R tais que:

$$\lambda_1 > \lambda_2 > \dots > \lambda_q > \dots > \lambda_p.$$

Para cada valor próprio λ_k , existe um vetor próprio \mathbf{u}_k associado:

$$\mathbf{u}_k = \begin{pmatrix} u_{k1} \\ u_{k2} \\ \vdots \\ u_{kp} \end{pmatrix}.$$

Os vetores próprios são normados, isto é, $\mathbf{u}'_k \mathbf{u}_k = 1$ e são ortogonais entre si, isto é, $\mathbf{u}'_k \mathbf{u}_l = 0$ para $k \neq l$. A primeira restrição assegura a unicidade (a menos de sinal) dos coeficientes e a segunda assegura que as componentes não são correlacionadas entre si.

A primeira componente principal de X é a combinação linear \mathbf{y}^1 expressa por:

$$\mathbf{y}^1 = u_{11}\mathbf{x}^1 + u_{12}\mathbf{x}^2 + \dots + u_{1p}\mathbf{x}^p, \text{ onde } s^2(\mathbf{y}^1) = \lambda_1 \text{ e } \mathbf{u}'_1 \mathbf{u}_1 = 1.$$

A segunda componente principal de X é a combinação linear \mathbf{y}^2 expressa por:

$$\mathbf{y}^2 = u_{21}\mathbf{x}^1 + u_{22}\mathbf{x}^2 + \dots + u_{2p}\mathbf{x}^p, \text{ onde } s^2(\mathbf{y}^2) = \lambda_2, \mathbf{u}'_2 \mathbf{u}_2 = 1 \text{ e } \mathbf{u}'_1 \mathbf{u}_2 = 0.$$

Utilizando o mesmo procedimento para as restantes componentes obtemos a generalização para a i -ésima componente principal \mathbf{y}^i .

Sendo \mathbf{u}'_i o vetor próprio correspondente ao valor próprio λ_i , a componente principal i é dada por:

$$\mathbf{y}^i = u_{i1}\mathbf{x}^1 + u_{i2}\mathbf{x}^2 + \dots + u_{ip}\mathbf{x}^p, \\ \text{onde } s^2(\mathbf{y}^i) = \lambda_i, \mathbf{u}'_i \mathbf{u}_i = 1 \text{ e } \mathbf{u}'_k \mathbf{u}_i = 0 \text{ para } i \neq k.$$

As componentes principais apresentam as seguintes propriedades:

- A variância da componente principal \mathbf{y}^i é igual ao valor próprio λ_i :

$$s^2(\mathbf{y}^i) = \lambda_i;$$

- As componentes principais estão ordenadas por ordem decrescente do valor da variância:

$$s^2(\mathbf{y}^1) > s^2(\mathbf{y}^2) > \dots > s^2(\mathbf{y}^q) > \dots > s^2(\mathbf{y}^p);$$

- O total de variância das variáveis originais é igual ao somatório dos valores próprios que é igual ao total da variância das componentes principais:

$$\sum_{i=1}^p s^2(\mathbf{x}^i) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p s^2(\mathbf{y}^i);$$

- As componentes principais não são correlacionadas entre si:

$$s(\mathbf{y}^i, \mathbf{y}^j) = 0, \text{ para } i \neq j.$$

2.5.1 Número de eixos a considerar

Inércia explicada pelos eixos

A importância de uma componente principal é avaliada em termos da sua contribuição, isto é, pela proporção de variância total por ela explicada.

A contribuição da k -ésima componente na explicação da variação total, ou por outras palavras, a percentagem de inércia explicada pelo eixo principal \mathbf{u}_k pode ser expressa por:

$$\frac{s^2(\mathbf{y}^k)}{\sum_{i=1}^p s^2(\mathbf{y}^i)} = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}. \quad (2.17)$$

A percentagem de inércia explicada pelo plano $[\mathbf{u}_k, \mathbf{u}_l]$ é dada por:

$$\frac{\lambda_k + \lambda_l}{\sum_{i=1}^p \lambda_i}.$$

A percentagem de inércia explicada pelos q primeiros eixos pode ser expressa por:

$$\frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^p \lambda_i}. \quad (2.18)$$

A soma dos q primeiros valores próprios representa a quantidade de informação retida na redução de p para q variáveis. Através desta informação irá ser determinado o número de componentes principais que serão utilizadas na análise.

Critérios para determinar o número de eixos a reter

A questão que se coloca neste momento é: "Qual o número de componentes principais que devem ser consideradas?"

Existem vários critérios na literatura que auxiliam na tomada de decisão, sendo que habitualmente o critério utilizado é o proposto por Pearson. Este consiste em reter as q primeiras componentes que expliquem determinada quantidade fixada, que normalmente corresponde a 80%:

$$\frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 \simeq 80\%$$

Cattell (1966) propõe a representação gráfica da percentagem de inércia explicada por cada componente, disposta por ordem decrescente. O ponto no qual o gráfico começa a ficar horizontal define o número de eixos a considerar.

No exemplo ilustrativo representado na figura 2.1, reter-se-iam 3 eixos.

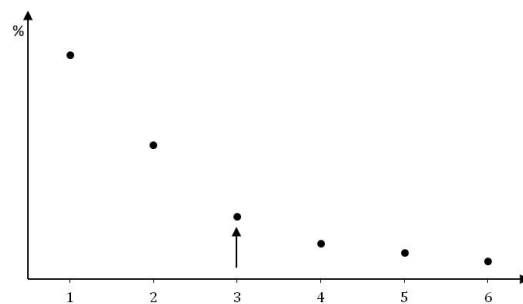


Figura 2.1: Scree plot.

Para dados centrados e reduzidos Kaiser (1958) propõe reter as componentes principais cujos valores próprios sejam superiores a 1.

2.6 Interpretação e qualidade da representação de resultados

2.6.1 Correlações entre as componentes principais e as variáveis iniciais

A interpretação dos eixos faz-se por intermédio do estudo das correlações entre a componente principal que define o eixo e as variáveis da tabela de dados inicial. Para dados centrados e reduzidos, a representação gráfica das variáveis pode ser feita utilizando o coeficiente de correlação linear entre \mathbf{y}^k e \mathbf{x}^j :

$$r_{x^j, y^k} = \sqrt{\lambda_k} u_{kj}. \quad (2.19)$$

Considerando os eixos associados a um par de componentes principais, $(\mathbf{y}^k, \mathbf{y}^l)$, as coordenadas da variável j serão representadas pelo ponto:

$$(r_{x^j, y^k}, r_{x^j, y^l}). \quad (2.20)$$

Sendo a norma das variáveis igual a 1, cada ponto representado pelas coordenadas descritas em (2.20) encontra-se posicionado no interior de um círculo de raio 1 centrado na origem. Neste caso, cada variável está posicionada à distância unitária da origem, isto é,

$$d^2(\mathbf{x}^j, O) = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i^j - \bar{x}^j}{s_j} \right)^2 = 1. \quad (2.21)$$

A distância entre as variáveis \mathbf{x}^j e $\mathbf{x}^{j'}$ é expressa por:

$$\begin{aligned} d^2(\mathbf{x}^j, \mathbf{x}^{j'}) &= \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i^j - \bar{x}^j}{s_j} - \frac{x_i^{j'} - \bar{x}^{j'}}{s_{j'}} \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i^j - \bar{x}^j}{s_j} \right)^2 + \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i^{j'} - \bar{x}^{j'}}{s_{j'}} \right)^2 - \\ &\quad - 2 \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i^j - \bar{x}^j}{s_j} \frac{x_i^{j'} - \bar{x}^{j'}}{s_{j'}} \right) \\ &\stackrel{(2.5), (2.21)}{=} 2(1 - r_{jj'}). \end{aligned} \quad (2.22)$$

Assim, as proximidades entre as variáveis podem ser interpretadas em termos das suas correlações. Para valores $r_{jj'} \simeq 1$ as variáveis apresentam uma correlação fortemente positiva encontrando-se próximas uma da outra.

Caso os dados se encontrem unicamente centrados o comprimento das variáveis é, como já foi referido, igual ao seu desvio padrão.

Na figura 2.2 estão projetadas 5 variáveis no plano $[\mathbf{y}^k, \mathbf{y}^l]$.

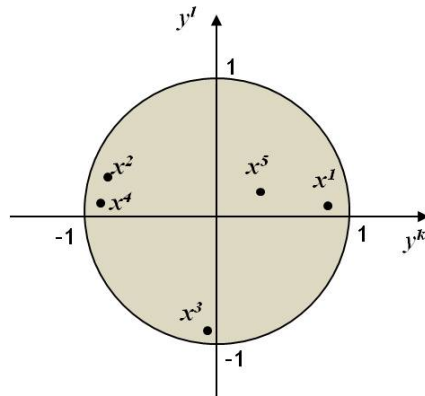


Figura 2.2: Círculo de correlações.

As variáveis \mathbf{x}^1 , \mathbf{x}^2 , \mathbf{x}^3 e \mathbf{x}^4 estão bem representadas no plano pois encontram-se próximas da circunferência unitária. As variáveis \mathbf{x}^2 e \mathbf{x}^4 estão fortemente correlacionadas entre si e apresentam uma correlação negativa forte com a variável \mathbf{x}^1 . Quanto à variável \mathbf{x}^5 , mal representada neste plano, nada se pode concluir.

Define-se:

- Contribuição relativa da componente \mathbf{y}^k para a variável \mathbf{x}^j :

$$CTR_j^k \stackrel{(2.16)}{=} r_{x^j, y^k}^2 \stackrel{(2.19)}{=} \left(\sqrt{\lambda_k} u_{kj} \right)^2;$$

- Contribuição relativa do plano $[\mathbf{y}^k, \mathbf{y}^l]$ para a variável \mathbf{x}^j :

$$CTR_j^{k,l} = r_{x^j, y^k}^2 + r_{x^j, y^l}^2 = CTR_j^k + CTR_j^l;$$

- Contribuição absoluta da variável \mathbf{x}^j para a formação da k -ésima componente principal:

$$CTA_k^j = \frac{u_{kj}^2}{s_j^2}.$$

2.6.2 Interpretação da posição dos indivíduos

Afirmar que a primeira componente principal y^1 apresenta uma correlação forte positiva com uma variável x^j , significa afirmar que indivíduos com uma coordenada positiva sobre o eixo u_k são caracterizados por um valor de x^j elevado.

A contribuição absoluta do indivíduo i para a formação do eixo principal u_k é:

$$CTA_i^k = \frac{p_i(y_{ik})^2}{\lambda_k}.$$

Note-se que:

- $\sum_{i=1}^n CTA_i^k = 1$;
- A contribuição absoluta de um indivíduo para a formação do eixo principal u_k , isto é, para a variância explicada pelo eixo u_k , permite evidenciar os indivíduos que apresentam características relacionadas com o fenómeno traduzido pela componente principal que lhe corresponde.

A qualidade da representação do indivíduo i no eixo u_k - contribuição relativa do eixo u_k para o indivíduo i - é avaliada a partir do ângulo θ que este forma com o eixo u_k :

$$CTR_i^k = \cos^2 \theta$$

Interessa também analisar a qualidade da representação dos indivíduos no plano principal $[u_k, u_l]$. Esta é avaliada a partir do ângulo, α que cada indivíduo forma com esse plano principal. Tem-se que:

$$CTR_i^{k,l} = CTR_i^k + CTR_i^l = \cos^2 \alpha.$$

CAPÍTULO 3

AS METODOLOGIAS STATIS E STATIS DUAL

3.1 Introdução

STATIS é um acrónimo utilizado para designar a expressão francesa '*Structuration de Tableaux à Trois Indices de la Statistique*' que pode, aproximadamente, ser traduzido como a estruturação estatística de tabelas de três dimensões.

A metodologia STATIS foi desenvolvida no laboratório de Probabilidades e Estatística da Universidade de Montpellier II por Yves Escoufier e sua equipa, com o primeiro trabalho publicado por L'Hermier des Plantes (1976) e posteriormente desenvolvido por Lavit (1988a, 1988b) e Lavit *et al.* (1994). Trata-se de um método exploratório de análise de dados, aplicável a variáveis de natureza quantitativa, que permite extrair informação de dados multidimensionais recolhidos em ocasiões diferentes e de uma das seguintes formas:

- i) T quadros de dados recolhidos em diferentes instantes de tempo ou circunstâncias sobre os mesmos indivíduos, podendo as variáveis ser ou não as mesmas em cada um dos quadros;
- ii) T quadros de dados compostos pelas mesmas variáveis medidas em conjuntos de indivíduos que podem diferir de quadro para quadro.

A cada uma destas situações está associada uma estratégia diferente. A primeira - STATIS - privilegia as posições relativas dos indivíduos. O procedimento consiste em comparar as nuvens de indivíduos através das matrizes dos produtos escalares. A segunda - STATIS Dual - privilegia as relações entre as variáveis. O procedimento consiste em comparar as nuvens de variáveis através das matrizes de correlações.

Independentemente da metodologia adotada, o objetivo passa por comparar os diversos quadros de dados que compõem o estudo, definir uma estrutura comum que resuma num só quadro a maior parte da informação analisada e, por fim, destacar os indivíduos (STATIS) ou as variáveis (STATIS Dual) que mais contribuíram para as semelhanças e dissemelhanças encontradas.

Podem então definir-se três etapas essenciais na metodologia:

1. Estudo da interestrutura cujo objetivo é proceder a uma comparação global da estrutura das T matrizes de dados.
2. Estudo da intraestrutura cujo objetivo é resumir num só objeto, designado por compromisso, a informação presente no conjunto de quadros analisados na interestrutura. O compromisso permitirá representar as posições compromisso dos indivíduos (ou variáveis) tal como são descritos pelo conjunto de quadros de dados - imagem euclidiana do compromisso.
3. Interpretação das trajetórias dos indivíduos (ou variáveis). A partir da imagem euclidiana do compromisso é traçada a trajetória de cada indivíduo (ou variável) que permitirá descrever o seu comportamento evolutivo. Esta representação irá destacar os indivíduos (ou variáveis) responsáveis pelas diferenças entre os quadros de dados.

A metodologia STATIS é popular para análise conjunta de quadros de dados e tem sido aplicada em estudos das mais variadas áreas de conhecimento. Niang *et al.* (2009) estudam o controlo multivariado de processos em lote; Scepi (2002) aplica a metodologia à área de controlo estatístico de qualidade; Fournier *et al.* (2009) investigam os transportes dentro do aquífero Karst; Gudmundsson *et al.* (2011) estudam padrões nos caudais baixos, médios e altos da Europa; Stanimirova *et al.* (2004) utiliza a metodologia em análise de dados ambientais; Thioulouse (2011) apresenta uma análise simultânea a pares de tabelas ecológicas; Figueiredo *et al.* (2012) estudam o processo de reestruturação de bancos portugueses após privatização; Oliveira *et al.* (2006) aplicam a metodologia a dados sobre HIV em Portugal. Os primeiros trabalhos realizados no âmbito desta metodologia destacam nomes como, L'Hermier des Plantes *et al.* (1977) com o estudo da precipitação na região do Mediterrâneo e Lavit (1988) que descreve a evolução das províncias espanholas entre 1960 e 1979 através da análise de dados económicos.

Na presença de quadros de dados compostos pelo mesmo número de variáveis e pelo mesmo número de indivíduos poder-se-á tirar partido da aplicação de ambas as abordagens STATIS.

3.2 A Metodologia STATIS Dual

3.2.1 Notações

A metodologia STATIS Dual aplica-se a T tabelas de dados compostas pelas mesmas p variáveis medidas em n_t conjuntos de indivíduos eventualmente diferentes, $t \in \{1, \dots, T\}$, conforme ilustra a figura 3.1.

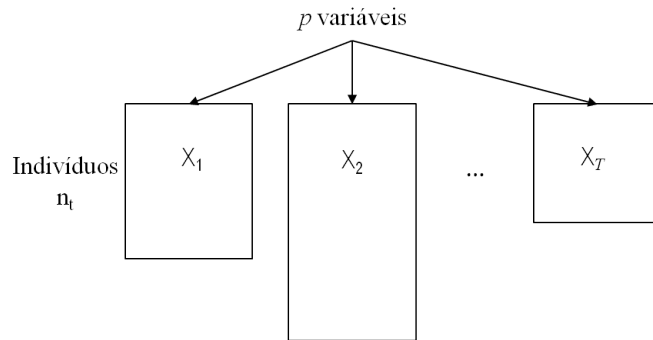


Figura 3.1: Esquema STATIS Dual.

Seja então $(X_t)_{n_t \times p}$ o quadro de dados associado ao t -ésimo momento no tempo ou circunstância, onde n_t representa o número de indivíduos no t -ésimo quadro e p representa o número de variáveis. O conjunto de dados é representado matricialmente por:

$$X_1 = \begin{pmatrix} (x_1^1)^1 & \cdots & (x_1^p)^1 \\ \vdots & \ddots & \vdots \\ (x_{n_1}^1)^1 & \cdots & (x_{n_1}^p)^1 \end{pmatrix}; X_2 = \begin{pmatrix} (x_1^1)^2 & \cdots & (x_1^p)^2 \\ \vdots & \ddots & \vdots \\ (x_{n_2}^1)^2 & \cdots & (x_{n_2}^p)^2 \end{pmatrix}; \cdots$$

Assim, no instante t , X_t é a matriz de dimensão $n_t \times p$ definida por:

$$X_t = \begin{pmatrix} (x_1^1)^t & (x_1^2)^t & \cdots & (x_1^p)^t \\ (x_2^1)^t & (x_2^2)^t & \cdots & (x_2^p)^t \\ \vdots & \vdots & \ddots & \vdots \\ (x_{n_t}^1)^t & (x_{n_t}^2)^t & \cdots & (x_{n_t}^p)^t \end{pmatrix}. \quad (3.1)$$

O i -ésimo indivíduo é o vector de \mathbb{R}^p

$$((x_i)^t)' = \left((x_i^1)^t, (x_i^2)^t, \dots, (x_i^p)^t \right)',$$

e a j -ésima variável é o vector de \mathbb{R}^{n_t}

$$(x^j)^t = \begin{pmatrix} (x_1^j)^t \\ (x_2^j)^t \\ \vdots \\ (x_{n_t}^j)^t \end{pmatrix}.$$

3.2.2 Estudo da interestrutura

O estudo da interestrutura corresponde a uma análise global dos T quadros de dados e pode ser decomposto numa sequência de três etapas:

- i) Definição de um objeto representativo para cada estudo;
- ii) Definição de uma distância (métrica) sobre os objetos representativos de cada estudo;
- iii) Definição de uma imagem euclidiana dos objetos representativos associada aos produtos escalares definidos em ii).

Definição dos objetos representativos

Conforme já foi introduzido na secção 2.4 do capítulo 2, o estudo t é um tripleto da forma $(X_t, Q, D_t)_{t=1, \dots, T}$ onde Q representa a métrica associada ao espaço dos indivíduos e D_t a métrica associada ao espaço das variáveis.

Na prática, $Q = I_p$ e $D_t = \text{diag}(1/n_1, \dots, 1/n_t, \dots, 1/n_T)$. O objeto representativo do estudo será a matriz V_t , de dimensão $p \times p$, definida por $V_t = X_t' D_t X_t$ - matriz de variâncias e de covariâncias do quadro X_t , onde X_t' representa a matriz transposta de X_t .

Definição do produto escalar

Com o objetivo de representar graficamente os T estudos, possibilitar a comparação dos T quadros de dados e averiguar sobre a existência de uma eventual estrutura comum entre as variáveis é necessário definir uma distância entre objetos V_t e $V_{t'}$. A metodologia STATIS Dual utiliza o produto escalar de *Hilbert-Schmidt*, HS , para induzir uma distância entre os objectos representativos V_t .

Definição 3.2.1 (Produto escalar de *Hilbert-Schmidt*) *O produto escalar de Hilbert-Schmidt entre dois objetos V_t e $V_{t'}$ é definido pela seguinte fórmula:*

$$\langle V_t, V_{t'} \rangle_{HS} = \text{Tr}(Q V_t Q V_{t'}) \quad (3.2)$$

onde $Tr(A)$ representa o traço da matriz A , isto é, a soma dos elementos da sua diagonal, e Q define a métrica no espaço dos indivíduos.

A partir deste produto escalar podemos definir:

- a norma de um objeto V_t :

$$\|V_t\|_{HS}^2 = \langle V_t, V_t \rangle_{HS} \stackrel{(3.2)}{=} Tr(QV_tQV_t) = Tr(QV_t)^2 = \sum_{l=1}^p (\lambda_l^t)^2, \quad (3.3)$$

onde λ_l^t representa o l -ésimo valor próprio de V_tQ ;

- a distância entre dois objetos V_t e $V_{t'}$:

$$d_{HS}(V_t, V_{t'}) = \|V_t - V_{t'}\|_{HS} = \sqrt{\|V_t\|_{HS}^2 + \|V_{t'}\|_{HS}^2 - 2\langle V_t, V_{t'} \rangle_{HS}}. \quad (3.4)$$

Este produto escalar permite determinar a matriz $Z = [Z_{tt'}]$, de dimensão $T \times T$, que é a matriz dos produtos escalares entre os quadros representados por V_t , cujo termo geral é dado por:

$$Z_{tt'} = \langle V_t, V_{t'} \rangle_{HS}. \quad (3.5)$$

Importa aqui sublinhar que quando os objetos V_t têm normas muito diferentes é aconselhável considerar os objetos normados $\frac{V_t}{\|V_t\|_{HS}}$ neutralizando desta forma o efeito do tamanho dos quadros de dados na análise. Obtemos, então uma nova matriz, \tilde{Z} , onde os quadros de dados são representadas por objetos normados e cujo termo geral é dado por:

$$\tilde{Z}_{tt'} = \frac{\langle V_t, V_{t'} \rangle_{HS}}{\|V_t\|_{HS}\|V_{t'}\|_{HS}}. \quad (3.6)$$

Z (ou \tilde{Z}) é uma matriz particularmente importante uma vez que é através da sua diagonalização que se obterá uma imagem euclidiana dos T estudos, conforme será referido posteriormente.

Coefficiente de correlação vetorial

Na análise da interestrutura recorre-se habitualmente ao coeficiente de correlação vetorial entre dois estudos t e t' , proposto por Robert e Escoufier (1976), e que é definido da seguinte forma:

Definição 3.2.2 (Coeficiente RV) A correlação vectorial entre dois estudos t e t' designa-se de coeficiente RV e é representada pela quantidade:

$$RV(t, t') = \left\langle \frac{V_t}{\|V_t\|_{HS}}, \frac{V_{t'}}{\|V_{t'}\|_{HS}} \right\rangle = \frac{Z_{tt'}}{\sqrt{Z_{tt}}\sqrt{Z_{t't'}}}. \quad (3.7)$$

Os coeficientes RV são bastante úteis para a interpretação da interestrutura uma vez que possuem as seguintes propriedades:

1. Permitem o cálculo da distância entre dois estudos normados t e t' através da relação:

$$d_{HS} \left(\frac{V_t}{\|V_t\|_{HS}}, \frac{V_{t'}}{\|V_{t'}\|_{HS}} \right) = \sqrt{2(1 - RV(t, t'))}; \quad (3.8)$$

2. Se $RV(t, t') = 1$ a distância descrita acima é nula, isto é, as estruturas em questão são coincidentes, e

$$\frac{V_t}{\|V_t\|_{HS}} = \frac{V_{t'}}{\|V_{t'}\|_{HS}}.$$

Definição da imagem euclidiana entre objetos

Para a construção da imagem euclidiana dos T objetos é necessário afetar a cada um destes objetos um peso, que serão aqui representados por π_t .

Consideremos a matriz Δ definida da seguinte forma:

$$\Delta = \begin{pmatrix} \pi_1 & & 0 \\ & \ddots & \\ 0 & & \pi_T \end{pmatrix}, \text{ onde } 0 \leq \pi_i \leq 1 \text{ e } \sum_{i=1}^T \pi_i = 1. \quad (3.9)$$

Caso seja atribuído o mesmo peso a todos os quadros de dados, tem-se que $\Delta = \frac{1}{T}I_T$, onde I_T representa a matriz identidade de ordem T . Um quadro que deva intervir para a determinação da imagem euclidiana e não para a análise e determinação dos eixos deve ser considerado suplementar e deverá ser-lhe atribuído um peso nulo.

Para comparar globalmente a estrutura das T matrizes de dados é possível obter uma representação euclidiana das mesmas submetendo a matriz Z (ou \tilde{Z} caso os objetos sejam normados) a uma ACP. Então, é necessário calcular os valores próprios e os vetores próprios da matriz $Z\Delta$ (ou $\tilde{Z}\Delta$).

Sejam:

- $\boldsymbol{\gamma}_i$ o i -ésimo vetor próprio associado ao i -ésimo maior valor próprio τ_i da matriz $Z\Delta$ (ou $\tilde{Z}\Delta$);
- M_1, \dots, M_T os pontos associados aos objetos V_1, \dots, V_T na imagem euclidiana.

As coordenadas de M_t , $t = 1, \dots, T$ sobre o i -ésimo eixo são as componentes do vetor $\sqrt{\tau_i}\boldsymbol{\gamma}_i$.

No espaço das componentes principais cada quadro vem representado por um ponto e a distância entre pontos no plano reflete o grau de semelhança entre quadros possibilitando a análise da interestrutura.

Na prática são considerados os dois primeiros eixos, ou seja, serão apenas considerados os dois primeiros vetores próprios associados aos dois maiores valores próprios, obtendo-se uma imagem euclidiana plana aproximada dos T estudos associada aos produtos escalares entre objetos.

A distância entre dois pontos M_t e $M_{t'}$ é a melhor aproximação possível no sentido da distância de *Hilbert-Schmidt* entre os objetos representativos dos quadros t e t' . A proximidade de dois pontos convenientemente representados no plano traduz a existência de uma estrutura comum entre os quadros correspondentes.

Graficamente, as posições dos pontos M_t na respectiva imagem euclidiana situar-se-ão todos no primeiro e quarto quadrante caso as coordenadas do primeiro vetor próprio sejam positivas - imagem da interestrutura não centrada.

Esta propriedade fica garantida com o **Teorema de Frobenius**.

Teorema 3.2.1 (Teorema de Frobenius) *Uma matriz simétrica com todos os seus termos positivos admite um vetor próprio associado ao maior valor próprio cujas coordenadas têm todas o mesmo sinal.*

Uma representação gráfica alternativa, designada por imagem euclidiana da interestrutura centrada, poderá ser obtida através da diagonalização da matriz $C\Delta$ (ou C , caso os pesos atribuídos aos objetos sejam iguais) com

$$C = (I_T - \mathbf{1}_T \mathbf{1}'_T \Delta) Z (I_T - \Delta \mathbf{1}_T \mathbf{1}'_T).$$

Caso sejam considerados os objetos normados $\frac{V_{t'}}{\|V_{t'}\|_{HS}}$, a matriz Z coincide com a matriz dos coeficientes RV . Estes interpretam-se como os cossenos dos ângulos entre os vetores que representam os quadros, pelo que, se existe uma estrutura comum, os ângulos são pequenos e a maior parte da variabilidade é explicada pelo primeiro eixo da representação. Por isso, unindo os pontos da imagem euclidiana com a origem do sistema de eixos, obtém-se uma estimação gráfica da correlação entre os quadros mediante

o ângulo que formam os vetores obtidos. Nesta imagem euclidiana podem-se incluir matrizes como elementos suplementares, as quais não intervêm na determinação dos eixos, mas para as quais se deseja observar o seu posicionamento na interestrutura.

3.2.3 Estudo da intraestrutura

Se a análise da interestrutura permite concluir que, efetivamente, os vários quadros de dados possuem uma estrutura similar, pode proceder-se à construção de um novo objeto, designado compromisso, que resuma a informação proveniente das configurações em estudo.

O compromisso não é mais que uma média ponderada das configurações V_t sendo, portanto, o objeto mais correlacionado com todos os quadros.

No caso de objetos não normados é definido por:

$$V = \sum_{t=1}^T \beta_t V_t;$$

e no caso de objetos normados, por:

$$V = \sum_{t=1}^T \beta_t \frac{V_t}{\|V_t\|_{HS}}.$$

Os critérios utilizados para determinar os coeficientes β_t são:

- O compromisso V é o objeto mais correlacionado (no sentido do produto escalar de *Hilbert-Schmidt*) com os objetos V_t ;
- O objeto compromisso V deverá ser um objeto da mesma natureza que os objetos V_t , isto é, o compromisso V é tal que a sua norma é igual à média ponderada das normas dos objetos V_t : $\|V\|_{HS} = \sum_{t=1}^T \pi_t \|V_t\|_{HS}$, onde π_t são os pesos definidos em 3.9.

Caso os objetos sejam normados, o compromisso V é tal que $\|V\|_{HS} = 1$, isto é, tem norma igual a 1.

Seja γ_1 o vetor próprio de $Z\Delta$ associado ao maior valor próprio τ_1 :

$$\gamma_1 = \begin{pmatrix} \gamma_1^{(1)} \\ \vdots \\ \gamma_1^{(T)} \end{pmatrix}.$$

Atendendo ao Teorema de Frobenius é possível escolher γ_1 de forma a que todas as suas coordenadas sejam positivas.

Os coeficientes β_t são determinados pela seguinte expressão:

$$\beta_t = \frac{1}{\sqrt{\tau_1}} \left(\sum_{l=1}^T \pi_l \|V_l\|_{HS} \right) \pi_t \gamma_1^t, \quad (3.10)$$

e no caso de objetos normados temos:

$$\beta_t = \frac{1}{\sqrt{\tau_1}} \pi_t \gamma_1^t.$$

Assim, a expressão que permite obter o objeto compromisso fica definida da seguinte forma:

- no caso de objetos V_t :

$$V = \sum_{t=1}^T \left[\frac{1}{\sqrt{\tau_1}} \left(\sum_{l=1}^T \pi_l \|V_l\|_{HS} \right) \pi_t \gamma_1^t \cdot V_t \right];$$

- no caso de objetos normados $\frac{V_t}{\|V_t\|_{HS}}$:

$$V = \sum_{t=1}^T \left(\frac{1}{\sqrt{\tau_1}} \pi_t \gamma_1^t \cdot \frac{V_t}{\|V_t\|_{HS}} \right).$$

Representação e interpretação do compromisso

Na imagem euclidiana dos estudos associados aos produtos escalares entre objetos, o compromisso situa-se sobre o primeiro eixo, isto é, todas as coordenadas do compromisso serão nulas à exceção da primeira:

- no caso dos objetos V_t , à distância:

$$\|V_t\|_{HS} = \sum_{t=1}^T \pi_t \|V_t\|_{HS} \text{ da origem};$$

- e no caso de objetos normados $\frac{V_t}{\|V_t\|_{HS}}$, à distância 1 da origem.

Para que a interpretação da imagem euclidiana dos objetos seja válida, os coeficientes RV de associação entre estudos devem ser elevados. Perante uma situação destas, a representação gráfica assemelha-se com a que mostra a figura 3.2.

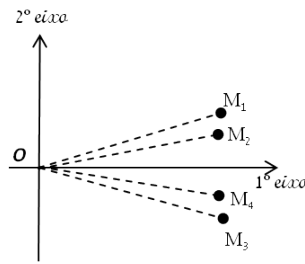


Figura 3.2: Representação de uma estrutura comum.

Os vetores \overrightarrow{OM}_t têm normas muito aproximadas e os coeficientes RV são elevados. Neste caso existe uma estrutura de variáveis comum aos vários quadros e o compromisso traduz corretamente esta estrutura.

Outras representações não refletem de forma tão adequada a estrutura dos dados.

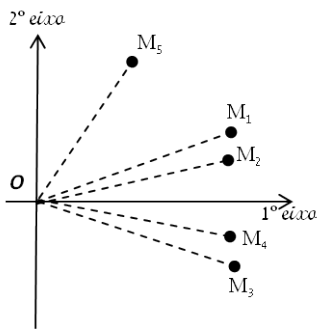


Figura 3.3: Objeto mal representado na estrutura.

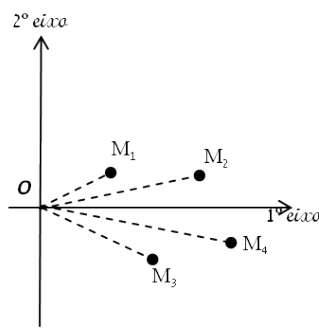


Figura 3.4: Objetos com normas bastante diferentes.

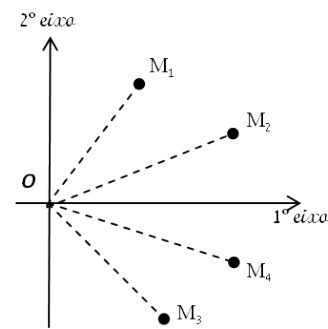


Figura 3.5: Inexistência de uma estrutura.

Na figura 3.3 o objeto M_5 intervém pouco na construção do compromisso, isto é, a tabela de dados X_5 possui uma estrutura diferente da dos restantes quadros.

Na figura 3.4 as normas são bastante diferentes. Apenas os objetos de norma mais elevada contribuirão para a construção do compromisso. Nesta situação é conveniente considerar os objetos normados $\frac{V_t}{\|V_t\|_{HS}}$ e comparar as imagens euclidianas normadas.

Por último, na situação representada na figura 3.5, os objetos são muito diferentes entre si. O compromisso é simplesmente uma média ponderada dos objetos.

A análise das representações no plano não deve ser efetuada de forma isolada sob pena de se tirarem conclusões menos corretas.

Submetendo a matriz compromisso V a uma ACP, obtém-se uma imagem euclidiana do compromisso aproximada das variáveis em estudo.

Sejam:

- μ_1, \dots, μ_p os valores próprios da matriz VQ associados aos vetores próprios $\mathbf{e}_1, \dots, \mathbf{e}_p$, respetivamente;

- B_1, \dots, B_p os pontos associados às variáveis na imagem euclidiana do compromisso.

As coordenadas de cada um destes pontos sobre o k -ésimo eixo são as componentes do vetor

$$\sqrt{\mu_k} \mathbf{e}_k, \quad k = 1, \dots, p. \quad (3.11)$$

O número de eixos a reter deverá ser definido com base nos métodos abordados no capítulo 2, página 16.

A distância entre os pontos B_i e B_j , d_{B_i, B_j} , na imagem euclidiana do compromisso interpreta-se como a distância média entre as variáveis i e j em cada estudo:

$$d_{B_i, B_j}^2 = \sum_{t=1}^T \beta_t \|x_i^t - x_j^t\|_{Q_t}^2. \quad (3.12)$$

3.2.4 Interpretação das trajetórias

Na interestrutura são evidenciadas as semelhanças e dissemelhanças entre os quadros de dados, sem as explicar. A representação e interpretação das trajetórias acompanhada de uma análise da decomposição do quadrado da distância entre pares de objetos sob a forma de percentagens de contribuições das variáveis para os afastamentos, permitirá explicar individualmente essas diferenças.

A expressão que representa a decomposição do quadrado da distância entre pares de objetos é dada por:

$$d_{HS}^2(V_t, V_{t'}) = \sum_{i=1}^p d_{ii} \sum_{j=1}^p d_{jj} [(V_t)^{ij} - (V_{t'})^{ij}]^2; \quad (3.13)$$

e a expressão que representa as percentagens de contribuições das variáveis para os afastamentos entre pares de objetos, por:

$$C_{var_i, d_{HS}^2} = \frac{d_{ii} \sum_{j=1}^p d_{jj} (V_t^{ij} - V_{t'}^{ij})^2}{d_{HS}^2(V_t, V_{t'})}, \quad (3.14)$$

onde:

- C_{var_i, d_{HS}^2} representa a contribuição da i -ésima variável para o quadrado da distância;
- d_{ii} representa o i -ésimo elemento da diagonal da matriz Q ;
- V_t^{ij} representa o elemento ij do objeto V_t .

A representação das trajetórias é efetuada na imagem euclidiana do compromisso através de uma expressão semelhante à definida em (3.11), mas considerando agora cada um dos objetos V_t :

$$\frac{1}{\sqrt{\mu_k}} V_t Q \mathbf{e}_k, \quad k = 1, \dots, T. \quad (3.15)$$

Obter-se-ão assim pT pontos que definirão p trajetórias cada uma representada por T pontos.

Caso os objetos V_t sejam normados a expressão (3.15) fica:

$$\frac{1}{\sqrt{\mu_k}} \frac{V_t}{\|V_t\|_{HS}} Q \mathbf{e}_k, \quad k = 1, \dots, T. \quad (3.16)$$

Propriedade 3.2.1 *O ponto compromisso B_i é o centro de gravidade dos pontos $B_i^1, B_i^2, \dots, B_i^T$ ponderados pelos coeficientes $\beta_1, \beta_2, \dots, \beta_T$; esta propriedade mantém-se na projeção.*

É possível distinguir dois casos em relação à forma que o sentido das trajetórias pode tomar:

- uma trajetória pouco extensa, em torno de si mesma e da sua posição compromisso, corresponde a uma variável com uma evolução muito próxima da evolução média;
- uma trajetória de grande amplitude reflete uma alteração na estrutura das variáveis de quadro para quadro, diferente da evolução média.

Tratando-se de representações gráficas aproximadas, a interpretação das trajetórias, como já foi referido, deverá ser sempre acompanhada de uma análise da decomposição do quadrado da distância entre pares de objetos sob a forma de percentagens de contribuições das variáveis para os afastamentos.

CAPÍTULO 4

A PROCRIAÇÃO MEDICAMENTE ASSISTIDA

4.1 Introdução

Para a Organização Mundial de Saúde (OMS), "Saúde" não se define simplesmente por um estado de ausência de doença, mas antes como algo que abrange um estado de completo bem-estar físico, mental e social.

Neste contexto, Saúde Reprodutiva pressupõe pessoas capazes de ter uma vida sexual responsável, satisfatória e segura, com a capacidade de reproduzir bem como a liberdade para decidir se, quando e quantas vezes fazê-lo. É portanto atribuído ao casal o direito à informação, ao acesso a métodos de fertilidade seguros, eficazes, acessíveis e aceitáveis, bem como o direito de acesso a cuidados de saúde adequados que permitam às mulheres uma gravidez saudável e aos casais as condições para terem um filho saudável (WHO, <http://www.who.int>).

Com base nesta abordagem, a OMS considera a infertilidade como uma doença que se define pela incapacidade temporária ou permanente de conceber um filho ou em levar a bom termo uma gravidez. Um casal é considerado infértil se, depois de um ano de relações sexuais regulares e sem contraceção, não ocorrer uma gravidez (Zegers-Hochschild *et al.*, 2009). Também se considera o casal infértil quando este apresenta abortamentos de repetição (3 ou mais consecutivos)(APF, <http://www.apfertilidade.org>).

Na nossa sociedade, a parentalidade é uma parte fundamental do projeto de vida de homens e mulheres, na qual as famílias e a sociedade em geral depositam grandes expectativas. Apesar de, geralmente, apenas um dos membros do casal ser alvo do diagnóstico de infertilidade, esta deve ser concetualizada como um problema de ambos, pois a "célula" do estudo é sempre o casal.

O problema de infertilidade tem vindo a assumir especial enfoque principalmente numa altura em que o contexto económico e social que vivemos se traduz, cada vez mais, numa planificação tardia da primeira gravidez por parte da mulher. Sabe-se que geralmente é a partir dos 35 anos de idade que a fertilidade feminina diminui consideravelmente.

Estima-se que a infertilidade afete cerca de 15% dos casais em idade reprodutiva em todo o mundo (WHO, 2010). Em 85% dos casos, deve-se a causas que podem ser diagnosticadas através de um estudo adequado e simultâneo do casal, cujos resultados permitirão direcionar para o tratamento mais adequado à situação. Os restantes 15% correspondem a situações em que não é possível definir uma causa - infertilidade idiopática, já que a fertilidade é afetada por fatores como a idade da mulher, de forma mais significativa a partir dos 35 anos, tipo e frequência das relações sexuais, consumo de tabaco, álcool ou drogas ilícitas, medicamentos utilizados, hábitos alimentares e estilos de vida, certos tipos de trabalho e/ou lazer, ou alterações significativas de peso (Santos *et al.* (2010)).

É importante salientar que cada uma das causas que alteram o processo reprodutivo poderá ter um tratamento concreto, que em muitas ocasiões se limita à prescrição de medicamentos ou a simples intervenções cirúrgicas para corrigir determinados defeitos anatómicos. Quando estes tratamentos, médicos e cirúrgicos, não conseguem resolver o problema recorre-se à Procriação Medicamente Assistida.

4.2 Procriação Medicamente Assistida (PMA)

Para que aconteça fecundação é necessário que ocorram relações sexuais próximas do dia da ovulação, isto é, quando o ovócito é libertado, e que haja encontro e reconhecimento entre os dois tipos de gâmetas (masculino - espermatozóide e feminino - ovócito ou ovo) para que se forme o ovo ou zigoto: o futuro embrião. É então necessário que a mulher produza um ovócito num dos seus ovários, que este seja libertado e migre até à trompa de Falópio e que ao mesmo tempo o homem produza um número suficiente de espermatozoides móveis e de qualidade para que atinjam a trompa onde se encontra o ovócito, e o consiga fecundar. Quando este processo falha sucessivas vezes por um período de um ano, o casal deverá procurar ajuda junto de um especialista em PMA.

PMA define-se pela reprodução resultante da intervenção médica, isto é, reprodução conseguida através da indução da ovulação, estimulação ovárica controlada, técnicas laboratoriais de PMA, inseminação intrauterina, intracervical e intravaginal inseminação com sémen do marido/parceiro ou dador (Zegers-Hochschild *et al.*, 2009).

O longo caminho da PMA foi possível graças a Robert Edwards (Prémio Nobel da Medicina em 2010) e Patrick Steptoe, em 1978, com o nascimento do primeiro bebé fruto

de uma Fertilização in Vitro (FIV), Louise Brown. Desde então que as técnicas de PMA experimentaram um forte desenvolvimento na maioria dos países desenvolvidos, existindo hoje várias técnicas que devolvem a esperança a milhões de casais.

Neste trabalho serão abordadas três dessas técnicas: a Inseminação artificial (IA), a Fertilização in Vitro (FIV) e a Microinjeção Intracitoplasmática de Espermatozóide (ICSI - sigla internacional) .

Inseminação Artificial

É uma técnica relativamente simples e indolor que consiste na colocação artificial de sêmen tratado laboratorialmente na cavidade uterina. O processo desenrola-se em três fases: a primeira é a estimulação ovárica e dura aproximadamente 10 dias; a segunda é a preparação do sêmen, em que se selecionam os espermatozóides móveis com maior capacidade fecundante; e a terceira diz respeito à inseminação propriamente dita, que é realizada no consultório, sem anestesia.

Geralmente é indicada para casais com infertilidade masculina ligeira, como nos casos com baixo volume de sêmen, baixa concentração ou mobilidade diminuída dos espermatozóides bem como para problemas de desempenho sexual. Mas a IA também pode ser utilizada para tratar casos de infertilidade feminina, como problemas do muco cervical ou fatores imunológicos.

A fertilização, neste caso, é *in vivo*, ocorrendo dentro das trompas de Falópio e as hipóteses de sucesso desta técnica são de aproximadamente 18 a 20%.

Fertilização in Vitro

Técnica através da qual um óvulo é fertilizado fora do corpo da mulher, ou seja, em meio laboratorial.

É um processo de quatro etapas, em que a primeira delas deve-se à hormona folículo-estimulante (FSH) que é utilizada para estimular o crescimento de vários folículos ováricos (contendo cada um deles um ovócito) ao mesmo tempo e de forma mais ou menos sincronizada. Tal desenvolvimento múltiplo de folículos aumenta as hipóteses de fertilização e de gravidez.

Na segunda etapa, a hCG (outro tipo de hormona) é usada para amadurecer e estimular a libertação dos ovócitos que são recolhidos dos ovários por via vaginal, utilizando-se uma agulha fina visualizada por ultra-som.

Na terceira etapa, os ovócitos são transferidos para o laboratório e aqui são colocados numa placa própria. Para que ocorra fecundação é-lhes adicionada uma concentração de espermatozóides seguindo-se vários dias de cultura embrionária.

Na etapa final, alguns embriões obtidos são transferidos para o interior do útero por meio de um cateter - transferência embrionária.

A FIV é indicada em casos de lesão das trompas, gravidez ectópica, laqueação irreversível das trompas de Falópio, endometriose, infertilidade masculina ligeira e em casos de infertilidade idiopática, isto é, sem causa aparente.

A taxa de sucesso associada a esta técnica varia entre os 25% e os 55%.

Microinjeção Intracitoplasmática de Espermatozóide

O procedimento é muito semelhante ao da FIV. A grande diferença reside no facto de um único espermatozóide ser microinjetado diretamente no interior do óvulo para possibilitar a fertilização nos casos em que o número de espermatozóides é muito baixo ou mesmo nos casos em que os espermatozóides não apresentam mobilidade (espermatozóides que não nadam efectivamente em direcção ao óvulo), sendo, por isso, especialmente indicada para o tratamento da infertilidade masculina grave.

Os ovócitos são sujeitos a vários procedimentos para que se obtenha a célula sexual isolada. Desta forma é possível avaliar o estado de maturação do gâmeta para que a ICSI se possa realizar, já que apenas os ovócitos maduros (em metáfase II) são seleccionados para serem microinjetados.

A taxa de sucesso desta técnica varia, como na FIV, entre os 25% e os 55%.

4.3 A PMA em Portugal

Segundo relatório do INE ("Estatísticas no Feminino, 2012), em 2010 a idade média da mulher no primeiro casamento era de 29,2 anos. O número anual de partos decresceu ao longo da primeira década do séc. XXI, tendo passado de 111616 em 2003 para 100280 partos em 2010. Neste mesmo período, a estrutura etária das parturientes alterou-se, perdendo importância relativa o grupo etário dos 25 aos 29 anos, e ganhando peso o grupo dos 30 aos 34 anos.

De facto, enquanto em 2003 a proporção de parturientes dos grupos etários 25-29 anos e 30-34 anos tinha peso idêntico (cerca de 30%), em 2010 o peso do primeiro destes grupos etários desceu para 26,5% e o do segundo subiu para 34,6%. Adicionalmente, a proporção de parturientes, com idades entre 35 e 39 anos, passou de 13,9% do total em 2003, para 18,0% em 2010.

Em Portugal, o primeiro ciclo terapêutico de FIV foi efectuado em Julho de 1985 no Hospital de Santa Maria/Faculdade de Medicina de Lisboa. A primeira criança portuguesa cuja fecundação ocorreu por FIV nasceu em Fevereiro de 1986.

No entanto, o recurso a procedimentos laboratoriais para o tratamento de casais inférteis iniciara-se no nosso país alguns meses antes da FIV, em Maio de 1985, com a execução da inseminação artificial intra-uterina - a menos complexa dessas técnicas - na Faculdade de Medicina do Porto pela mão do Prof. Doutor Alberto Barros.

Seguiu-se a introdução de variantes, como a transferência intra-tubária de gâmetas em 1986 nos Hospitais da Universidade de Coimbra, ou de técnicas sucedâneas como a criopreservação de embriões na Maternidade Dr. Alfredo da Costa, em 1990 e a ICSI pela equipa dirigida pelo Prof. Doutor Alberto Barros, em 1994 (CNPMA, <http://www.cnpma.org.pt>).

A regulamentação da PMA em Portugal data de 2006 com a publicação da Lei n.º 32/2006, de 26 de Julho que criou o Conselho Nacional de Procriação Medicamente Assistida (CNPMA) como entidade reguladora da prática desta atividade.

O ponto 2 do artigo 4.º define as condições de admissibilidade:

- 2- "A utilização de técnicas de PMA só pode verificar-se mediante diagnóstico de infertilidade ou ainda, sendo caso disso, para tratamento de doença grave ou do risco de transmissão de doenças de origem genética, infecciosa ou outras."

e o artigo 6.º define os beneficiários:

- 1- "Pessoas casadas que não se encontrem separadas judicialmente de pessoas e bens ou separadas de facto ou as que, sendo de sexo diferente, vivam em condições análogas às dos cônjuges há pelo menos dois anos."
- 2- "Pessoas com, pelo menos, 18 anos de idade que não se encontrem interditos ou incapacitados por anomalia psíquica."

Para 2010, o Ministério da Saúde fixou, entre outros, os critérios de acessibilidade a técnicas de PMA relativos à idade feminina baseados em razões de ordem clínica e de custo/benefício em saúde materna e infantil. Assim:

- Para consulta de apoio à fertilidade são admitidas todas as mulheres independentemente da idade, desde que referenciadas pelo Médico de Família;
- Para técnicas de PMA de 1ª linha (indução de ovulação e inseminação intra-uterina) são admitidas todas as mulheres que não ultrapassem os 42 anos;
- Para técnicas de PMA de 2ª linha (FIV e ICSI) são admitidas todas as mulheres que não ultrapassem os 40 anos.

Desde 2011 que o Serviço Nacional de Saúde assume o pagamento até 3 ciclos das técnicas de PMA de 1ª e 2ª linha supracitadas.

O artigo 10.º desta lei autoriza que se recorra à dádiva de ovócitos, de espermatozóides ou de embriões, prática que, até início de 2011, só era possível recorrendo a serviços privados, muito caros, ou, situação mais comum, recorrendo a clínicas espanholas. Com a criação do primeiro Banco Público de Gâmetas, autorizado pelo Despacho n.º 3219/2011 de 17 de Fevereiro, no Centro Hospitalar do Porto, EPE e financiado por verbas do Serviço Nacional de Saúde (SNS) dá-se a possibilidade de acesso a todos os utentes SNS.

Conforme informação disponível no site do CNPMA (<http://www.cnpma.org.pt>), existem no país 29 centros autorizados a ministrar técnicas de PMA. Destes, 10 são públicos sendo os restantes 19 privados. Na figura 4.1 está representada a distribuição de centros PMA, públicos e privados, no País.

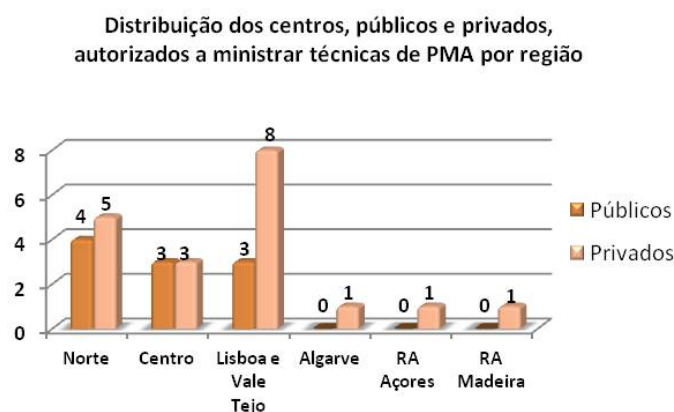


Figura 4.1: Distribuição dos centros de PMA autorizados pelas regiões do país.

Observamos que é na região Norte e na região de Lisboa e Vale do Tejo que se encontra a maior oferta. Na região do Algarve, RA Madeira e RA Açores não existem centros públicos.

Esta realidade aliada às listas de espera nos centros públicos e aos custos elevados associados aos tratamentos de PMA nos centros privados fazem da experiência da parentalidade um sonho não acessível a todos.

A título informativo e conforme informação disponibilizada pela APF, os custos variam entre 75€ e os 130€ para uma primeira consulta de infertilidade, 350€ a 570€ para uma IA, 2500€ a 3300€ para uma FIV e 3250€ a 4000€ para uma ICSI. A estes valores somam outros respeitantes a exames complementares de diagnóstico e sempre que se inicie um novo ciclo por insucesso do anterior, os valores repetem-se.

CAPÍTULO 5

ANÁLISE PRELIMINAR DOS DADOS

5.1 Apresentação dos dados

As bases de dados que irão ser analisadas contêm informação sobre 413 casais que recorreram à Unidade de Procriação Medicamente Assistida do CHP, EPE - Unidade MJD, no período de Setembro de 2010 até Outubro de 2011, e são da responsabilidade dos Técnicos Laboratoriais (Biólogos-embriologistas) do CPMA.

Foram disponibilizadas três bases de dados diferentes, cada uma registando informação sobre casais submetidos a diferentes técnicas de PMA: FIV, ICSI ou IA.

A cada uma destas bases de dados corresponde o conjunto de variáveis listado na tabela 5.1 apresentando-se uma breve descrição sobre o significado e importância destas em anexo (Anexo 1).

Conforme se pode verificar, para as bases de dados referentes às técnicas FIV e ICSI as variáveis são, na sua maioria, as mesmas, verificando-se apenas ausência da variável "concentração do esperma" (Esperma []) na base de dados de ICSI. Trata-se de duas técnicas de PMA muito semelhantes no que respeita ao estudo do casal, à estimulação ovárica, à colheita de ovócitos (vulgarmente conhecida como punção folicular) e à transferência embrionária. A diferença reside na técnica laboratorial aplicada para a punção dos gametas femininos e masculinos.

A base de dados de IA contém um número de variáveis bastante inferior, quando comparado com as outras duas bases de dados já referidas. Esta diferença justifica-se com os procedimentos e características associadas a esta técnica. Enquanto a FIV e a ICSI são técnicas "in vitro", isto é, técnicas em que a fecundação ocorre fora do corpo materno, no laboratório, a IA é uma técnica "in vivo" e não pressupõe a colheita de ovócitos nem transferência de embriões, visto a fecundação ser interna.

	FIV	ICSI	IA		FIV	ICSI	IA
IdadeF	X	X	X	EmbTBQ.	X	X	-
IdadeM	X	X	X	EmbTMQ.	X	X	-
Causa	X	X	X	EmbTMaQ.	X	X	-
Esperma[]	X	-	X	Peso	X	X	-
%EspermaB.	X	X	X	Altura	X	X	-
Fol.> 15mm	-	-	X	Grav.	X	X	X
OvObt.	X	X	-	SwimUp	-	-	X
OvInsem.	X	X	-	FSHBasal	X	X	-
Ov.NFec.	X	X	-	LHBasal	X	X	-
OvFec.	X	X	-	E2Basal	X	X	-
TxFec.	X	X	-	Gona.	X	X	-
EmbObt.	X	X	-	DiasEstim.	X	X	-
DiaTransf.	X	X	-	Estradiol	X	X	-
EmbTransf.	X	X	-	Endométrio	X	X	-

Tabela 5.1: Variáveis disponíveis em cada uma das bases de dados.

No decorrer do trabalho procedeu-se à categorização de algumas variáveis e à introdução de novas, calculadas a partir das já existentes. Fazer-se-á referência a estes dois aspetos sempre que a análise assim o exija.

Sempre que possível, isto é, sempre que as variáveis sejam comuns e sempre que a característica em estudo o justifique, serão efetuadas análises ao conjunto das bases de dados como um todo.

Serão aplicados os testes de independência do Qui-Quadrado (ou de Fisher no caso de tabelas 2×2) e o teste t para as médias entre duas amostras (FIV e ICSI) independentes, (Pestana *et al.*(2003)). O nível de significância considerado foi de 5% e foi utilizado o IBM SPSS Statistics 20 e o Microsoft Office Excel 2007 como software de suporte à análise e construção gráfica.

5.2 Análise Univariada e Bivariada

Idade Feminina

Dos 413 casais em estudo, 95 iniciaram ciclo FIV, 131 ciclo ICSI e os restantes 187 submeteram-se a IA. Originalmente cada uma das bases de dados continha um número maior de ciclos, no entanto, após uma análise preliminar, foram eliminadas linhas que continham valores em falta.

Sabemos que a fertilidade feminina é especialmente afetada pelo aumento da idade.

Na tabela 5.2 encontram-se as estatísticas descritivas da variável idade feminina observadas em cada uma das bases de dados.

Técnica de Fertilidade	Total de casais	IDADE FEMININA					
		Média	Moda	Mediana	Desvio Padrão	Mínimo	Máximo
FIV	95	33,2	33	33	3,7	20	40
ICSI	131	33,3	36	34	4,1	22	40
IA	187	31,6	30 e 34	32	3,8	22	41
TOTAL	413	32,5	34	33	4,0	20	41

Tabela 5.2: Estatísticas descritivas da variável "IdadeF".

A idade média das mulheres submetidas a FIV e a ICSI são bastante próximas, 33,2 e 33,3 respetivamente, embora na base de dados ICSI esta variável apresente valores superiores para a moda, mediana e desvio padrão.

Na IA verifica-se que a idade média feminina é inferior: 31,6 anos de idade. Sendo esta uma técnica menos invasiva é sempre preferida a outras quando no casal não se diagnosticam fatores de infertilidade masculina graves, nem a idade feminina se encontra próxima dos 35.

O limite máximo de idade observado em cada uma das técnicas encontra-se de acordo com os critérios fixados pela ACSS na Circular Normativa n.º9 de 2010 (12/Agosto).

Uma vez que recorrem a esta Unidade casais de diferentes culturas é natural que se registem idades bastante jovens, 20 anos de idade no caso dos dados analisados, mas que já registam infertilidade conjugal superior a 1 ano.

Para estudar a relação entre a idade da mulher e o sucesso de um ciclo foi criada a variável "GIdade" que resulta da categorização da variável idade em grupos definidos com base em estudos realizados sobre o tema. Sabe-se que a partir dos 35 anos a fertilidade feminina diminui, sendo esta mais marcante depois dos 37 anos e tornando-se muito acentuada depois dos 40 anos de idade.

Uma vez que, como já foi referido, no CPMA não encontraremos muitas mulheres com idade igual ou superior a 40, decidiu-se pela criação das 4 categorias para a variável idade listadas na tabela 5.3.

Uma primeira análise considerou o total de casais em estudo no conjunto das três técnicas.

Foi aplicado o teste de independência do Qui-Quadrado às variáveis "GIdade" e "Grav" e não foi encontrada evidência estatística que permita sustentar uma relação de dependência entre estas duas variáveis, a um nível de significância de 5%. Podendo esta observação estar associada ao critério definido para a categorização da variável, fez-se

Categoria	Critério de agrupamento
Gid1 - ≤ 30	IdadeF menor ou igual a 30
Gid2 - $]30, 35[$	IdadeF superior a 30 e inferior a 35
Gid3 - $[35, 38[$	IdadeF igual ou superior a 35 e inferior a 38
Gid4 - ≥ 38	IdadeF igual ou superior a 38

Tabela 5.3: Categorias da variável "IdadeF".

também uma análise considerando apenas 3 categorias: < 35 anos, $[35; 38[$ e ≥ 38 anos, no entanto os resultados continuaram no sentido da independência entre a idade e o resultado do ciclo.

Seguidamente procedeu-se à mesma análise mas considerando cada base de dados individualmente.

Para casais submetidos a FIV os resultados do teste obtidos, que se encontram no tabela 5.4, são no sentido de rejeitar a hipótese de independência entre a idade da mulher e a obtenção de gravidez, para o nível de significância usual de 5%, ($p\text{-value} = 0,042 < 0,05$).

Chi-Square Tests				
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	8,216 a.	3	,042	0,0425226
Likelihood Ratio	8,820	3	,032	0,0379293
Fisher's Exact Test	8,1558			0,0428404
N of Valid Cases	95			

a. 0 cells (,0%) have expected count less than 5.

The minimum expected count is 25,53.

Tabela 5.4: Teste Qui-Quadrado: variáveis "Grav" vs "IdadeF" para casais em FIV.

No caso da FIV que sabemos estar mais indicada para casais cuja causa de infertilidade é essencialmente feminina a relação de dependência obtida justifica-se por si só, isto é, se uma mulher apresenta fatores menos favoráveis à obtenção de uma gravidez é natural que estes se acentuem e sejam mais difíceis de contornar com o avançar da idade. A análise dos resíduos ajustados, tabela 5.5, permite-nos constatar que, quando submetidas a FIV, mulheres com idades entre os 30 e os 34 anos são mais favoráveis a conseguir uma gravidez e que mulheres com idade superior a 38 anos terão mais dificuldade em consegui-lo. Esta análise resulta da observação das células que apresentam o valor dos resíduos ajustados superior, em módulo, a 1,96.

No entanto, observando os valores obtidos para as medidas de associação baseadas no Qui-Quadrado, tabela 5.6, verificamos que o grau de associação entre estas duas variáveis

	Resíduos Ajustados	
	Resultado	
	Grav.	Não Grav.
Gid1	0,6	-0,6
Gid2	2,1	-2,1
Gid3	-1,4	1,4
Gid4	-2,1	2,1

Tabela 5.5: Resíduos ajustados: variáveis "Grav" vs "IdadeF" para casais em FIV.

é fraco. Repare-se que caso estivesse a ser considerado um nível de significância de 1% a interpretação seria diferente, isto é, a hipótese de independência não seria rejeitada.

Symmetric Measures		
	Value	Approx. Sig.
Phi	0,2941	0,0417
Cramer's V	0,2941	0,0417
Contingency Coefficient	0,2821	0,0417
N of Valid Cases	95	

Tabela 5.6: Medidas de associação baseadas no Qui-Quadrado.

Somos então levados a concluir que, na maioria dos casos, a idade feminina superior a 35 anos é um fator importante para o resultado de um ciclo se associada a um conjunto de outros fatores que dificultam o processo.

Além da idade da mulher, os valores hormonais basais (isto é, valores medidos no início do ciclo) como o FSH basal, o LH basal e o E2 basal, os valores registados após tratamento para o nível de estradiol, a espessura do endométrio, as gonadotrofinas, a percentagem de espermatozóides com morfologia normal e o número de ovócitos inseminados devem também ser analisados quando se pretende avaliar as possibilidades de sucesso num tratamento de medicina reprodutiva.

A tabela 5.7 apresenta as estatísticas descritivas para estas variáveis.

Relembrando que a ICSI é uma técnica especialmente indicada para o tratamento de casais com fator de infertilidade masculina grave, seria de esperar que a percentagem de espermatozóides com morfologia normal (%EspermaB) fosse, em média, inferior quando comparada com os valores obtidos para a FIV.

Outras diferenças que se destacam entre o valor destas variáveis registado em cada uma das técnicas são, na dose de gonadotrofinas administrada ("Gona"), "E2Basal" e no número de ovócitos inseminados ("OvInsem").

Para perceber se as diferenças entre as médias são significativas, procedeu-se à aplicação do teste *t* para duas amostras independentes.

CAPÍTULO 5. ANÁLISE PRELIMINAR DOS DADOS

	Média		Desvio Padrão		Moda		Mediana		Máximo		Mínimo	
	FIV	ICSI	FIV	ICSI	FIV	ICSI	FIV	ICSI	FIV	ICSI	FIV	ICSI
%EspermaB	8,4	3,9	5,0	3,2	5,0	2,0	7,0	3,0	26,0	15,0	1,0	0,0
OvInsem	8,2	6,3	4,5	3,6	8,0	6,0	8,0	6,0	21,0	19,0	1,0	1,0
FSHBasal	7,3	6,9	3,5	2,4	6,5	6,6	6,7	6,5	36,0	22,1	3,0	2,1
LHBasal	6,0	5,8	3,4	2,4	3,0	4,9	5,3	5,1	30,0	15,6	2,0	1,9
E2Basal	61,2	57,0	49,9	38,8	52,8	42,0	48,0	44,7	319,0	253,0	13,0	11,4
Gona.	2083,1	1941,6	1746,8	1008,1	1200,0	1350,0	1575,0	1575,0	12375,0	5400,0	675,0	800,0
Estradiol	1651,6	1657,2	1088,3	929,2	1332,0	1477,0	1417,0	1381,0	9157,0	4300,0	92,6	414,3
Endométrio	10,4	10,3	2,1	1,9	12,5	12,0	10,2	10,2	19,0	15,8	6,0	6,4

Tabela 5.7: Estatísticas descritivas de algumas variáveis quantitativas.

Obteve-se significância estatística quando comparadas as médias entre as variáveis ”%EspermaB” e ”OvInsem” em ambos os grupos, tabela 5.8 e tabela 5.9 respetivamente.

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval	
								Lower	Upper
Equal variances assumed	21,448	,000	5,630	224	,000	3,160	,561	2,054	4,266
Equal variances not assumed			5,250	146,530	,000	3,160	,602	1,970	4,350

Tabela 5.8: Teste *t* para a diferença das médias entre os grupos FIV e ICSI relativamente à variável ”%EspermaB”.

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval	
								Lower	Upper
Equal variances assumed	3,833	,051	3,288	224	,001	1,769	,538	,709	2,829
Equal variances not assumed			3,175	174,503	,002	1,769	,557	,669	2,869

Tabela 5.9: Teste *t* para a diferença das médias entre os grupos FIV e ICSI relativamente à variável ”OvInsem”.

As diferenças de 3,160 e de 1,769 para as médias observadas nas duas técnicas quanto à variável ”%EspermaB” e à variável ”OvInsem”, respetivamente, são estatisticamente

significativas ($p\text{-value} < 0,0001 < 0,05$ e $p\text{-value} = 0,001 < 0,05$, respetivamente), registando-se valores superiores para ambas em FIV.

No que diz respeito ao número de ovócitos inseminados também não é de estranhar esta diferença para a média entre as duas técnicas. A inseminação em FIV não permite avaliar inicialmente o estado de maturação do ovócito pelo que normalmente são inseminados mais do que na técnica ICSI, onde só se inseminam ovócitos que apresentem aspetos morfológicos que indiquem a sua maturidade.

Taxa de gravidez

A PMA representa uma esperança para os casais que não respondem a outros tratamentos e que veem, desta forma, dificultado o processo de conceber um filho. Independentemente da técnica de fertilidade adotada, o objetivo é um: responder de forma positiva às expetativas dos casais proporcionando-lhes a tão desejada gravidez.

As taxas de sucesso de um ciclo variam consoante a técnica de fertilidade adotada. Estudos indicam que as taxas de sucesso associadas a FIV e a ICSI são semelhantes e variam entre os 24% e os 55%, já para IA a taxa diminui significativamente estando avaliada entre os 18% e os 20%.

Na figura 5.1 está representada a distribuição da taxa de gravidez associada a cada base de dados analisada neste trabalho.

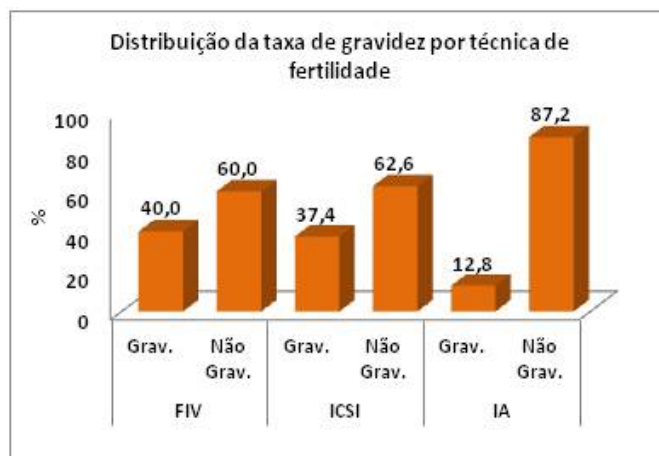


Figura 5.1: Distribuição da taxa de sucesso por técnica de fertilidade.

Conforme se pode verificar, e de acordo com o que refere a literatura, das três técnicas de fertilidade representadas, a IA é aquela que apresenta menor taxa de sucesso por ciclo, 12,8%. Para as técnicas FIV e ICSI a taxa de gravidez obtida é bastante semelhante embora ligeiramente superior na FIV, 40% e 37,4% respetivamente.

Apesar da taxa de sucesso associada, como já foi referido, a IA é, das três técnicas de PMA aqui referidas, a menos invasiva e a que mais se aproxima da conceção natural.

Mostrou-se interesse em perceber se, para os dados analisados, existe evidência estatística que permita relacionar a técnica de fertilidade com o resultado obtido no final do ciclo: gravidez ou não gravidez. O teste de independência do Qui-Quadrado devolveu os resultados que se observam nas tabelas 5.10 e 5.11.

Com um $p\text{-value} < 0,0001 < 0,05$, somos levados a rejeitar a hipótese de independência, isto é, a rejeitar que o sucesso do ciclo é independente da técnica de fertilidade adotada, a 5% nível de significância. Neste caso a rejeição da hipótese de independência mantém-se considerando um nível de significância de 1%.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	34,476 ^a	2	,000	0,00
Likelihood Ratio	36,355	2	,000	0,00
Fisher's Exact Test	36,182			0,00
N of Valid Cases	413			

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 25,53.

Tabela 5.10: Qui-Quadrado: variáveis "Técnica" vs "Grav".

		Obteve Gravidez?		Total	
		Grav.	Não Grav.		
Técnica fertilidade	FIV	Count	38	57	95
		Expected Count	25,5	69,5	95,0
		Adjusted Residual	3,3	-3,3	
	IA	Count	24	163	187
		Expected Count	50,3	136,7	187,0
		Adjusted Residual	-5,9	5,9	
	ICSI	Count	49	82	131
		Expected Count	35,2	95,8	131,0
		Adjusted Residual	3,3	-3,3	
Total	Count	111	302	413	
	Expected Count	111,0	302,0	413,0	

Tabela 5.11: Resíduos ajustados: variáveis "Técnica" vs "Grav".

Os valores dos resíduos ajustados obtidos para a IA são aqueles que mais contribuem para a rejeição da hipótese de independência, sendo de esperar que casais submetidos a esta técnica tenham menor taxa de sucesso como, aliás, já foi referido.

Houve interesse, por parte da equipa do CPMA, em verificar se existiria alguma relação entre:

- o número de dias de estimulação ovárica e a obtenção de gravidez;
- o dia de transferência de embriões e a obtenção de gravidez.

Para ambas as questões foram analisados os casais submetidos às técnicas FIV e ICSI, no seu conjunto e individualmente, já que as variáveis de interesse fazem parte destas duas bases de dados.

Para a primeira análise, isto é, para a relação existente entre o número de dias de estimulação e a obtenção de gravidez procedeu-se à categorização da variável "DiasEstim" em dois grupos:

- Grupo 1: constituído pelos casais cuja estimulação da ovulação durou 7 ou menos dias.
- Grupo 2: constituído pelos casais cuja estimulação da ovulação durou 8 ou mais dias.

Esta divisão foi sugerida pelas Embriologistas da Unidade no sentido de averiguar se em ciclos denominados por curtos, os dias de estimulação são suficientes para originar ovócitos com viabilidade e obter, desta forma, a gravidez pretendida, uma vez que se julga que uma estimulação inferior a 7 dias não dá tempo suficiente para a síntese de moléculas essenciais ao bom desenvolvimento embrionário.

No conjunto dos casais em FIV e em ICSI não foi encontrada evidência estatística que permita rejeitar a hipótese de independência, ou seja, o número de dias de estimulação é independente do sucesso do ciclo.

Quando analisada cada base de dados individualmente, o teste de Fisher forneceu os resultados que se encontram na tabela 5.12 sendo o $p\text{-value} = 0,039 < 0,05$ para casais em ICSI.

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,469 ^a	1	,035	,039	,029
Continuity Correction ^b	3,625	1	,057		
Likelihood Ratio	4,367	1	,037	,057	,029
Fisher's Exact Test				,039	,029
N of Valid Cases	131				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10,00.
b. Computed only for a 2x2 table

Tabela 5.12: Qui-Quadrado: variáveis "DiasEstim" vs "Grav".

Somos, assim, levados a rejeitar a hipótese de independência e a concluir que o dia em que é feita estimulação da ovulação influencia o sucesso do ciclo em casais submetidos a ICSI.

Nesta análise, considerando um nível de significância de 1%, a conclusão seria no sentido da aceitação da hipótese de independência.

Relativamente à relação entre o dia da transferência dos embriões e a obtenção de gravidez, nenhuma das estatísticas de teste obtidas, analisando FIV e ICSI separadamente e analisando o total das duas, permite rejeitar a hipótese de independência.

Causas de infertilidade

A decisão pela técnica de PMA a que um casal irá ser submetido não pode ser eleita com base nas taxas de sucesso. Esta depende de vários fatores clínicos entre os quais a causa que determina a infertilidade do casal.

Relembrando, a técnica FIV está mais indicada para casais cuja infertilidade tem causa essencialmente feminina. Já para a técnica ICSI verifica-se o oposto, um casal é proposto a ICSI quando a causa de infertilidade é predominantemente masculina e grave. Em IA estão associadas causas masculinas e femininas de grau ligeiro.

Consoante o diagnóstico realizado a cada casal para avaliar a causa de infertilidade associada, foi atribuída uma de quatro categorias definidas da seguinte forma:

Categoria	Descrição
1 - Masc.	Fatores exclusivamente masculinos (ex: alterações no espermograma).
2 - Masc.+ Fem.	Associação de fatores masculinos e femininos.
3 - Fem.	Fatores exclusivamente femininos (ex: obstrução tubar, fatores ovulatórios).
4 - Idiopática	Quando a causa de infertilidade não está diagnosticada.

Tabela 5.13: Categorias da variável "Causa".

Na figura 5.2 está representada a distribuição das causas associadas a cada casal para cada uma das três técnicas em análise.

Tal como seria de esperar, está evidenciada uma prevalência das causas de infertilidade masculinas nas técnicas ICSI e IA e de causas predominantemente femininas em FIV.

Índice de Massa Corporal

No âmbito da infertilidade, sabe-se que um peso fora do normal, por excesso ou por defeito, é um fator que pode contribuir para o não sucesso de um ciclo. Normalmente, mulheres obesas ou excessivamente magras têm uma ovulação deficiente, tendo por isso maiores dificuldades em engravidar (Brannian (2102), Luke (2011)).

Para explorar esta situação nos dados em estudo foi calculado o Índice de Massa Corporal, IMC, de cada mulher a partir das variáveis Peso e Altura: $IMC = \frac{Peso}{Altura^2}$.

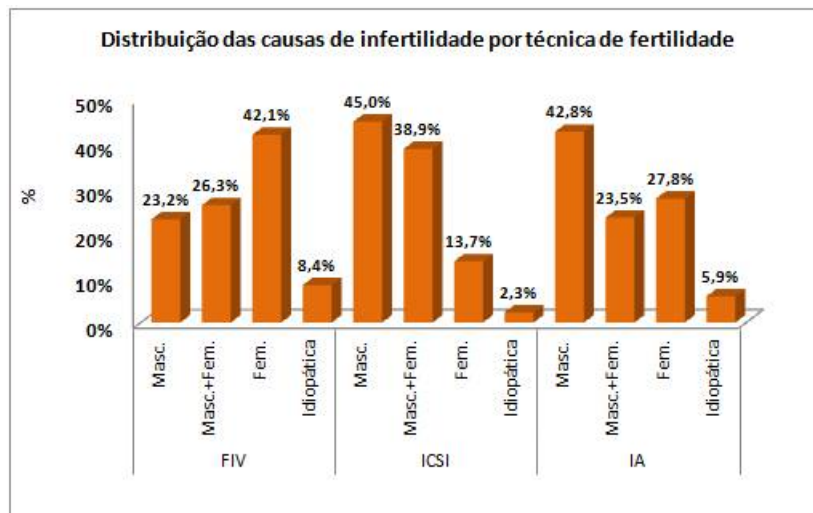


Figura 5.2: Distribuição das causas de infertilidade por técnica de fertilidade.

Precedeu-se, de seguida, à categorização desta nova variável em 4 grupos, conforme tabela 5.14 .

Categoria	Critério de agrupamento
Demasiado magra	IMC < 18,5
Peso normal	IMC ∈ [18,5; 25[
Peso elevado	IMC ∈ [25; 30[
Peso excessivo	IMC ≥ 30

Tabela 5.14: Categorias da variável "IMC".

No caso dos casais analisados neste trabalho não foi encontrada evidência estatística que permita confirmar a existência de associação entre o peso da mulher e o sucesso do ciclo. Foram analisados os casais submetidos a FIV e ICSI em conjunto, casais em FIV e casais em ICSI separados e foram ainda consideradas apenas duas classes na categoria de IMC: uma para Peso Normal e outra para Peso fora do Normal que engloba as restantes três categorias: Demasiado magro, Peso elevado e Peso excessivo. Em nenhuma destas situações foi encontrada evidência estatística que permita ir ao encontro do que refere a literatura.

Mais uma vez fica evidenciado que fatores de risco isolados não ditam o insucesso de um ciclo, sendo este resultado de uma conjugação de vários fatores. Assim se justifica a necessidade e a importância de uma análise conjunta às variáveis de interesse na tentativa de perceber quais as responsáveis pelas diferenças entre quadros de dados.

6.1 Introdução

Neste capítulo irá ser efetuada uma análise multivariada aos dados apresentados no capítulo anterior à luz da metodologia STATIS Dual.

Considerar-se-ão quatro quadros de dados construídos com base no cruzamento das variáveis "Grav" (que pode assumir uma das categorias: engravidou ou não engravidou) e "Técnica" (categorizada entre FIV ou ICSI): FIVGrav, FIVNaoGrav, ICSIGrav e ICSINaoGrav.

Pretende-se averiguar sobre uma possível estrutura comum às matrizes que compõem cada estudo, bem como perceber que variáveis contribuem mais para as semelhanças e dissemelhanças entre os vários quadros de dados.

A metodologia STATIS Dual implementada no SPAD (CISIA, 1997) considera os dados centrados e reduzidos, no entanto, não trabalha com os objetos normados. Para ultrapassar esta questão foi implementado um programa em linguagem R, que devolve os resultados mais importantes de cada uma das fases da metodologia considerando os objetos normados, caso a análise o indique.

Na fase inicial de implementação foi desenvolvido em paralelo um exemplo de uma análise em MSEXcel, que considerou objetos normados e objetos não normados, cujos resultados foram sendo confirmados quer pelos resultados do programa em R quer pelos resultados do SPAD, respetivamente. Desta forma foi possível detetar erros de cálculo e adquirir um maior entendimento sobre a metodologia. Para construção gráfica foram utilizados os programas R e MSEXcel.

O estudo inicia com uma análise preliminar aos quadros de dados que o compõem passando-se posteriormente para a aplicação da metodologia STATIS Dual considerando

cada uma das etapas que a definem.

6.2 Caraterização dos quadros de dados

A Análise STATIS Dual foi efetuada a quatro grupos de dados. Cada grupo corresponde a um resultado obtido no final do ciclo: gravidez ou não gravidez, associado a uma de duas técnicas: FIV e ICSI. A base de dados de IA foi deixada de parte uma vez que, conforme referido no início do capítulo 5, é composta por um número de variáveis muito inferior ao registado para as outras duas técnicas.

Nesta análise pretendeu-se perceber quais as variáveis que se destacam quando se comparam grupos de casais que conseguiram a gravidez com os que não engravidaram tendo sido submetidos a duas técnicas de fertilidade diferentes.

Construíram-se, assim, quatro grupos:

- Casais em FIV que engravidaram (FIVGrav);
- Casais em FIV que não engravidaram (FIVNaoGrav);
- Casais em ICSI que engravidaram (ICSIGrav);
- Casais em ICSI que não engravidaram (ICSINaoGrav).

Esta abordagem considera um total de 226 casais distribuídos pelos quatro grupos em análise conforme tabela 6.1:

Técnica Fertilidade	Resultado do ciclo		Total
	Não Grav.	Grav.	
FIV	57	38	95
ICSI	82	49	131
Total	139	87	226

Tabela 6.1: N° de casais por técnica e por resultado do ciclo.

O grupo de casais submetidos a ICSI representa 58% do total de casais em análise, correspondendo os restantes 42% a casais em FIV. A percentagem de casais que engravidou é de 35,5%.

Considerando o total de casais por técnica verificamos que a percentagem de casais que engravidaram é muito próxima, 40,0% (38/95) e 37,4% (49/131) para FIV e ICSI, respetivamente.

Para cada quadro os dados foram centrados e reduzidos eliminando-se desta forma os efeitos das diferentes escalas de medida das variáveis.

A partir destes, e para estudar as relações estatísticas entre as variáveis, foram calculadas as matrizes de correlações que, neste caso, coincidem com a matriz de variâncias e covariâncias para cada um dos grupos, e que se encontram em anexo (Anexo 2).

Da análise das matrizes de correlação destaca-se:

- A forte correlação negativa entre o par de variáveis "EmbTBQ" e "EmbTMQ" nos grupos FIVGrav (-0,846) e ICSIGrav (-0,849).

Os embriões obtidos são classificados quanto à sua qualidade em bons, médios ou maus. Para aumentar a probabilidade de sucesso do ciclo serão seleccionados para transferir o maior número de embriões de boa qualidade que existam (no máximo 2 e em casos excepcionais 3). Naturalmente quanto maior o número de embriões de boa qualidade transferidos menor será o número de embriões de qualidade média ou má transferidos e vice-versa, justificando desta forma a correlação negativa observada entre estas duas variáveis. A variável que regista o número de embriões transferidos de má qualidade, "EMBTMaQ", foi eliminada nesta análise uma vez que para o quadro de dados de FIVGrav esta apresentava o mesmo valor, zero, para os 38 registos;

- A correlação positiva entre as variáveis "OvInsem" e "OvFec" nos grupos ICSIGrav (0,827) e ICSINaoGrav (0,792).

De entre os procedimentos que caracterizam a técnica ICSI, aquele que diferencia esta da técnica FIV prende-se com o facto de serem microinjetados (inseminados) apenas os ovócitos maduros. O "trabalho" do espermatozóide é manipulado, isto é, um único espermatozóide é injetado diretamente no ovócito. Pressupõe-se portanto que quanto maior o número de ovócitos inseminados maior será o número de ovócitos fecundados.

Na FIV, além dos ovócitos serem recolhidos sem se saber qual o seu estado de maturação, são colocados vários espermatozóides e espera-se que um se funde com o ovócito para que ocorra a fecundação. Assim se explica que em FIV não se verifique uma correlação forte entre as duas variáveis.

Para analisar e comparar os valores que cada variável toma em cada um dos grupos, procedeu-se ao cálculo da média de cada uma e à sua representação gráfica. Utilizaram-se gráficos estrela, figura 6.1, já que permitem uma fácil visualização comparativa do comportamento da mesma variável observada em cada um dos 4 grupos.

Observa-se que a variável idade feminina ("IdadeF") apresenta valores médios ligeiramente superiores nos grupos em que não se obteve a gravidez, registando o seu maior valor médio no grupo FIVNaoGrav.

CAPÍTULO 6. ANÁLISE STATIS DUAL

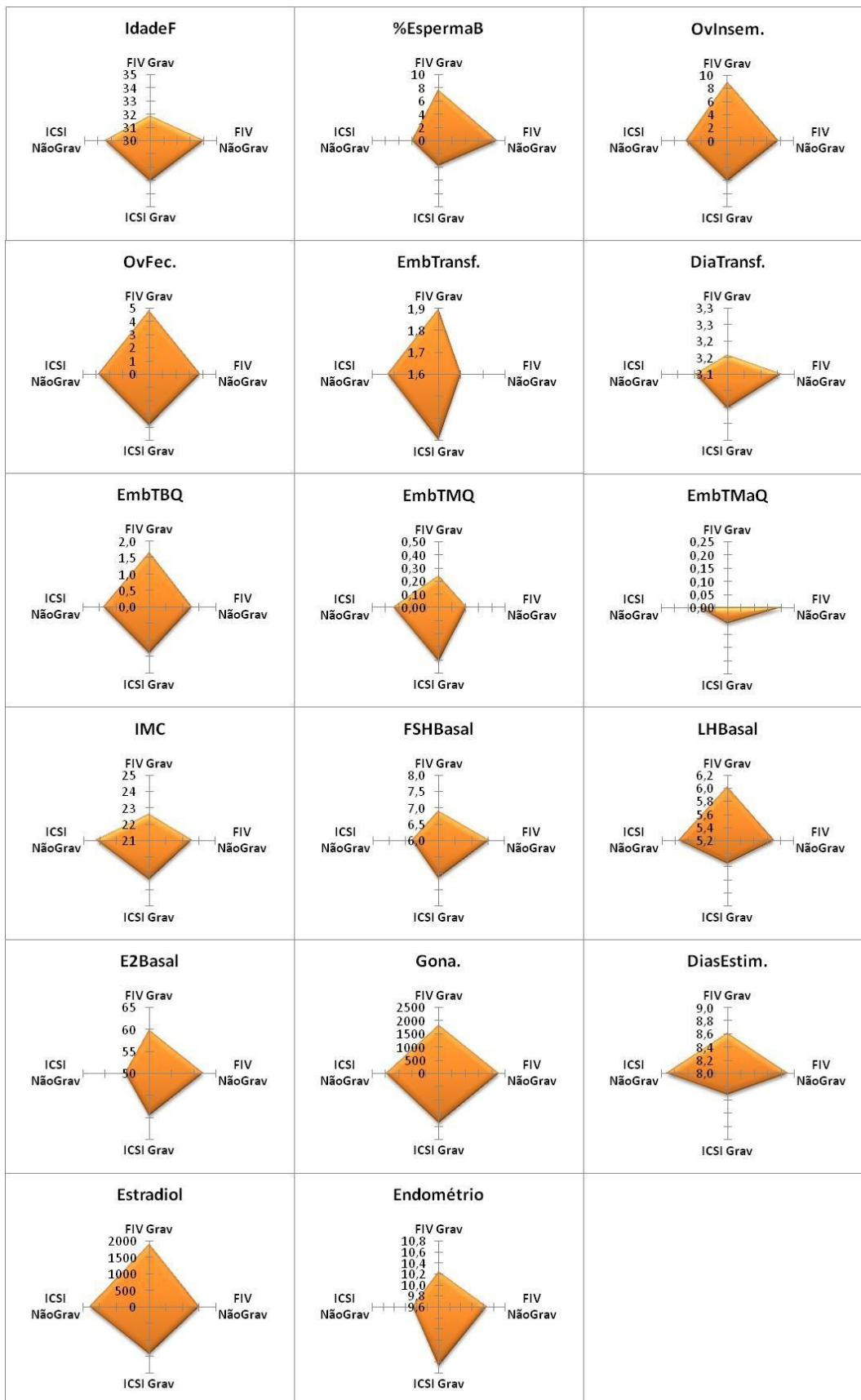


Figura 6.1: Gráficos Estrela para as variáveis em estudo.

Também no que respeita ao número de embriões de má qualidade transferidos ("EmbT-MaQ"), dose de gonadotrofinas administrada ("Gona") e número de dias de estimulação ovárica ("DiasEstim") os maiores valores médios registam-se nos dois grupos em que o ciclo não resultou em gravidez.

Já se analisarmos as variáveis "FSHBasal" e "E2Basal", comparando ainda os mesmos dois grupos, verificamos que o menor valor médio destas se regista no grupo ICSI-NaoGrav, e que o maior regista-se no grupo FIVNaoGrav.

O índice de massa corporal feminino ("IMC") apresenta valores médios superiores em ICSINaoGrav. Neste grupo observa-se o menor valor médio para a espessura do endométrio ("Endométrio") opondo-se a ICSIGrav onde se regista o maior valor médio para a mesma variável.

Nos grupos compostos por casais que conseguiram obter uma gravidez destacam-se os maiores valores médios para o número de embriões transferidos ("EmbTransf") acompanhados por um número elevado de embriões de boa qualidade ("EmbTBQ") e um número baixo (nulo no caso de FIV) para embriões transferidos de má qualidade ("EmbTMaQ").

A variável "LHBasal" apresenta valores médios semelhantes para os dois grupos em que não se verificou a gravidez, e valores extremos para os grupos em que se obteve gravidez, isto é, o valor médio mais elevado no grupo FIVGrav e o mais baixo em ICSIGrav. Se analisarmos a variável "E2Basal" verificamos que devolve uma análise semelhante à anterior alterando no entanto os grupos correspondentes no que diz respeito à obtenção de gravidez, ou seja, esta variável apresenta valores médios semelhantes para os grupos onde se obteve a gravidez e valores extremos nos grupos onde não se obteve gravidez, sendo o maior correspondente ao grupo FIVNaoGrav e o menor ao grupo ICSINaoGrav.

É no número de ovócitos inseminados ("OvInsem") e na percentagem de espermatozóides de morfologia normal ("%EspermaB") que se notam as maiores diferenças entre as duas técnicas (independentemente do resultado obtido), valores estes superiores em FIV e inferiores em ICSI e que se justificam com as características e finalidades a que cada técnica se propõe, conforme já referido anteriormente.

6.2.1 Análise STATIS Dual

Esta primeira análise é constituída por quatro quadros de dados compostos por 17 variáveis cada, sendo o número de observações diferente de quadro para quadro, conforme ilustra a figura 6.2. Os estudos são do tipo (X_t, Q, D_t) , com $t \in \{1,2,3,4\}$ correspondendo a FIVGrav, FIVNaoGrav, ICSIGrav, ICSINaoGrav, respetivamente. X_t representa o quadro de dados na circunstância t , Q é a métrica associada aos indivíduos, a matriz identidade I_{17} no nosso caso, e D_t é a métrica associada às variáveis, $D_t = \text{diag}(1/n_t)$, $t \in \{1,2,3,4\}$

e $n_1=38$, $n_2=57$, $n_3=49$ e $n_4=82$).

Uma vez que se pretende dar a mesma importância a cada um dos estudos considerou-se a matriz dos pesos Δ da forma $\Delta = \frac{1}{4}I_4$, sendo I_4 a matriz identidade de ordem 4×4 .

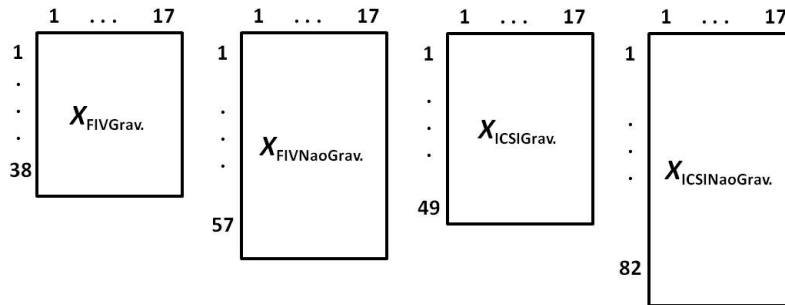


Figura 6.2: Estrutura da análise.

Interestruturura

Conforme descrito no capítulo 3 o estudo da interestruturura pretende evidenciar as relações entre os diferentes quadros de dados através da comparação global da estrutura dos mesmos.

Cada objeto representa um técnica (FIV ou ICSI) associada a um resultado (gravidez ou não gravidez); assim, o estudo da interestruturura permitirá perceber quais os pares de quadros que mais se aproximam e quais os que mais se distanciam, como por exemplo como se relacionam técnicas diferentes associadas ao mesmo resultado e técnicas iguais associadas a resultados diferentes.

As normas dos objetos V_t , tabela 6.2, são da mesma ordem de grandeza variando entre 5,24 e 5,65 o que nos permite a implementação da metodologia STATIS Dual a objetos não normados.

$\ V_1\ $	$\ V_2\ $	$\ V_3\ $	$\ V_4\ $
5,425	5,365	5,646	5,236

Tabela 6.2: Normas dos objetos.

A correlação vetorial entre estudos, tabela 6.3, e a matriz das distâncias, tabela 6.4, permitem-nos perceber quais os quadros mais próximos e quais os mais afastados.

O maior coeficiente RV , 0,857, verifica-se entre os grupos FIVNaoGrav e ICSINaoGrav, sendo este o par de quadros de dados que se encontra mais próximo.

Os grupos mais afastados correspondem aqueles cujo coeficiente RV é menor, 0,751 (e cuja distância euclidiana é maior, 3,91), isto é, correspondem aos grupos de casais

	FIVGrav	FIVNaoGrav	ICSIGrav	ICSINaoGrav
FIVGrav	1			
FIVNaoGrav	0,818	1		
ICSIGrav	0,751	0,783	1	
ICSINaoGrav	0,794	0,857	0,808	1

Tabela 6.3: Coeficientes *RV*.

	FIVGrav	FIVNaoGrav	ICSIGrav	ICSINaoGrav
FIVGrav	0,00			
FIVNaoGrav	3,25	0,00		
ICSIGrav	3,91	3,64	0,00	
ICSINaoGrav	3,43	2,84	3,40	0,00

Tabela 6.4: Distâncias euclidianas.

que obtiveram sucesso no final do ciclo tendo sido submetidos a técnicas de fertilidade diferentes, FIVGrav e ICSIGrav.

As distâncias entre os quadros de dados apresentam valores razoavelmente elevados uma vez que os objetos não foram normados.

Na tentativa de induzir uma ordem dos pares de quadros mais semelhantes para os menos semelhantes, poderemos concluir que no topo da lista encontramos os grupos que resultaram em não gravidez, independentemente da técnica a que foram submetidos, seguindo-se os quadros que correspondem à mesma técnica mas cujos resultados obtidos foram diferentes, (FIVGrav e FIVNaoGrav a uma distância 3,25 e ICSIGrav e ICSINaoGrav à distancia de 3,4) e terminando com os grupos de casais que apresentam maiores diferenças entre eles, ou seja, os que conseguiram a gravidez.

Esta análise por si só demonstra a complexidade associada à problemática em análise. Repare-se que são mais semelhantes grupos que correspondem a técnicas diferentes com resultados diferentes (ICSINaoGrav com FIVGrav ou ICSIGrav com FIVNaoGrav) do que os grupos que correspondem ao mesmo resultado: gravidez (FIVGrav e ICSIGrav).

A imagem euclidiana da interestrutura não centrada, figura 6.3, obtida através da diagonalização da matriz dos produtos escalares entre objetos V_i , evidencia a oposição entre o grupo FIVGrav e ICSIGrav sobre o segundo eixo. Uma vez que a maior parte da variabilidade é explicada pelo primeiro eixo da representação, 85% e que os coeficientes *RV* são relativamente elevados somos conduzidos a afirmar que existe uma estrutura comum entre os estudos.

Também se procedeu à representação da imagem euclidiana da interestrutura centrada, figura 6.4. Os quadros de dados relativos a FIVGrav e a ICSIGrav são os que se distanciam mais do comportamento médio uma vez que são os mais afastados da origem dos eixos,

isto é, os registos verificados em cada um destes grupos apresentam maior variabilidade. Por outro lado, parece existir um comportamento similar entre os grupos de casais que não engravidaram mesmo tendo sido submetidos a técnicas de fertilidade diferentes.

Imagem euclidiana não centrada da interestrutura

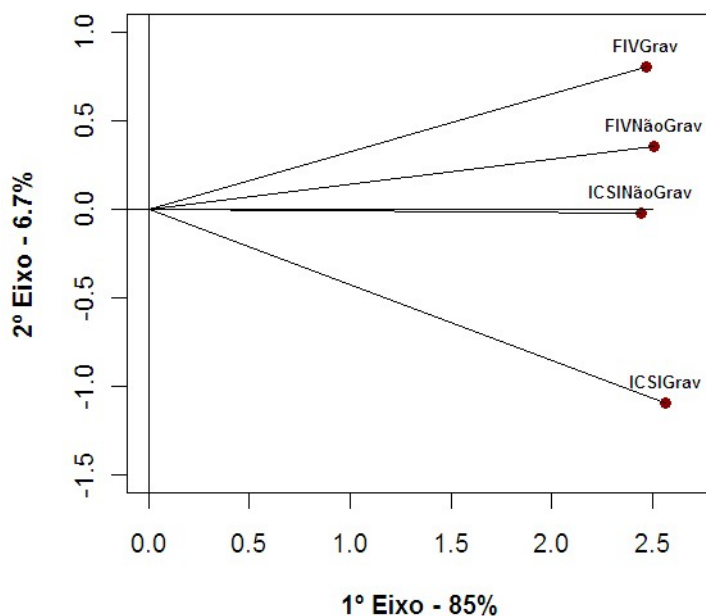


Figura 6.3: Interestrutura não centrada.

Intraestrutura

Uma vez evidenciadas as semelhanças e as diferenças entre as estruturas definidas pelos quatro quadros de dados, procederemos à construção de um novo quadro, o compromisso, e à representação das posições compromisso das variáveis na imagem euclidiana do compromisso.

Os coeficientes de ponderação β_t , definidos em (3.10), das matrizes de correlação iniciais utilizados para a definição da matriz compromisso são, como se pode observar na tabela 6.5, próximos uns dos outros. Esta observação permite-nos concluir que a contribuição de cada objeto para a formação do objeto compromisso é semelhante e portanto este representa um bom resumo da estrutura dos dados.

β_1	β_2	β_3	β_4
0,247	0,251	0,257	0,245

Tabela 6.5: Coeficientes de ponderação β'_k s.

Imagem euclidiana centrada da interestrutura

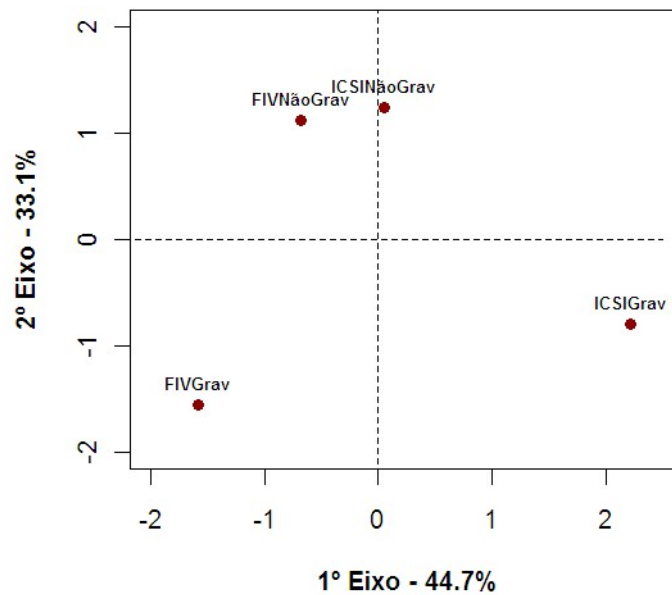


Figura 6.4: Interestrutura centrada.

A matriz compromisso, que se encontra em anexo (Anexo 3), fica então definida por:

$$V = 0,247V_1 + 0,251V_2 + 0,257V_3 + 0,245V_4.$$

Os objetos que mais se aproximam do compromisso, com distâncias muito semelhantes, correspondem aos quadros cujo resultado do ciclo foi a não gravidez, tabela 6.6.

	FIVGrav	FIVNaoGrav	ICSIGrav	ICSINaoGrav
Distância	2,25	1,89	2,35	1,86

Tabela 6.6: Distância dos objetos ao compromisso.

Os valores próprios da matriz compromisso assim como a percentagem de inércia explicada e a percentagem de inércia acumulada em cada eixo encontram-se na tabela 6.7.

Para se obter uma variabilidade total próxima dos 80% seria preciso reter, no mínimo, nove eixos. Segundo o critério de Cattell, figura 6.5, reter-se-iam cinco eixos.

Optou-se pela análise a cinco eixos, obtendo uma variabilidade de cerca de 53%, no entanto, como se verificará no decorrer da análise, a qualidade de representação das variáveis, principalmente a partir do terceiro eixo, é muito fraca.

As coordenadas das variáveis no objeto compromisso para os cinco primeiros eixos, e as respetivas contribuições encontram-se na tabela 6.8.

Eixo	Valor Próprio	Inércia Explicada	Inércia Acumulada
1	2,9907	17,59	17,59
2	1,7714	10,42	28,01
3	1,6962	9,98	37,99
4	1,3769	8,10	46,09
5	1,2119	7,13	53,22
6	1,2015	7,07	60,29
7	1,0924	6,43	66,71
8	0,9110	5,36	72,07
9	0,8702	5,12	77,19
10	0,8315	4,89	82,08
11	0,7723	4,54	86,62
12	0,6459	3,80	90,42
13	0,5583	3,28	93,71
14	0,4782	2,81	96,52
15	0,4331	2,55	99,07
16	0,0859	0,51	99,57
17	0,0725	0,43	100,00

Tabela 6.7: Inércia explicada por cada eixo.

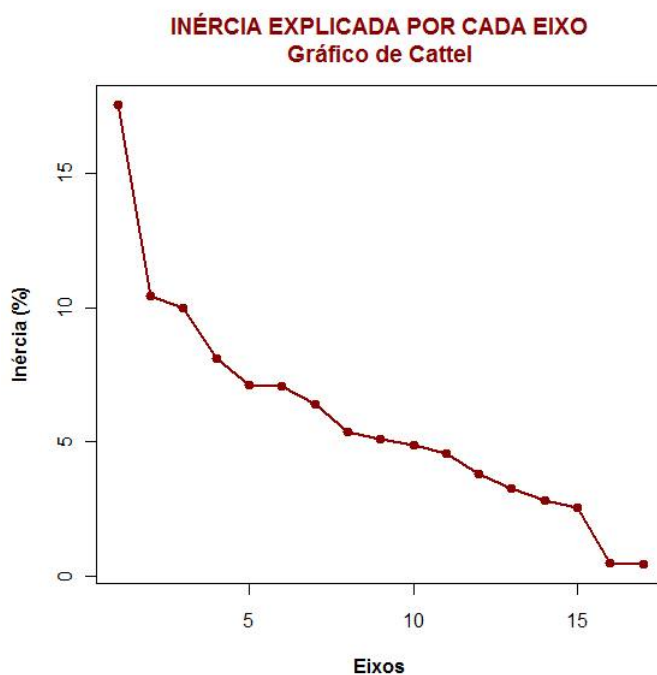


Figura 6.5: Gráfico de Cattell.

As variáveis que mais contribuem para a formação do primeiro eixo são "OvFec", "OvInsem", "EmbTBQ" e "DiaTransf", correspondendo a 68% das contribuições absolutas para a formação do mesmo.

	Coordenadas					Contribuições Absolutas					Contribuições Relativas				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
IdadeF	-0,24	0,56	0,07	0,06	-0,08	0,02	0,18	0,00	0,00	0,01	0,06	0,31	0,00	0,00	0,01
%EspermaB	-0,01	0,16	-0,02	0,02	0,22	0,00	0,01	0,00	0,00	0,04	0,00	0,03	0,00	0,00	0,05
OvInsem.	0,79	-0,06	-0,41	-0,07	-0,03	0,21	0,00	0,10	0,00	0,00	0,62	0,00	0,17	0,01	0,00
OvFec.	0,84	0,18	0,09	0,10	0,13	0,23	0,02	0,01	0,01	0,01	0,70	0,03	0,01	0,01	0,02
TxFec.	0,01	0,29	0,71	0,25	0,17	0,00	0,05	0,29	0,04	0,02	0,00	0,08	0,50	0,06	0,03
EmbTransf.	0,46	0,43	0,14	0,48	0,31	0,07	0,10	0,01	0,17	0,08	0,22	0,18	0,02	0,23	0,10
DiaTransf.	0,52	-0,15	-0,11	-0,08	-0,12	0,09	0,01	0,01	0,00	0,01	0,27	0,02	0,01	0,01	0,01
EmbTBQ	0,68	0,02	0,58	-0,09	-0,08	0,15	0,00	0,20	0,01	0,01	0,46	0,00	0,34	0,01	0,01
EmbTMQ	-0,39	0,28	-0,57	0,44	0,37	0,05	0,04	0,19	0,14	0,12	0,15	0,08	0,32	0,19	0,14
IMC	-0,18	0,13	0,06	-0,28	0,51	0,01	0,01	0,00	0,06	0,21	0,03	0,02	0,00	0,08	0,26
FSHBasal	-0,22	0,60	0,04	-0,09	-0,54	0,02	0,20	0,00	0,01	0,24	0,05	0,36	0,00	0,01	0,29
LHBasal	0,21	0,34	-0,28	0,40	-0,32	0,02	0,06	0,05	0,12	0,08	0,05	0,11	0,08	0,16	0,10
E2Basal	-0,12	-0,15	0,06	0,12	0,22	0,01	0,01	0,00	0,01	0,04	0,02	0,02	0,00	0,01	0,05
Gona.	-0,11	0,62	-0,06	-0,48	0,04	0,00	0,21	0,00	0,17	0,00	0,01	0,38	0,00	0,23	0,00
DiasEstim.	0,25	0,36	-0,23	-0,44	0,03	0,02	0,07	0,03	0,14	0,00	0,06	0,13	0,05	0,20	0,00
Estradiol	0,49	0,04	-0,41	0,09	-0,04	0,08	0,00	0,10	0,01	0,00	0,24	0,00	0,16	0,01	0,00
Endométrio	0,24	0,11	-0,11	-0,40	0,40	0,02	0,01	0,01	0,12	0,13	0,06	0,01	0,01	0,16	0,16

Tabela 6.8: Coordenadas das variáveis no compromisso.

No entanto, analisando a qualidade de representação destas no eixo verificamos que apenas as variáveis "OvInsem" e "OvFec" estão bem representadas, uma vez que apresentam contribuições relativas superiores a 0,50.

Atendendo à definição destas variáveis, poderemos inferir que o primeiro eixo representa os procedimentos manipulados em laboratório e/ou da responsabilidade dos embriologistas, e que são característicos das técnicas de fertilização em análise.

Salienta-se que todas as variáveis citadas se encontram na parte positiva do primeiro eixo. Embora se coloquem questões relativamente à qualidade de representação das variáveis "EmbTBQ" e "DiaTransf", optou-se por incluí-las uma vez que permitem reforçar o significado atribuído ao eixo.

As variáveis que melhor explicam o segundo eixo, isto é, as que estão mais correlacionadas, são "Gona", "FSHBasal" e "IdadeF"; no entanto, a qualidade de representação destas é considerada fraca (contribuições relativas de 0,38, 0,36 e 0,31, respetivamente). Todas estas variáveis encontram-se representadas na parte positiva do segundo eixo.

O valor da hormona FSH no início do ciclo fornece informação sobre a reserva ovárica da mulher. Com o avanço da idade feminina assiste-se a uma variação na FSHBasal. A mulher produz menos ovócitos sendo, por isso, necessário administrar uma dose mais elevada de gonadotrofinas para permitir que esta responda com uma produção de ovócitos que possa, de certa forma, ser comparada aos obtidos por uma mulher em idade jovem e com menos dose de gonadotrofinas.

O segundo eixo poderá assim explicar o tratamento de estimulação ovárica a aplicar à mulher.

A imagem euclidiana do compromisso no primeiro e segundo eixos encontra-se representada na figura 6.6.

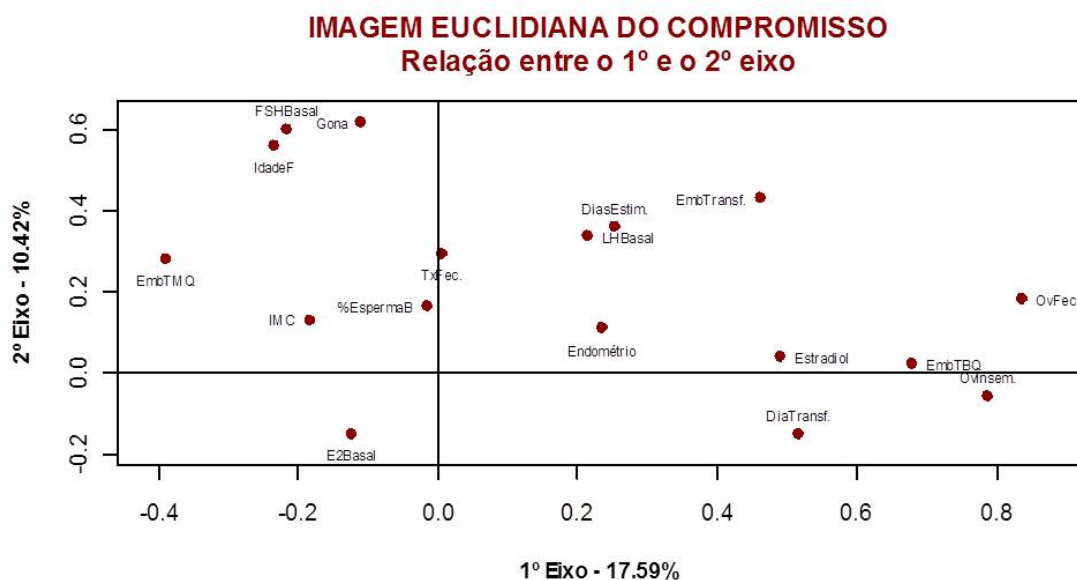


Figura 6.6: Imagem euclidiana do compromisso no 1º e 2º eixos.

No terceiro eixo e com uma contribuição absoluta que totaliza 88% temos as variáveis "TxFec" e "EmbTBQ", no lado positivo e as variáveis "EmbTMQ" e "OvInsem" no lado negativo do eixo.

Destas, apenas a variável "TxFec" se encontra relativamente bem representada no terceiro eixo. No entanto, considerando a representação no plano formado pelo primeiro e terceiro eixos, figura 6.7, distinguimos as variáveis "OvInsem" e "EmbTBQ" com uma boa qualidade de representação, 0,89 e 0,80, respetivamente.

Atendendo à natureza das variáveis poderemos relacionar o terceiro eixo com as variáveis que facilitam a obtenção de gravidez. Este opõe casais com taxa de fecundação e número de embriões de boa qualidade transferidos elevados, a casais com um número de embriões de média qualidade transferidos e número de ovócitos inseminados baixos.

As variáveis que melhor explicam o quarto eixo são "Gona", na parte positiva do eixo, e "EmbTransf", na parte negativa do eixo.

Quando é necessário administrar doses elevadas de gonadotrofinas é sinal de que a resposta ovárica é mais difícil existindo um número inferior de ovócitos disponíveis e, conseqüentemente, um número inferior de embriões para transferir.

Assim, este eixo poderá explicar a resposta ovárica da mulher ao tratamento.

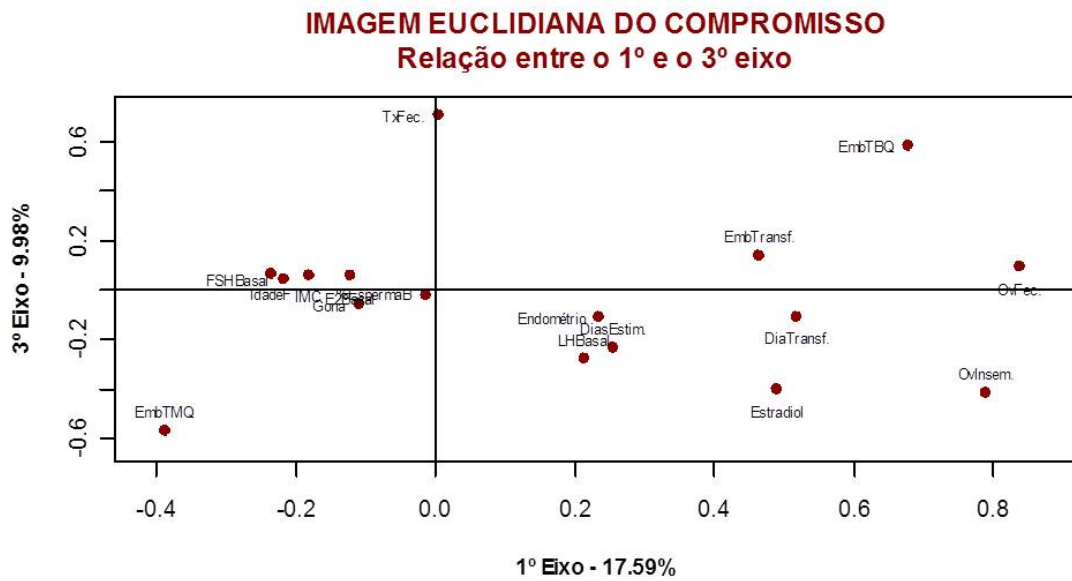


Figura 6.7: Imagem euclidiana do compromisso no 1º e 3º eixos.

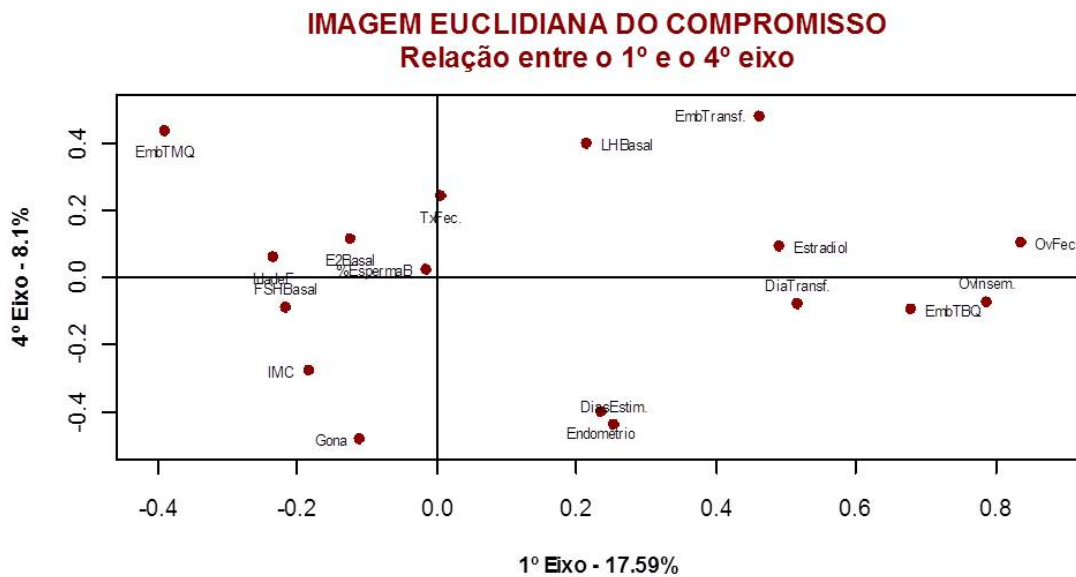


Figura 6.8: Imagem euclidiana do compromisso no 1º e 4º eixos.

Relativamente ao quinto eixo, cuja representação gráfica se encontra na figura 6.9, as variáveis que melhor o explicam são "IMC" do lado positivo do eixo e "FSHBasal" do lado negativo. Estas contribuem com 45% para a formação do quinto eixo, apresentando no entanto uma fraca qualidade de representação.

Sabe-se que um elevado IMC altera a parte hormonal influenciando, entre outros, os valores de FSHBasal. Nestes casos é sempre aconselhado à mulher uma perda de

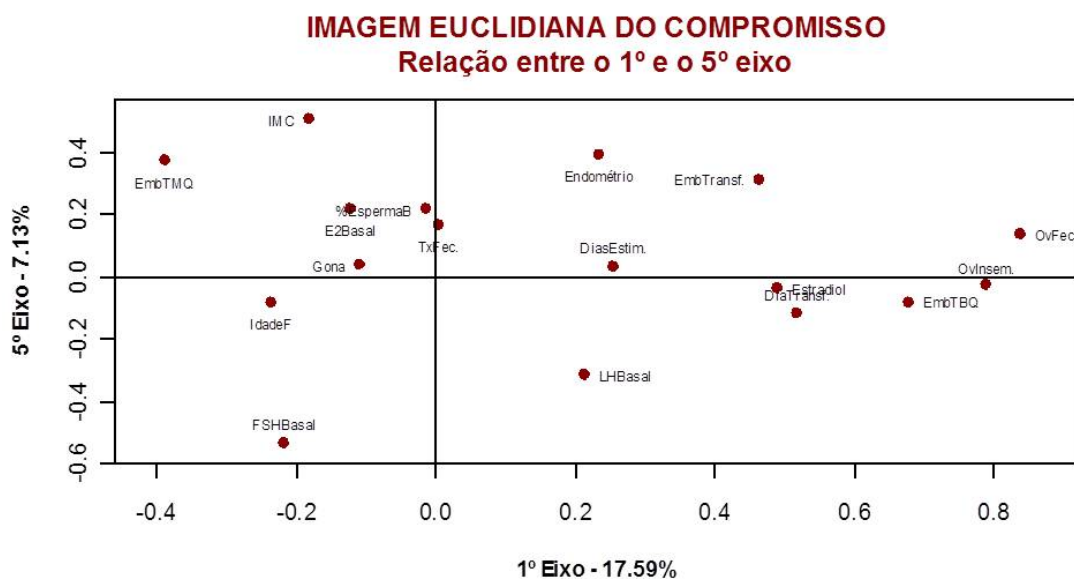


Figura 6.9: Imagem euclidiana do compromisso no 1º e 5º eixos.

peso que uma vez conseguida culmina, muitas vezes, com a obtenção de uma gravidez espontânea justificada por um melhor funcionamento do aparelho reprodutor. Este eixo poderá estar relacionado com mulheres com diagnóstico de síndrome de ovários poliquísticos (que atinge cerca de 20% da população feminina da zona mediterrânica, (Moureira *et al.* (2010)), no entanto muitos outros diagnósticos podem ser apontados como, por exemplo, diabetes.

O quinto eixo poderá explicar diagnósticos femininos relacionados com um IMC elevado e que produzem alteração nos valores da hormona FSH.

Trajatórias

A representação das trajetórias das variáveis é efetuada nos dois primeiros eixos do compromisso. Para cada variável são definidos quatro pontos correspondentes à posição dessa variável em cada um dos quatro estudos na imagem euclidiana do compromisso. Foi ainda considerado um quinto ponto, assinalado a vermelho, que representa o objeto compromisso. As coordenadas das variáveis em cada eixo, bem como as coordenadas do objeto compromisso encontram-se em anexo (Anexo 4).

A interpretação das trajetórias deve ser complementada com a decomposição da soma das distâncias ao quadrado entre pares de objetos, tabela 6.9, que nos indica quais as variáveis que mais contribuem para as diferenças globais entre os quadros de dados, e com a decomposição das distâncias entre pares de objetos, tabela 6.10, em contribuições de variáveis, que dá a indicação de quais as variáveis responsáveis pelas diferenças entre pares de quadros de dados.

Todas as distâncias se decompõem de forma a poderem ser interpretadas. Neste sentido, considerando o conjunto dos quadros de dados, a soma dos quadrados das distâncias entre cada par de quadros é uma medida de dispersão dos pontos na nuvem global. Esta quantidade decompõe-se segundo as variáveis dando estas contribuições uma explicação global da dispersão e dos principais responsáveis por elas.

Variáveis	$\sum_t \sum_{t'} d_{HS}^2(V_t, V_{t'})$	Variáveis	$\sum_t \sum_{t'} d_{HS}^2(V_t, V_{t'})$
Gona	10,49%	%EspermaB	5,53%
Endométrio	7,98%	IMC	5,5%
DiasEstim	6,94%	TxFec	5,25%
EmbTBQ	6,75%	E2Basal	4,69%
LHBasal	6,42%	EmbTransf	4,69%
DiaTransf	6,3%	OvInsem	4,63%
FSHBasal	6,13%	OvFec	4,44%
EmbTMQ	5,64%	Estradiol	3,06%
IdadeF	5,55%		

Tabela 6.9: Decomposição da soma das distâncias ao quadrado entre pares de objetos.

	FIVGrav			FIVNaoGrav		ICSI
	FIV	ICSI	ICSI	ICSI	ICSI	ICSI
	NaoGrav	Grav	NaoGrav	Grav	NaoGrav	NaoGrav
IdadeF	3,0	6,4	6,6	5,0	6,9	5,4
%EspermaB	6,9	5,0	8,1	4,6	5,7	3,3
OvInsem.	5,0	3,3	7,6	3,4	5,2	4,1
OvFec.	4,5	3,1	4,5	5,8	5,0	4,1
TxFec.	6,0	4,5	3,9	6,8	8,1	3,2
EmbTransf.	3,9	7,5	3,3	5,6	3,9	2,6
DiaTransf.	7,0	6,9	3,9	6,8	3,4	8,8
EmbTBQ	5,5	6,2	3,4	10,9	6,6	7,4
EmbTMQ	5,8	4,4	4,9	5,6	5,2	8,3
IMC	3,5	5,7	7,0	3,9	10,5	3,9
FSHBasal	9,4	6,8	4,0	4,3	7,2	5,8
LHBasal	6,1	3,7	4,6	9,6	5,3	9,4
E2Basal	4,4	4,6	6,5	3,8	6,0	3,2
Gona.	7,8	13,9	11,3	9,5	6,8	11,4
DiasEstim.	9,3	7,6	7,4	6,1	4,8	6,0
Estradiol	4,5	3,2	4,6	1,3	2,9	2,3
Endométrio	7,5	7,3	8,5	7,3	6,4	10,6

Tabela 6.10: Decomposição das distâncias entre pares de objetos em contribuições de variáveis para os afastamentos.

A decomposição das somas das distâncias ao quadrado entre pares de objetos permite distinguir as variáveis cujas correlações com as restantes são mais instáveis, contribuindo

desta forma para as diferenças entre os quadros de dados.

Neste aspeto, e conforme a tabela 6.9, destacam-se as variáveis gonadotrofinas (10,49%) e espessura do endométrio (7,98%), logo seguidas pelas variáveis número de dias de estimulação ovárica (6,94%), número de embriões transferidos de boa qualidade (6,75%) e hormonas LH e FSH medidas no início do ciclo (6,42% e 6,13%, respetivamente).

Analisando as trajetórias destas variáveis, figura 6.10, verificamos que se apresentam bastante irregulares e marcadas por uma elevada amplitude. Esta observação reflete uma alteração na estrutura das variáveis em questão nos diversos quadros.

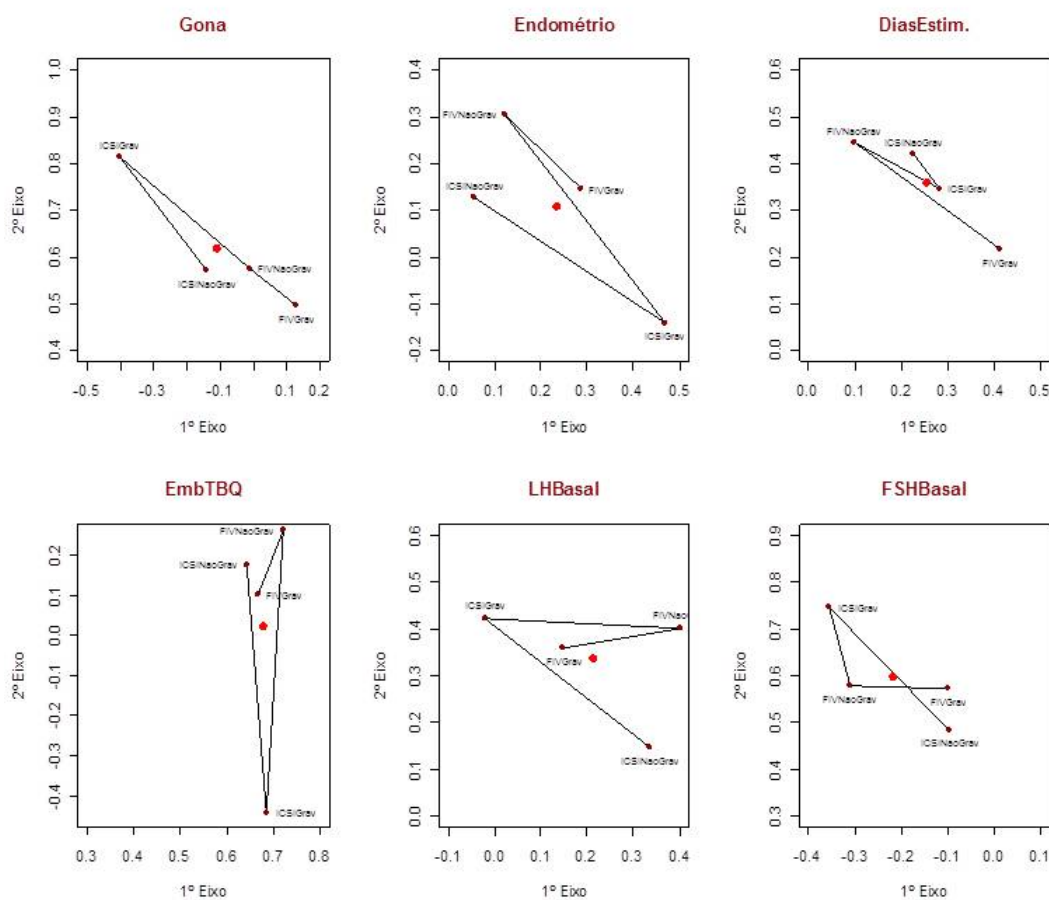


Figura 6.10: Trajetórias das variáveis "Gona", "Endométrio", "DiasEstim.", "EmbTBQ", "LHBasal", "FSHBasal".

A contribuição mais significativa da variável "Gona" dá-se entre os grupos em que se verificou gravidez com 13,9%, e a menos significativa entre os grupos em que não se verificou a gravidez (6,8%). Observando a trajetória desta variável verificamos que os quadros correspondentes a FIVNeoGrav e ICSINeoGrav são os que mais se aproximam do objeto compromisso.

A maior contribuição da variável "Endométrio", com 10,6%, dá-se entre os quadros correspondentes à técnica ICSI. As variáveis que mais contribuem para as diferenças entre

os dois grupos representados por esta técnica são "Gona"(11,4%), "Endométrio"(10,6%), "LHBasal"(9,4%), "DiaTransf"(8,8%) e "EmbTBQ"(7,4%). As variáveis "FSHBasal" e número de dias de estimulação ovárica ("DiasEstim") são as que mais contribuem para o afastamento entre os dois grupos representados pela técnica FIV, com 9,4% e 9,3% respetivamente. Se a estas juntarmos as variáveis "Gona"(7,8%), "Endométrio"(7,5%), "DiaTransf"(7,0%) e "%EspemaB"(6,9%) obtemos uma contribuição de aproximadamente 48% para as diferenças entre os grupos FIVGrav e FIVNaoGrav.

O estradiol, hormona que nivela o crescimento de folículos, medido após tratamento, é a variável que se apresenta mais estável sendo a que menos contribui, com 3,06%, para o afastamento entre os quadros de dados. De facto, se observarmos a sua trajetória (figura 6.11) verificamos que esta se apresenta mais fechada e em trono do objeto compromisso que qualquer outra.

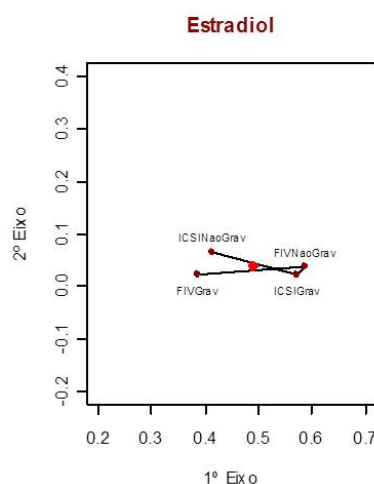


Figura 6.11: Trajetória da variável "Estradiol".

As variáveis número de ovócitos fecundados ("OvFec"), número de ovócitos inseminados ("OvInsem"), e estradiol no início do ciclo ("E2Basal") contribuem de forma semelhante para os afastamentos entre os quadros (4,44%, 4,63% e 4,69%, respetivamente). As trajetórias destas variáveis podem ser analisadas na figura 6.12 Estas três variáveis em conjunto com a variável "Estradiol", são as que menos contribuem para as diferenças entre os quadros.

Pode ainda definir-se um outro grupo de variáveis que contribuem de forma semelhante para os afastamentos: "TxFec"(5,25%), "IMC"(5,50%), "%EspermaB"(5,53%), "IdadeF"(5,55%) e "EmbTMQ"(5,64%), e cujas trajetórias se apresentam na figura 6.13

A trajetória da variável "TxFec" realça uma situação de maior instabilidade no quadro FIVNaoGrav. De facto se analisarmos a tabela 6.10 verificamos que as maiores distâncias relativas a esta variável estão associadas ao quadro FIVNaoGrav. Para a variável "%Es-

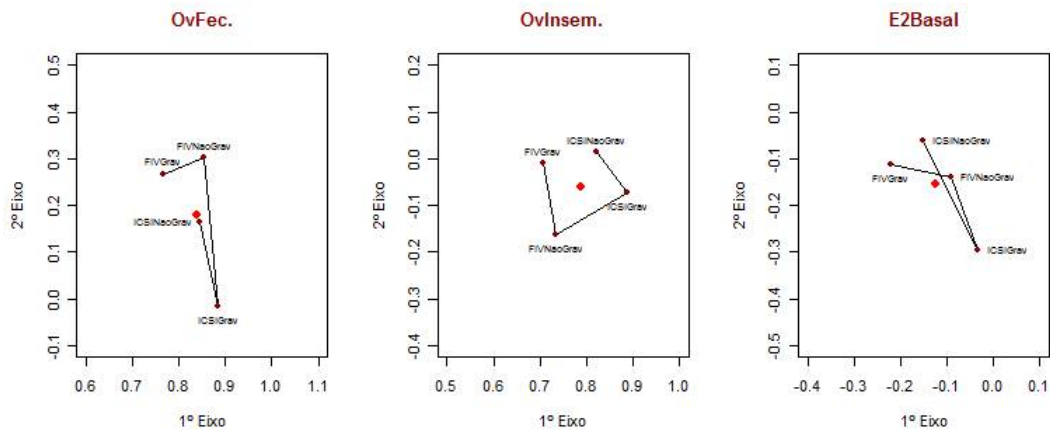


Figura 6.12: Trajetórias das variáveis "OvFec", "OvInsem" e "E2Basal".

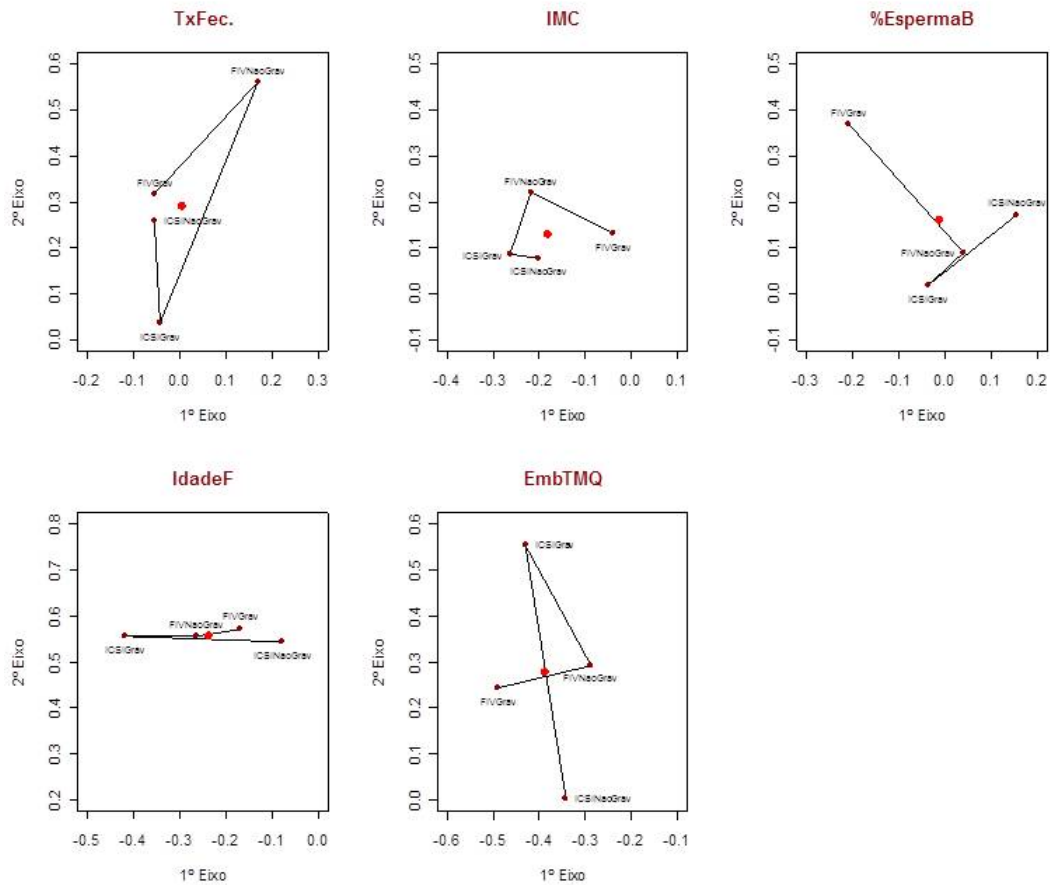


Figura 6.13: Trajetórias das variáveis "TxFec", "IM", "%EspermaB", "IdadeF" e "EmbTMQ".

permaB" a instabilidade verifica-se no quadro de FIVGrav.

Para a variável "OvInsem" verificamos uma situação de menor estabilidade entre FIV-Grav e ICSINaoGrav, cuja contribuição na decomposição $d_{HS}^2(V_{FIVGrav}, V_{ICSINaoGrav})$ é de 7,6%.

Foram elaborados vários estudos preliminares considerando as três bases de dados e a divisão destas nos mais variados grupos possíveis: segundo a "IdadeF", o "IMC", as "Causas de infertilidade", entre outras, associando variáveis como, por exemplo, "IdadeF" e "Grav", "Causas" e "Grav", "Técnica" e "Grav", não tendo sido encontrados resultados significativos ou diferentes dos apresentados nesta tese, em nenhum dos casos. A qualidade de representação das variáveis no espaço euclidiano do compromisso revelou-se fraca em todas as simulações realizadas, sendo que se optou por escolher apenas a análise que, no conjunto de todas, se mostrou ligeiramente mais interpretável e com algum interesse para os profissionais que trabalham na área de PMA.

7.1 Conclusão final

A análise efetuada revelou uma maior proximidade entre os grupos compostos por casais que não engravidaram e um maior distanciamento entre os grupos que engravidaram, quando consideramos casais submetidos a técnicas de fertilidade diferentes.

O objeto compromisso obtido representa um bom resumo da estrutura de dados, os objetos mais próximos deste foram os dois grupos em que não se verificou a gravidez, FIVNaoGrav e ICSINaoGrav, e aqueles que mais se distanciam são os grupos em que se observou a gravidez: FIVGrav e ICSIGrav.

A percentagem de inércia explicada por cada eixo não é elevada sendo necessário a análise a nove eixos para explicar, aproximadamente, 77% da variabilidade total.

A opção pela análise a apenas 5 eixos justifica-se com a fraca qualidade de representação das variáveis a partir do quinto eixo, tendo sido obtida explicação para apenas 53% da variabilidade total.

O primeiro eixo é explicado pelos procedimentos laboratoriais efetuados pelos embriologistas e que caracterizam as técnicas de fertilidade em análise.

O segundo eixo explica o tratamento de estimulação ovárica a aplicar à mulher.

O terceiro eixo opõe variáveis que estão diretamente relacionadas com a obtenção de gravidez.

O quarto eixo poderá explicar a resposta ovárica da mulher ao tratamento.

O quinto eixo poderá explicar diagnósticos femininos relacionados com um IMC elevado e que produzem alteração nos valores da hormona FSH.

As variáveis que mais contribuem para a diferença global entre os quadros de dados são: "Gona", "Endométrio", "DiasEstim", "EmbTBQ", "LHBasal" e "DiaTransf".

Considerando as diferenças entre pares de quadros verificamos que as variáveis que mais contribuem para as diferenças entre:

- FIVGrav e FIVNaoGrav são: "FSHBasal", "DiasEstim", "Gona", "Endométrio", "DiaTransf" e "%EspermaB";
- FIVGrav e ICSIGrav são: "Gona", "DiasEstim", "EmbTransf", "Endométrio", "DiaTransf", e "FSHBasal";
- FIVGrav e ICSINaoGrav são: "Gona", "Endométrio", "%EspermaB", "Ov.Insem", "DiasEstim" e "IMC";
- FIVNaoGrav e ICSIGrav são: "EmbTBQ", "LHBasal", "Gona", "Endométrio", "DiaTransf" e "TxFec".
- FIVNaoGrav e ICSINaoGrav são: "IMC", "TxFec", "FSHBasal", "IdadeF" e "Gona";
- ICSIGrav e ICSINaoGrav são: "Gona", "Endométrio", "LHBasal", "DiaTransf", "EmbTMQ" e "EmbTBQ".

Convém salientar que dada a complexidade da questão em análise não se poderia esperar que os dados devolvessem resultados inequívocos. Isto é, um casal que, à partida, se julgue reunir as condições para chegar com sucesso ao fim do ciclo poderá não obtê-lo e este verificar-se num casal em que as expectativas de sucesso por parte da equipa de PMA seriam praticamente nulas. Variáveis aleatórias como questões psicológicas e estilos de vida adotados contribuem, ainda que na maioria das vezes de uma forma subliminar, para o (in)sucesso de um ciclo de PMA.

No entanto, julga-se que o trabalho realizado e os resultados obtidos contribuem para destacar aspetos importantes relacionados com a infertilidade, sublinhando que se desconhecem trabalhos similares, isto é, que recorram a técnicas de análise multivariada, realizados nesta área.

7.2 Trabalho Futuro

Dada a relevância do tema em análise considera-se pertinente desenvolver outras análises e/ou manipulações nos dados que permitam acrescentar informações úteis à compreensão do problema.

As bases de dados analisadas poderão ser reorganizadas no sentido de permitir avaliar o mesmo casal a cada novo ciclo que inicia por insucesso do anterior. Desta forma poder-se-ia tirar partido da aplicação de ambas as metodologias STATIS e STATIS Dual, bem como de outras técnicas da análise conjunta de quadros de dados, como é o caso da Dupla Análise em Componentes Principais, ou da Análise Fatorial Múltipla.

Uma análise às alterações ocorridas no mesmo casal a cada ciclo, nomeadamente as alterações de um ciclo que resulte em insucesso para um ciclo que resulte em sucesso, permitirá evidenciar as variáveis que mais contribuíram para as diferenças.

Uma maior discriminação das causas de infertilidade diagnosticadas (no grande grupo das causas femininas ou masculinas existem diversos subgrupos) que permitisse a comparação entre casais que engravidaram e casais que não engravidaram, quando submetidos à mesma técnica de fertilidade, poderia igualmente revelar associações interessantes.

ANEXOS

Anexo 1

Tabela I: Descrição das variáveis.

Variáveis	Descrição
%EspermaB	Percentagem de espermatozoides com morfologia normal.
Altura	Altura, em m, da mulher.
AnosInfert	Tempo de infertilidade (anos). Nº de anos que o casal tenta a gravidez.
Causa	Causa diagnosticada para a infertilidade no casal.
DiasEstim	Número de dias de estimulação ovárica.
DiaTransf	Contagem de dias desde a punção folicular até ao dia da transferência (Punção folicular= dia 0).
E2Basal	Estradiol no início do ciclo.
EmbObt	Nº de embriões obtidos.
EmbTBQ	Nº de embriões, de boa qualidade, transferidos.
EmbTMaQ	Nº de embriões, de má qualidade, transferidos.
EmbTMQ	Nº de embriões, de qualidade média, transferidos.
EmbTransf	Nº de embriões transferidos.
Endométrio	Espessura, em mm, da mucosa que reveste a parede uterina.
Esperma[]	Concentração de esperma (10^6 /ml).
Estradiol	Hormona feminina que nivela o desenvolvimento dos folículos. Esta depende do número de folículos em crescimento. Esta variável regista o nível de estradiol após tratamento.
Fol>15mm	Número de folículos com mais de 15 mm na altura da punção folicular. São os que apresentam maior probabilidade de maturação.
FSHBasal	Hormona Folículo-Estimulante medida no início do ciclo. Na mulher, a FSH estimula o crescimento dos folículos ováricos.
Gona	Dose de Gonadotrofinas administrada. FSH e LH Hormonas de estimulação ovárica.
Grav	Resultado do ciclo: Obteve gravidez ou não obteve gravidez.
IdadeF	Idade feminina.
IdadeM	Idade masculina.
LHBasal	Hormona Luteinizante medida no início do ciclo. Na mulher, a LH é necessária para a produção de estrogénio.
OvFec	Nº de ovócitos que foram fecundados.
OvInsem	Nº de ovócitos que se submeteram a Técnica FIV ou ICSI.
OvObt	Nº de ovócitos obtidos após punção folicular.
Peso	Peso, em kg, da mulher.
SwimUp	Nº de espermatozóides progressivos rápidos obtidos após tratamento do esperma.
TxFec	Taxa de fecundação= $\frac{OvFec}{OvInsem} \times 100$.

Anexo 2

Tabela II: Matrizes de Correlação dos quadros de dados FIVGrav e FIVNaoGrav.

FIVGrav	IdadeF	%EspermaB	OvInsem	OvFec	TxFec	EmbTransf	DiaTransf	EmbTBQ	EmbTMQ	IMC	FSHBasal	LHBasal	E2Basal	Gona	DiasEstim	Estradiol	Endométrio
IdadeF	1,000																
%EspermaB	0,036	1,000															
OvInsem	-0,177	-0,166	1,000														
OvFec	0,031	-0,124	0,579	1,000													
TxFec	0,287	-0,009	-0,504	0,205	1,000												
EmbTransf	0,250	0,081	0,131	0,487	0,328	1,000											
DiaTransf	0,135	0,006	0,362	0,517	0,099	0,146	1,000										
EmbTBQ	-0,073	-0,172	0,190	0,437	0,206	0,542	0,146	1,000									
EmbTMQ	0,245	0,257	-0,143	-0,210	-0,037	-0,009	-0,081	-0,846	1,000								
IMC	0,234	-0,041	0,098	0,105	-0,022	0,075	0,101	-0,182	0,264	1,000							
FSHBasal	0,274	0,199	-0,224	-0,008	0,186	0,072	0,181	-0,015	0,064	-0,239	1,000						
LHBasal	-0,035	0,030	0,167	0,006	-0,123	0,202	0,106	-0,120	0,271	-0,052	0,155	1,000					
E2Basal	-0,120	-0,007	-0,181	-0,014	0,281	0,047	-0,208	-0,100	0,149	-0,039	-0,097	-0,061	1,000				
Gona	-0,010	0,561	0,202	0,142	-0,080	0,128	0,080	0,142	-0,087	0,004	0,299	0,004	0,046	1,000			
DiasEstim	0,051	-0,269	0,296	0,184	-0,065	0,294	0,076	0,257	-0,119	0,048	-0,224	0,322	-0,187	-0,121	1,000		
Estradiol	-0,221	-0,141	0,278	0,143	-0,163	0,069	0,173	0,098	-0,072	-0,040	0,160	0,119	-0,209	0,024	0,025	1,000	
Endométrio	-0,041	0,141	0,392	0,190	-0,189	-0,210	0,214	0,010	-0,145	0,233	0,002	-0,002	-0,276	0,335	0,203	0,129	1,000

FIVNaoGrav	IdadeF	%EspermaB	OvInsem	OvFec	TxFec	EmbTransf	DiaTransf	EmbTBQ	EmbTMQ	IMC	FSHBasal	LHBasal	E2Basal	Gona	DiasEstim	Estradiol	Endométrio
IdadeF	1,000																
%EspermaB	-0,156	1,000															
OvInsem	-0,347	0,009	1,000														
OvFec	-0,111	0,136	0,577	1,000													
TxFec	0,300	-0,035	-0,358	0,441	1,000												
EmbTransf	0,113	0,134	0,220	0,605	0,368	1,000											
DiaTransf	-0,162	-0,059	0,389	0,356	0,010	0,019	1,000										
EmbTBQ	0,026	-0,055	0,235	0,623	0,459	0,630	0,181	1,000									
EmbTMQ	0,187	0,051	-0,006	-0,082	-0,101	0,204	-0,179	-0,440	1,000								
IMC	0,224	-0,230	-0,125	0,071	0,166	-0,031	-0,070	-0,187	0,277	1,000							
FSHBasal	0,196	-0,018	-0,326	-0,177	0,240	-0,150	-0,278	-0,052	-0,012	0,096	1,000						
LHBasal	0,034	0,234	0,128	0,328	0,178	0,281	0,188	0,277	-0,120	-0,043	0,179	1,000					
E2Basal	-0,051	-0,135	-0,131	-0,173	-0,136	0,078	-0,048	0,049	-0,073	-0,162	-0,045	-0,022	1,000				
Gona	0,157	-0,028	0,038	0,052	0,037	0,139	-0,299	0,090	0,065	-0,003	0,149	-0,015	0,113	1,000			
DiasEstim	-0,078	-0,122	0,119	0,039	0,049	-0,010	-0,216	-0,019	0,068	-0,078	0,349	0,111	0,008	0,265	1,000		
Estradiol	-0,033	-0,177	0,487	0,352	-0,137	0,230	0,185	0,267	-0,039	-0,152	-0,246	0,266	0,077	0,092	0,106	1,000	
Endométrio	0,063	0,004	-0,108	0,078	0,196	0,119	-0,074	0,008	0,055	0,007	0,012	0,104	-0,199	0,108	0,274	0,123	1,000

Tabela III: Matrizes de Correlação dos quadros de dados ICSIGrav e ICSINaoGrav.

ICSIGrav	IdadeF	%EspermaB	OvInsem	OvFec	TxFec	EmbTransf	DiaTransf	EmbTBQ	EmbTMQ	IMC	FSHBasal	LHBasal	E2Basal	Gona	DiasEstim	Estradiol	Endométrio
IdadeF	1,000																
%EspermaB	-0,086	1,000															
OvInsem	-0,237	-0,049	1,000														
OvFec	-0,248	-0,158	0,827	1,000													
TxFec	-0,053	-0,164	-0,294	0,228	1,000												
EmbTransf	-0,211	0,104	0,271	0,257	0,009	1,000											
DiaTransf	-0,196	0,193	0,441	0,424	-0,102	-0,177	1,000										
EmbTBQ	-0,350	0,039	0,369	0,384	0,067	0,185	0,197	1,000									
EmbTMQ	0,244	0,057	-0,132	-0,207	-0,186	0,267	-0,219	-0,849	1,000								
IMC	0,045	-0,045	-0,229	-0,144	0,172	-0,115	-0,074	0,006	-0,080	1,000							
FSHBasal	0,247	-0,037	-0,144	-0,153	-0,081	-0,059	-0,322	-0,244	0,231	0,039	1,000						
LHBasal	0,103	-0,168	0,039	-0,060	-0,112	0,311	-0,403	-0,210	0,302	-0,267	0,259	1,000					
E2Basal	-0,099	0,248	0,023	-0,082	-0,102	-0,140	0,036	0,051	-0,110	0,073	-0,187	-0,075	1,000				
Gona	0,449	0,079	-0,135	-0,140	-0,052	-0,129	-0,074	-0,463	0,411	0,222	0,638	0,045	-0,127	1,000			
DiasEstim	0,083	0,129	0,363	0,274	-0,290	-0,077	0,360	-0,052	0,078	-0,009	0,172	-0,049	-0,063	0,345	1,000		
Estradiol	0,041	-0,128	0,428	0,392	-0,118	0,226	0,248	0,130	0,039	-0,171	-0,218	0,180	-0,032	-0,145	0,167	1,000	
Endométrio	-0,234	-0,051	0,299	0,332	-0,113	0,144	0,103	0,248	-0,123	-0,221	-0,142	0,073	0,133	-0,242	0,005	0,333	1,000

ICSINaoGrav	IdadeF	%EspermaB	OvInsem	OvFec	TxFec	EmbTransf	DiaTransf	EmbTBQ	EmbTMQ	IMC	FSHBasal	LHBasal	E2Basal	Gona	DiasEstim	Estradiol	Endométrio
IdadeF	1,000																
%EspermaB	0,048	1,000															
OvInsem	-0,026	0,133	1,000														
OvFec	0,030	0,173	0,792	1,000													
TxFec	0,075	0,064	-0,333	0,219	1,000												
EmbTransf	0,075	0,130	0,281	0,361	0,113	1,000											
DiaTransf	-0,142	-0,005	0,458	0,410	-0,135	-0,031	1,000										
EmbTBQ	0,008	0,203	0,329	0,437	0,058	0,445	0,155	1,000									
EmbTMQ	-0,021	-0,042	-0,158	-0,180	-0,027	0,222	-0,129	-0,638	1,000								
IMC	-0,209	0,008	-0,236	-0,243	-0,142	-0,047	-0,112	0,047	-0,022	1,000							
FSHBasal	0,241	-0,007	-0,004	-0,061	0,000	0,003	0,109	0,012	-0,140	-0,040	1,000						
LHBasal	0,007	-0,053	0,294	0,238	-0,135	0,124	0,242	0,142	-0,073	-0,154	0,233	1,000					
E2Basal	0,055	0,120	-0,219	-0,196	-0,047	-0,067	-0,207	-0,061	0,059	-0,037	-0,128	0,171	1,000				
Gona	0,349	0,009	-0,192	-0,149	0,071	-0,075	-0,138	0,089	-0,241	0,335	0,198	-0,238	-0,040	1,000			
DiasEstim	-0,020	0,125	0,106	0,121	0,026	0,113	-0,076	0,144	-0,157	0,137	0,028	0,007	0,077	0,444	1,000		
Estradiol	0,180	-0,042	0,408	0,301	-0,236	0,127	0,092	0,012	0,080	-0,151	-0,158	0,194	0,135	-0,032	0,151	1,000	
Endométrio	-0,026	0,175	-0,123	-0,015	0,058	0,036	-0,136	0,106	-0,031	0,223	-0,253	-0,110	0,277	0,231	0,317	-0,014	1,000

Anexo 3

Tabela IV: Matriz Compromisso.

	IdadeF	%EspermaB	Ovinsem	OvFec	TxFec	EmbTransf	DiaTransf	EmbTBQ	EmbTMQ	IMC	FSHBasal	LHBasal	EZBasal	Gona	DiasEstim	Estradiol	Endométrio
IdadeF	1,000																
%EspermaB	-0,041	1,000															
Ovinsem	-0,198	-0,019	1,000														
OvFec	-0,077	0,005	0,694	1,000													
TxFec	0,151	-0,037	-0,372	0,274	1,000												
EmbTransf	0,054	0,112	0,226	0,426	0,203	1,000											
DiaTransf	-0,092	0,035	0,413	0,427	-0,032	-0,013	1,000										
EmbTBQ	-0,100	0,003	0,282	0,470	0,198	0,448	0,170	1,000									
EmbTMQ	0,165	0,080	-0,110	-0,170	-0,089	0,172	-0,153	-0,694	1,000								
IMC	0,074	-0,078	-0,124	-0,053	0,045	-0,030	-0,039	-0,079	0,109	1,000							
FSHBasal	0,239	0,034	-0,175	-0,101	0,085	-0,034	-0,081	-0,076	0,038	-0,035	1,000						
LHBasal	0,028	0,010	0,155	0,127	-0,048	0,231	0,029	0,021	0,096	-0,130	0,207	1,000					
EZBasal	-0,055	0,057	-0,125	-0,116	-0,002	-0,021	-0,105	-0,015	0,005	-0,040	-0,115	0,002	1,000				
Gona	0,238	0,154	-0,022	-0,024	-0,006	0,015	-0,108	-0,039	0,042	0,139	0,324	-0,049	-0,003	1,000			
DiasEstim	0,009	-0,033	0,222	0,155	-0,072	0,078	0,038	0,081	-0,031	0,023	0,083	0,097	-0,041	0,234	1,000		
Estradiol	-0,008	-0,122	0,401	0,298	-0,163	0,164	0,176	0,127	0,002	-0,129	-0,117	0,190	-0,007	-0,016	0,113	1,000	
Endométrio	-0,061	0,066	0,116	0,148	-0,013	0,024	0,027	0,094	-0,061	0,057	-0,095	0,017	-0,016	0,104	0,198	0,145	1,000

Anexo 4

Tabela V: Coordenadas das trajetórias das variáveis no 1º eixo.

	Coordenadas			
	FIVGrav	FIVNaoGrav	ICSIGrav	ICSINaoGrav
IdadeF	-0,1692	-0,2652	-0,4189	-0,0791
%EspermaB	-0,2099	0,0391	-0,0380	0,1546
OvInsem	0,7086	0,7340	0,8892	0,8229
OvFec	0,7669	0,8534	0,8841	0,8443
TxFec	-0,0535	0,1703	-0,0435	-0,0539
EmbTransf	0,4740	0,5641	0,3772	0,4413
DiaTransf	0,5148	0,5129	0,5479	0,4950
EmbTBQ	0,6668	0,7213	0,6864	0,6421
EmbTMQ	-0,4922	-0,2895	-0,4294	-0,3432
IMC	-0,0395	-0,2198	-0,2628	-0,2028
FSHBasal	-0,0993	-0,3115	-0,3551	-0,0961
LHBasal	0,1482	0,4006	-0,0211	0,3361
E2Basal	-0,2223	-0,0902	-0,0328	-0,1507
Gona	0,1285	-0,0123	-0,4039	-0,1399
DiasEstim	0,4129	0,0981	0,2807	0,2260
Estradiol	0,3867	0,5862	0,5728	0,4115
Endométrio	0,2865	0,1216	0,4691	0,0552

Tabela VI: Coordenadas das trajetórias das variáveis no 2º eixo.

	Coordenadas			
	FIVGrav	FIVNaoGrav	ICSIGrav	ICSINaoGrav
IdadeF	0,5733	0,5569	0,5556	0,5451
%EspermaB	0,3686	0,0892	0,0194	0,1729
OvInsem	-0,0105	-0,1636	-0,0733	0,0141
OvFec	0,2672	0,3019	-0,0162	0,1656
TxFec	0,3161	0,5599	0,0351	0,2593
EmbTransf	0,5751	0,4939	0,2588	0,3944
DiaTransf	0,2031	-0,3478	-0,3297	-0,1318
EmbTBQ	0,1022	0,2634	-0,4430	0,1760
EmbTMQ	0,2437	0,2941	0,5564	0,0034
IMC	0,1320	0,2195	0,0849	0,0780
FSHBasal	0,5726	0,5774	0,7460	0,4837
LHBasal	0,3590	0,4032	0,4238	0,1460
E2Basal	-0,1121	-0,1386	-0,2960	-0,0601
Gona	0,4966	0,5756	0,8154	0,5718
DiasEstim	0,2159	0,4439	0,3462	0,4216
Estradiol	0,0217	0,0385	0,0234	0,0665
Endométrio	0,1451	0,3053	-0,1410	0,1275

Tabela VII: Coordenadas do compromisso no 1º e 2º eixos.

	Coordenadas	
	Eixo 1	Eixo 2
IdadeF	-0,2354	0,5577
%EspermaB	-0,0140	0,1608
OvInsem	0,7894	-0,0590
OvFec	0,8377	0,1782
TxFec	0,0051	0,2912
EmbTransf	0,4638	0,4292
DiaTransf	0,5180	-0,1542
EmbTBQ	0,6795	0,0207
EmbTMQ	-0,3887	0,2779
IMC	-0,1821	0,1286
FSHBasal	-0,2176	0,5966
LHBasal	0,2141	0,3346
E2Basal	-0,1229	-0,1533
Gona	-0,1094	0,6168
DiasEstim	0,2541	0,3570
Estradiol	0,4907	0,0373
Endométrio	0,2354	0,1075

BIBLIOGRAFIA

- [APF] APF. Associação Portuguesa de Fertilidade (<http://www.apfertilidade.org>).
- [2] Assembleia da República (2006). Lei n.º32/2006 de 26 de julho. *Diário da República - N.º143/06, 1.ª série - Procriação Medicamente Assistida*.
- [3] Brannian, J. D. (2011). Obesity and Fertility. *South Dakota Medicine: The Journal Of The South Dakota State Medical Association*, 64(7):251 – 254.
- [4] Cailliez, F. and Pàges, J. (1976). *Introduction à l'analyse des données*. Smash.
- [5] Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2):245–276.
- [6] Chatfield, C. and Collins, A. (1980). *Introduction to multivariate analysis*. Chapman and Hall, London, repr. edition.
- [7] CISIA. Centre International de Statistique et d'informatique Appliquées (1997). *SPAD.TM - Analyse des Tableaux Multiples, Manuel de Reference*.
- [CNPMA] CNPMA. Centro Nacional de Procriação Medicamente Assistida (<http://www.cnpma.org.pt>).
- [9] Dazy, F. and Le Barzic, J. (1996). *L'analyse des données évolutives: méthodes et applications*. Editions Technip.
- [10] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39(11):27–34.
- [11] Figueiredo, A., Figueiredo, F., Monteiro, N., and Straume, O. (2012). Restructuring in privatised firms: A Statis aproach. *Structural Change and Economic Dynamics*, 23(1):108–116.

- [12] Fournier, M., Motelay-Massei, A., Massei, N., Aubert, M., Bakalowicz, M., and Dupont, J. (2009). Investigation of transport processes inside ,Karst aquifer by means of STATIS. *Ground Water*, 47(3):391–400.
- [13] Gourvéneç, S., Stanimirova, I., Saby, C.-A., Airiau, C. Y., and Massart, D. L. (2005). Monitoring batch processes with the STATIS approach. *Journal of Chemometrics*, 19(5-7):288–300.
- [14] Gudmundsson, L., Tallaksen, L. M., and Stahl, K. (2011). Spatial cross-correlation patterns of European low, mean and high flows. *Hydrological Processes*, 25(7):1034–1045.
- [15] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441.
- [16] Instituto Nacional de Estatística (2012). *Estatísticas no feminino: Ser Mulher em Portugal 2001-2011*. INE, Lisboa.
- [17] Jolliffe, I. (1986). *Principal Component Analysis*. Springer Verlag.
- [18] Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200.
- [19] Lavit, C. (1988a). *Analyse conjointe de tableaux quantitatifs*. Méthode + programmes. Masson.
- [20] Lavit, C. (1988b). Présentation de la méthode STATIS permettant l’analyse conjointe de plusieurs tableaux de données quantitatives. *Cachiers de la Recherche Développement*, 18:49–60.
- [21] Lavit, C., Escoufier, Y., Sabatier, R., and Traissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis*, 18:97–119.
- [22] L’Hermier des Plantes, H. (1976). *Structuration des Tableaux à Trois Indices de la Statistique*. Thèse de 3^{eme} cycle. Université de Montpellier II.
- [23] L’Hermier des Plantes, H. and Thiébaud, B. (1977). Étude de la pluviosité au moyen de la méthode S.T.A.T.I.S. *Revue de Statistique Appliquée*, 25(2):57–81.
- [24] Luke, B. (2011). The Role of Obesity in Reproduction. In Racowsky, C., Schlegel, P. N., Fauser, B. C., and Carrell, D. T., editors, *Biennial Review of Infertility*, pages 35–42. Springer US.

- [25] Ministério da Saúde (2011). Despacho n.º3219/2011 de 17 de fevereiro. *Diário da República - N.º34/11, 2.ª série*.
- [26] Moreira, S., Soares, E., Tomaz, G., Maranhão, T., and Azevedo, G. (2010). Síndrome dos ovários poliquísticos - enfoque psicossocial. *Acta Médica Portuguesa*, 23(2):237–242.
- [27] Niang, N., Fogliatto, F., and Saporta, G. (2009). Contrôle multivarié de procédés par lots à l'aide de statis. In *41èmes Journées de Statistique, SFdS, Bordeaux, Bordeaux, France*.
- [28] Oliveira, M. M. and Mexia, J. T. (2007). Modelling series of studies with a common structure. *Computational Statistics and Data Analysis*, 51(12):5876–5885.
- [29] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572.
- [30] Pestana, M. and Gageiro, J. (2003). *Análise de Dados para Ciências Sociais-SPSS: A complementaridade do SPSS*. 3ª Edição. Sílabo.
- [31] Robert, P.; Escoufier, Y. (1976). A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient. *Applied Statistics*, 25:257–265.
- [32] Santos, T. and Ramos, M. (2010). *Esterilidade e Procriação Medicamente Assistida*. Estado da arte. Universidade de Coimbra.
- [33] Scepi, G. (2002). *Parametric and non parametric multivariate quality control charts.*, pages 163–189. Heidelberg: Physica-Verlag.
- [34] Stanimirova, I., Walczak, B., Massart, D., Simeonov, V., Saby, C., and Crescenzo, E. D. (2004). STATIS, a three-way method for data analysis. application to environmental data. *Chemometrics and Intelligent Laboratory Systems*, 73(2):219 – 233.
- [35] Thioulouse, J. (2011). Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *The Annals of Applied Statistics*, 5(4):2300–2325.
- [WHO] WHO. World Health Organization (<http://www.who.int>).
- [37] WHO (2010). Mother or nothing: the agony of infertility. *Bulletin World Health Organization*, 88(12):881–2.

- [38] Zegers-Hochschild, F., Adamson, G. D., de Mouzon, J., Ishihara, O., Mansour, R., Nygren, K., Sullivan, E., and Vanderpoel, S. (2009). International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology. *Fertility and Sterility*.