

Teresa Isabel Cardoso Valada
Gestão de Resíduos



Teresa Isabel Cardoso Valada

Gestão de Resíduos

Mestrado em Química

Departamento de Química

FCTUC

Julho 2014



UNIVERSIDADE DE COIMBRA

• U  C • 2014

Teresa Isabel Cardoso Valada

GESTÃO DE RESÍDUOS

**Dissertação apresentada para provas de Mestrado em Química, Área de especialização em
Controle de Qualidade e Ambiente**

Professor Doutor Jorge Luís Gabriel Ferreira da Silva Costa Pereira

Doutora Helena Margarida Mesquita de Carvalho

Julho 2014

Universidade de Coimbra



 **bluepharma**

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota”.

Madre Teresa de Calcutá

Agradecimentos

Este trabalho não poderia ter sido desenvolvido sem a ajuda e apoio de alguns intervenientes a quem quero agradecer.

Ao Conselho de Administração de empresa Bluepharma, ao Presidente, Doutor Paulo Barradas e à Vice-presidente, Doutora M^a Isolina Mesquita, pela autorização concedida para poder efectuar o meu Estágio Curricular.

À Doutora Helena Margarida de Carvalho por toda a ajuda, simpatia, profissionalismo e cooperação que manifestou durante a minha permanência na empresa.

Ao meu orientador de estágio do Departamento de Química, Doutor Jorge Costa Pereira, pela fundamental e imprescindível orientação, compreensão, colaboração e disponibilidade ao longo de toda a elaboração da dissertação de mestrado.

Ao Eng. Ricardo Alves e à Dr^a. Matilde Ribeiro por toda a disponibilidade, ajuda e confiança que depositaram em mim e no meu trabalho.

Aos responsáveis de secção dos diferentes departamentos da empresa pela disponibilidade, ajuda e colaboração no decurso do meu projeto.

A todos os colaboradores da empresa Bluepharma, em especial aos colaboradores das pesagens, do armazém, da Gestão da Qualidade, à D.Gina, D. Isabel, à Susana, à Kátia, à Lia, à D. Ana, à Marília, ao Hugo, à Filipa e à Joana pela ajuda e disponibilidade durante a minha permanência na empresa.

À Doutora Ana Cristina Ribeiro por tudo o que aprendi e pelo apoio que deu em todas as decisões que tomei.

Ao Dr. Luís Veríssimo, ao Doutor Francisco Ferreira e à Dr^a. Filomena Pinto pela disponibilidade, apoio e força que sempre transmitiram.

À Dr^a Aline pela amizade demonstrada ao longo destes anos.

A todos os meus amigos do Departamento de Química, em especial à Sara Pereira, à Filipa Correia, à Ana Freitas, à Joana Brito, à Joana Pragana, à Joana Virgínio, à Filipa Pires, à Ângela Varela e ao Nelson Fonseca por todas as palavras de incentivo.

Aos meus amigos, Cláudio Valada, João Amado, Sofia Neves, Jacqueline Janeiro, Rita Oliveira, Mariana Janeiro, André Simões, Pedro Guerra e Elisabete Silva pela paciência, apoio e ânimo ao longo da realização deste trabalho.

Ao meu padrinho e à minha madrinha pela força e coragem que me deram no decorrer destes anos.

Por ultimo, mas não sendo menos importante, aos meus pais e avó pela força, coragem, paciência e ensinamento, mas sobretudo por serem os primeiros e os últimos a aplaudir. Dedico-vos esta e as restantes vitórias.

A ti Coimbra, eterna cidade dos estudantes, por tudo o que me ensinaste e por todos os momentos inesquecíveis que me proporcionaste.

A todos, bem-haja!

Objectivos

O tema proposto pela indústria farmacêutica Bluepharma - Gestão de Resíduos - para o desenvolvimento deste projecto de mestrado teve como objectivo principal identificar as causas mais relevantes na produção de resíduos da empresa, a fim de se definirem estratégias para a sua redução.

Neste sentido pretende-se, por um lado, evitar os danos ambientais associados ao tratamento dos resíduos, por outro lado, minorar os desperdícios obtidos ao longo do processo produtivo, uma vez que mantendo as perdas sob controlo, é possível expandir o rendimento económico e financeiro da empresa.

Com os métodos ao dispor, será feita uma análise aprofundada de toda a informação recolhida, para determinar quais as etapas do processo e os agentes responsáveis pela maior quantidade de resíduos produzidos ao longo do processo de fabrico e embalamento.

Será ainda dado especial relevo aos detalhes do processo, que podem de alguma forma influenciar o resultado final - volume de resíduos gerado.

Resumo

O trabalho desenvolvido incidiu na aplicação de ferramentas da qualidade e de métodos quimiométricos no tratamento da informação multivariada, referente aos detalhes dos processos de fabrico e embalagem, assim como dos respectivos resíduos gerados.

A informação foi obtida no decorrer do processo de produção e de embalagem dos diversos produtos, no período de Janeiro a Abril de 2014. Esta informação foi processada e interpretada da seguinte forma: inicialmente, e para se ter uma perspectiva gráfica, construiu-se um diagrama circular e um diagrama de Pareto, tendo como objectivo concluir quais os resíduos que constituíam maior preocupação para a empresa, quer em termos de massa, ou seja, em quantidade produzida, quer em termos de custos associados. Tendo esta informação disponível, procedeu-se, então, à análise multivariada que se revela um instrumento rigoroso e forte para analisar a informação relativa a inúmeros produtos que passam por diferentes etapas no processo produtivo.

Foram, então, usados métodos quimiométricos para avaliar o impacto que os detalhes do processo teriam sobre a quantidade de resíduos gerada. É necessário ter em conta que cada etapa do processo foi analisada em separado, restringindo-se o trabalho aos detalhes e aos resíduos nela contida.

Os resultados revelam que através do método por Mínimos Quadrados Parciais (PLS) e pelo método de Mínimos Quadrados Multivariado (MOLS) é possível determinar qual ou quais os detalhes que geram maior quantidade de resíduos.

Para finalizar, conclui-se que os métodos utilizados identificaram os grupos de resíduos mais problemáticos e permitiram retirar conclusões úteis e interessantes.

Palavras-chave:

Ambiente, Indústria Farmacêutica, Resíduos Industriais, Gestão de Resíduos, Diagrama de Pareto, Mínimos Quadrados Parciais, Mínimos Quadrados Multivariado e Redução de Resíduos.

Abstract

The work focused on the application of quality tools and chemometric methods in the treatment of multivariate information regarding the details of the manufacturing and packaging processes, as well as the produced waste.

The information was obtained during the production and packaging process of various products, in the period from January to April 2014. This information was processed and interpreted as follows: initially, and to have a graphical perspective, we constructed a circular diagram and a Pareto diagram, aiming to conclude the residues which constitute major concern for the company, both in terms of mass, ie, quantity produced and in terms of costs. Having this information available, we proceeded, then to multivariate analysis that reveals a rigorous and powerful tool to analyze the information concerning numerous products that go through different stages in the production process.

Were then used chemometric methods to assess the impact that the details of the case would have on the amount of waste generated. It is necessary to take into account that each step of the process was analyzed separately, restricting the work to the details and waste contained in it.

The results show that using the method of partial least squares (PLS) and the method of multiple ordinary least squares (MOLS) it is possible to detail which process steps generate more waste.

Finally, it is concluded that methods used identified the most problematic groups of waste and got to interesting and useful conclusions.

Keywords:

Environmental, pharmaceutical industry, industrial waste, waste management, Pareto chart, partial least squares, multiple ordinary least squares and waste reduction.

Preâmbulo

Esta dissertação encontra-se organizada em 6 capítulos, que são de seguida descritos.

O capítulo 1 é constituído pelo enquadramento do trabalho.

O capítulo 2 diz respeito à fundamentação do trabalho, ou seja, aborda a caracterização e os tipos dos resíduos, a política ambiental, a gestão de resíduos, assim como as formas de prevenção, as técnicas de tratamento de resíduos perigosos e o tratamento de dados que serve de base para o capítulo 4.

O capítulo 3 refere-se à parte experimental, isto é, descreve-se a metodologia adoptada para alcançar os objectivos propostos.

O capítulo 4 corresponde à análise e discussão dos resultados obtidos, referentes aos resíduos gerados no processo produtivo da indústria.

No capítulo 5 apresenta-se uma síntese conclusiva dos resultados obtidos e sugestões para futuro.

No capítulo 6 encontram-se listadas as referências bibliográficas que serviram de suporte teórico e metodológico ao presente trabalho de investigação.

Nomenclatura

- # - iteração
- $\%J_y$ - justificação da resposta em y
- $\%R_x$ - informação residual relativa do sub-espço x
- $\%R_y$ - informação residual relativa do sub-espço y
- $\%RSE$ - erro padrão relativo (sob a forma de valor percentual)
- $\%U_x$ - quantidade relativa de informação de x
- α - nível de significância ou de diferença significativa (traduz a probabilidade de erro tipo I - rejeição abusiva)
- β - potência do teste (traduz a probabilidade de erro tipo II - aceitação abusiva)
- $\Delta\nu$ - variação no número de graus de liberdade introduzida por uma alteração voluntária (não aleatória)
- ΔSS - variação na soma de quadrados introduzida por uma alteração voluntária (não aleatória) -
- η_i - eficiência do factor latente
- ν - número de graus de liberdade (número de valores estatisticamente independentes)
- ρ_{xy} - correlação entre x e y
- $\sigma(\beta)$ - erro padrão do parâmetro
- σ_{pe} - estimativa da contribuição puramente aleatória (dispersão)
- σ_{xy}^2 - covariância entre as variáveis x e y
- σ_x^2 - variância de x
- σ_y^2 - variância de y
- $\sum(\%J_y)$ - cumulativo da informação justificada em y
- $\sum(\%U_x)$ - cumulativo da informação utilizada em x
- a_3 - estimador de simetria ($a_3 = \sum(z_i)^3/N$)

- a_4 - estimador de curtose ($a_3 = \sum(z_i)^4/N$)
- Corr* - correlação
- Corr*[x, y] - correlação entre x e y
- CoV*[x, y] - covariância entre as variáveis x e y
- H_0 - hipótese nula, aquela que se assume a priori como correcta. No formalismo estatístico convencional a H_0 contempla sempre a situação de conformidade e igualdade
- H_1 - hipótese alternativa, aquela que complementa a hipótese nula no universo do evento. No formalismo estatístico convencional a H_1 contempla a desigualdade que se pretende verificar
- $p[H_0]$ - valor de prova; traduz a probabilidade de aceitação da hipótese nula (H_0)
- p - número de parâmetros do modelo
- SS* - soma de quadrados; do Inglês, Sum of Squares
- SS_x - soma de quadrados da variável x ($SS_x = (x_i - \bar{x})^2$); do Inglês, Sum of Squares
- SS_y - soma de quadrados da variável y ($SS_y = (y_i - \bar{y})^2$); do Inglês, Sum of Squares
- TV* - valor de teste calculado (do Inglês, Test Value)
- $V[x]$ - variância de x
- x_i - variável independente
- y_i - variável dependente
- ANR - Autoridade Nacional de Resíduos
- APA - Agência Portuguesa do Ambiente
- CQ - Controlo de Qualidade
- e-GAR - Guia de Acompanhamento de Resíduos electrónica
- EMAS - Sistema Comunitário de Eco-Gestão e Auditoria (do Inglês, Environmental Management and Audit Scheme)

- ETARI - Estações de Tratamento de Águas Residuais Industriais
- FL - factor latente (estrutura latente em PLS)
- GMP - Boas Práticas de Fabrico (do Inglês, Good Manufacturing Practices)
- IDAG - Investigação e Desenvolvimento Analítico e Galénico
- ISO - International Organization for Standardization
- ISO 14001 - Norma ISO relativa à gestão ambiental
- ISO 9001 - Norma ISO sobre a Gestão da Qualidade
- LER - Lista Europeia de Resíduos
- MOLS - Mínimos Quadrados Multivariado (do Inglês, Multiple Ordinary Least Squares)
- Nm. - nome do modelo
- OHSAS - Segurança e Saúde Ocupacional (do Inglês, Occupational Health and Safety Advisory Services)
- OLS - Método dos Mínimos Quadrados não Ponderados (do inglês, Ordinary Least Squares)
- ONU - Organização das Nações Unidas
- PDCA - Ciclo de Deming: Plan (planear) - estabelecer objectivos e metodologias com vista à satisfação do cliente; Do (executar) - implementar as metodologias aplicadas anteriormente; Check (verificar) - monitorizar e medir processos ou produtos e relatar os dados; Act (actuar) - exercer acções de melhoria contínua
- PLS - Mínimos Quadrados Parciais (do Inglês, Partial Least Squares ou ainda Projection over Latent Structure)
- PVC - polímero plástico feito à base de cloreto de polivinil (do Inglês, polyvinyl chloride)
- PVDC - Polímero de plástico feito à base de cloreto de polivinilideno (do Inglês, Polyvinylidene chloride)
- RH - Resíduos Hospitalares
- RI - Resíduos Industriais
- RIP - Resíduos Industriais Perigosos
- RSU - Resíduos Sólidos Urbanos
- SGA - Sistema de Gestão Ambiental
- SGI - Sistema de Gestão Integrado
- SGQ - Sistema de Gestão da Qualidade

SIRER - Sistema Integrado de Registo Electrónico de Resíduos

SST - Segurança e Saúde no Trabalho

UE - União Europeia

WLS Método dos Mínimos Quadrados Ponderados (do inglês, Weighted Least Squares)

GAR - Guias de Acompanhamento de Resíduos

PDCA - Ciclo de Deming (do Inglês, Plan, Do, Check and Act)

TQM - Gestão Total da Qualidade

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Contextualização	2
2	Fundamentação	4
2.1	Resíduos	4
2.1.1	Gestão de Resíduos	5
2.1.2	Resíduos perigosos	6
2.1.2.1	Classificação de resíduos perigosos	6
2.1.3	Operadores licenciados de gestão de resíduos industriais	8
2.1.3.1	Preenchimento de guias de acompanhamento de resíduos	8
2.1.4	Técnicas para tratamento de resíduos perigosos	10
2.1.4.1	Tratamento físico-químico	11
2.1.4.2	Unidades de estabilização	12
2.1.4.3	Unidades de tratamento de resíduos orgânicos	12
2.1.4.4	Incineração e Co-Incineração	13
2.1.4.5	Aterros	14
2.1.5	Resíduos em Indústrias Farmacêuticas	16
2.2	Sistema de Gestão Ambiental	16
2.2.1	Resíduos produzidos na Bluepharma	18
2.2.1.1	Principais resíduos produzidos na Bluepharma	18
2.2.1.2	Resíduos farmacêuticos	18
2.2.1.3	Resíduos líquidos e sólidos provenientes dos laboratórios	19
2.2.1.4	Resíduos de processo de embalagem não contaminados	20
2.2.1.5	Resíduos de material de embalagem impresso	20
2.2.1.6	Óleos e absorventes com óleos	21
2.2.1.7	Areias com carvão activado	21
2.2.1.8	Resíduos para reciclar	21
2.2.1.9	Materiais eléctricos e electrónicos	21
2.2.1.10	Resíduos pontuais	21
2.2.1.11	Agulhas	22
2.2.1.12	Resíduos de meios de dissolução	22
2.2.1.13	Separação e encaminhamento de resíduos	22
2.2.2	Sistema de Gestão da Qualidade	23
2.2.2.1	ISO 9000	23

2.2.2.2	ISO 9001:2008	24
2.2.2.3	ISO 9004	25
2.2.3	Sistema Comunitário de Ecogestão e Auditoria	25
2.2.3.1	ISO 14001 e EMAS	26
2.3	Sistema de Gestão de Segurança e Saúde no Trabalho	26
2.3.1	OHSAS 18001:2007	27
2.3.2	OHSAS 18002:2008	27
2.4	Tratamento de dados	28
2.4.1	Ferramentas da Qualidade	29
2.4.1.1	Análise de Pareto	29
2.4.2	Testes estatísticos	30
2.4.2.1	Nível de confiança e nível de significância	30
2.4.2.2	Procedimento	31
2.4.2.3	Testes baseados nas distribuições estatísticas	33
2.4.2.4	Outros testes	35
2.4.3	Análise de interacção	36
2.4.4	Modelação Implícita	38
2.4.5	Modelação Explícita	46
2.4.5.1	Pressupostos	46
2.4.5.2	Tipos de estimativas	46
3	Parte experimental	50
3.1	Descrição das operações de produção e seus resíduos	50
3.1.1	Pesagem da matéria-prima	50
3.1.2	Processo de granulação	51
3.1.2.1	Granulação húmida	51
3.1.2.2	Granulação a seco	52
3.1.2.3	Compressão directa	52
3.1.3	Mistura	53
3.1.4	Compressão	54
3.1.4.1	Compressora rotativa	54
3.1.4.2	Compressora excêntrica	54
3.1.5	Revestimento	55
3.2	Tratamento de dados	56
3.3	Critérios	57
4	Resultados e discussão	59
4.1	Análise preliminar	59
4.1.1	Impacto material	59
4.1.2	Impacto financeiro	61
4.2	Análise detalhada	63
4.2.1	Resposta y13	64
4.2.1.1	Análise de correlação	64
4.2.1.2	Abordagem PLS	65
4.2.1.3	Modelação explícita	68
4.2.2	Resposta y7	74
4.2.2.1	Análise de correlação	74
4.2.2.2	Abordagem PLS	75
4.2.2.3	Modelação explícita	77
4.2.3	Resposta y1	83

4.2.3.1	Análise de correlação	83
4.2.3.2	Abordagem PLS	84
4.2.3.3	Modelação explícita	86
4.2.4	Resposta y12	91
4.2.4.1	Análise de correlação	91
4.2.4.2	Abordagem PLS	92
4.2.4.3	Modelação explícita	95
4.2.5	Resposta y9	100
4.2.5.1	Análise de correlação	100
4.2.5.2	Abordagem PLS	101
4.2.5.3	Modelação explícita	104
4.2.6	Resposta y4	107
4.2.6.1	Análise de correlação	108
4.2.6.2	Abordagem PLS	108
4.2.6.3	Modelação explícita	111
4.2.7	Resposta y5	116
4.2.7.1	Análise de correlação	116
4.2.7.2	Abordagem PLS	116
4.2.7.3	Modelação explícita	118
5	Conclusões	122
A	Tabelas	129

Capítulo 1

Introdução

Neste capítulo vamos abordar algumas questões relacionadas com a produção de resíduos provenientes das actividades humanas e industriais, o impacto ambiental causado por acidentes industriais e as suas consequências para a saúde pública.

1.1 Motivação

A gestão de resíduos é um assunto preocupante para o meio ambiente e para a saúde pública e, por isso, merece o devido destaque.

Nas últimas décadas, o tratamento dado aos diferentes resíduos tem evoluído de forma satisfatória, atendendo a que um dos objectivos é minimizar a poluição ambiental. Recentemente e devido à crescente industrialização e competitividade, as questões ambientais passaram a ser mais do que um aspecto relacionado com a qualidade e o conceito de geração de resíduos adquiriu uma nova definição, em que as questões administrativas se agregam à diversidade das questões ambientais. Tornou-se, então, essencial o desenvolvimento de estratégias e técnicas de protecção ambiental, tendo em conta as constantes alterações da legislação e normas cada vez mais rígidas e ao mesmo tempo a localização, as condições económicas e operacionais da indústria, a fim de a tornar mais competitiva no mercado.

Actualmente existe uma consciência crescente para a importância de minimizar o impacto dos resíduos industriais, tendo como finalidade criar processos mais eficientes, trazendo melhores resultados económicos e financeiros, para competir e sobreviver no mercado.

Os resíduos podem ser transformados em bens rentáveis se forem tratados de forma adequada.

1.2 Contextualização

Apesar de a indústria ser uma actividade com largos anos de existência, apenas há sensivelmente pouco tempo surgiu a preocupação com a gestão dos seus resíduos. A actividade industrial origina anualmente uma enorme quantidade de resíduos, com as mais variadas propriedades e características [1].

A crescente competitividade das empresas no mercado tem levado ao desenvolvimento de técnicas de análise que visam a melhoria de processos internos, diminuindo as falhas e perdas do processo produtivo e, conseqüentemente, reduzindo os custos de produção. [2]. Essas análises são feitas utilizando as Ferramentas da Qualidade que constituíram um passo importante para a evolução do processo produtivo nas empresas.

No entanto, durante muitos anos essa evolução não teve em conta as questões relacionadas com o meio ambiente [3]. Os resíduos e emissões eram apenas lançados no ambiente, sem qualquer preocupação com os possíveis impactos ambientais e as conseqüências para a saúde humana daí decorrentes. As águas e o ar eram vistos como locais de deposição dos resíduos [4] e foram vários os acidentes ocorridos na indústria química ao longo dos anos. Nesse sentido, a Organização das Nações Unidas (ONU) realizou a primeira Conferência Mundial sobre o Meio Ambiente Humano, em Estocolmo, em 1972 [5], com o intuito de orientar a população mundial e os dirigentes das nações na prevenção deste tipo de catástrofes e sensibilizar para as problemáticas ambientais. Ainda assim, alguns dos mais graves acidentes industriais ocorreram nos anos subsequentes: Bhopal, na Índia, em 1984, que envolveu libertação tóxica, San Juanico, Cidade do México, no mesmo ano, com a explosão em terra e Piper Alpha, em 1988, numa plataforma petrolífera, em pleno Mar do Norte [6].

Mais tarde, as entidades reguladoras introduziram vários normativos que utilizam o conceito dos 3R, Reduzir, Reutilizar e Reciclar, para regulamentar o uso de substâncias químicas, para reduzir as emissões de gases, as descargas de efluentes líquidos e os resíduos sólidos. Deste modo, foram implementadas as normas International Organization for Standardization (ISO) e as Legislações Ambientais, que vão sofrendo alterações sempre que necessário [7], para que as empresas e instituições se tornem conscientes do seu impacto ambiental, ao mesmo tempo que adquirem responsabilidades nessa área.

É portanto necessário um constante ajustamento no funcionamento das empresas de forma a que se adaptem a novas condições e, conseqüentemente, se tornem mais competitivas.

Neste sentido, é cada vez mais importante as empresas possuírem o certificado do Sistema de Gestão da Qualidade (SGQ), do Sistema de Gestão Ambiental (SGA) e da Segurança e Saúde Ocupacional (OHSAS), à medida que os clientes começam a exigir padrões de qualidade cada vez mais elevados. Estas certificações são compatíveis, pois todas as empresas têm como objectivo melhorar o seu desempenho, reduzir o impacto ambiental, bem como reduzir o consumo de energia, a emissão de materiais e resíduos, tendo assim uma oportunidade de se diferenciarem e tornarem mais competitivas [8].

Actualmente existe uma maior preocupação nas diversas áreas de actividade industrial no que diz respeito à quantidade de resíduos gerados numa empresa. Se, por um lado, a quantidade de resíduos produzidos não pára de aumentar e a empresa tem um impacto profundo na sua contabilidade, por outro lado, os resíduos podem ter implicações no meio ambiente e, conseqüentemente, na

saúde pública [9].

Capítulo 2

Fundamentação

Neste capítulo vamos abordar toda a complexidade dos resíduos: a sua formação, as medidas viáveis tomadas na gestão de resíduos, as condições adequadas de armazenamento dentro das empresas, as técnicas de tratamento e a implementação de normas internas para minimizar os resíduos gerados, tendo como objectivo reunir elementos que contribuam para a análise posterior do nosso tema.

2.1 Resíduos

A crescente urbanização e industrialização das sociedades modernas tem levado a um aumento significativo de resíduos. Assim, a necessidade de minimizar a produção de resíduos e de assegurar a sua gestão sustentável transformou-se num dever de cidadania. Segundo o artigo 3^o do Decreto-Lei 178/2006 de 5 de Setembro, os resíduos são definidos como “*qualquer substância ou objecto de que o detentor se desfaz ou tem a intenção ou a obrigação de se desfazer, (...)*” e podem ser classificados de acordo com a sua origem e características físico-químicas. De um modo geral, os resíduos são classificados como urbanos (resíduos sólidos produzidos pelas actividades residenciais), industriais (gerados pelas indústrias), hospitalares (são resíduos sem valor e são considerados perigosos no hospital, pois podem conter agentes causadores de doenças) e agrícolas (provenientes das explorações agrícolas). Qualquer que seja a classificação que se considere, há resíduos inertes, resíduos não perigosos e resíduos perigosos para o Homem e restantes seres vivos. Estes últimos designam-se por resíduos perigosos em função do seu carácter nocivo, tóxico, radioactivo e são essencialmente produzidos na indústria, mas também na saúde e na agricultura. Trata-se de resíduos para os quais as indústrias e os agricultores não conseguem assumir a inteira responsabilidade de os eliminar, com o mínimo de riscos para a saúde pública e para o ambiente.

Devido à sua perigosidade, estes devem ter um tratamento adequado e a sua gestão deve ser o mais correcta possível, ou seja, diferentes categorias de resíduos não devem ser misturadas entre si, a não ser que depois de conhecidas as características, a origem e a quantidade destes, o procedimento de eliminação ou valorização possa ser realizado de forma mais segura. Actualmente, o país possui um conjunto de operadores licenciados para a gestão de resíduos e

um conjunto de entidades gestoras do fluxo de resíduos, tendo como finalidade encaminhar os diferentes tipos de resíduos para os devidos locais, de modo a maximizar a reciclagem e a valorização. São da responsabilidade dos operadores os danos ambientais causados pelo exercício da sua actividade, traduzindo uma maior consciencialização e consequentemente um aumento do nível de prevenção e precaução nas suas instalações [10]. Este princípio tornou-se mais concreto quando a União Europeia (UE), em 1996, implementou uma estratégia de prevenção de resíduos, que estabelece uma hierarquia preferencial das operações de gestão de resíduos: 1) prevenção; 2) reutilização; 3) reciclagem; 4) outros tipos de valorização; 5) eliminação. A prevenção é a primeira das opções da estratégia de gestão de resíduos. Segue-se a reutilização e a reciclagem, com a igual finalidade de reduzir a produção de resíduos. Quando as opções de prevenção, reutilização e reciclagem não são viáveis, escolhem-se outros tipos de valorização. Esgotadas todas as soluções anteriores, segue-se então a eliminação, em que o resíduo segue para um destino final adequado. [11]

2.1.1 Gestão de Resíduos

Nos últimos tempos, a gestão de resíduos tem sido um tema importante no que se refere à política ambiental. Em 1985 tentou-se definir o conceito da política nacional de gestão de resíduos. Porém, só em 1990 é que se assistiu aos primeiros passos para a implementação de uma política ambiental integrada, estando os resultados ainda aquém do esperado. Em 1996, foram elaborados planos estratégicos para os resíduos sólidos urbanos (RSU), os resíduos hospitalares (RH), os resíduos industriais (RI) e os resíduos agrícolas, estando actualmente em constante alteração [12]. De acordo com o artigo 2º do Decreto-Lei nº 178/2006 de 5 de Setembro, a gestão de resíduos é descrita como: *“toda e qualquer operação de recolha, transporte, armazenagem, triagem, tratamento, valorização e eliminação de resíduos, bem como às operações de descontaminação de solos e à monitorização dos locais de deposição após o encerramento das respectivas instalações”*. É necessário que estas actividades sejam executadas de forma ambientalmente correcta, constituindo o menor risco possível para a saúde ambiental e pública, assim como, recorram às melhores técnicas disponíveis que não impliquem custos excessivos. No mesmo seguimento, a gestão de resíduos deve ser efectuada por agentes autorizados, assegurando que o transporte e o tratamento de resíduos sejam realizados em local apropriado, sendo proibido o abandono de resíduos, a incineração de resíduos no mar, a queima a céu aberto, bem como a descarga em locais não licenciados para a realização de tratamento de resíduos [10].

A elaboração de um plano de gestão de resíduos dentro de uma empresa tem como finalidade definir metas e prazos para que estas se organizem de forma a alcançar melhorias contínuas neste sector. O desafio neste sentido é adoptar estratégias preventivas a fim de reduzir ou eliminar os resíduos gerados. Uma produção mais limpa tem como objectivo aumentar a eficiência no uso de matérias-primas, água, energia, assim como os benefícios ambientais. É do interesse das empresas procederem a algumas regras básicas na gestão dos seus resíduos, tais como:

1. Seleccionar os resíduos, fazendo a sua separação de acordo com a sua origem e características, para posteriormente serem submetidos ao trata-

mento mais adequado,

2. Armazenar temporariamente os resíduos que não possuem risco para os operadores, nem para o meio ambiente,
3. Entregar cada tipo de resíduos ao operador licenciado, assegurando que o seu transporte e o seu devido tratamento são realizados,
4. Emitir ou solicitar a emissão de guias de transporte para cada tipo de resíduos e organizá-las num dossier,
5. Efectuar o registo no Sistema Integrado de Registo Electrónico de Resíduos (SIRER) da Agência Portuguesa do Ambiente (APA) , 30 dias úteis após a data de início da respectiva actividade,
6. Preencher anualmente o mapa de resíduos, até ao dia 31 de Março do ano seguinte a que diga respeito.

A caracterização e quantificação de resíduos são ferramentas essenciais na gestão de resíduos, visto que são indispensáveis para o planeamento económico e eficiente da gestão dos sistemas de recolha, transporte, armazenamento, triagem, tratamento, valorização e eliminação. A quantificação tem a finalidade de estimar a quantidade, em massa, dos resíduos gerados no processo produtivo, enquanto que a caracterização consiste em analisar o tipo de resíduos gerados (vidro, plástico, papel. . .) [11].

Em suma, são estratégias de gestão que tornam as empresas mais competitivas, pois ao reduzirem a produção de resíduos, reduzem também os seus impactos ambientais e os custos associados ao seu tratamento [10].

2.1.2 Resíduos perigosos

As controvérsias geradas em torno do destino e do tratamento adequado a dar aos Resíduos Industriais Perigosos (RIP) e a implementação de estratégias para os reduzir ou evitar representam relações problemáticas entre ciência/sociedade e tecnologia/democracia [4].

Os resíduos industriais em processos produtivos são uma forma de poluição que suscita uma crescente preocupação, tanto ao nível dos diversos impactos ambientais, como relativamente ao custo associado ao tratamento destes resíduos. O armazenamento destes resíduos não é a melhor opção e a sua destruição também não é a mais satisfatória, pois implica a formação de emissões, resultando na produção de resíduos concentrados e poluentes. A melhor opção é reduzir, prevenir ou até mesmo evitar a produção de resíduos ao longo do processo produtivo e escolher o melhor destino final a dar-lhes. Assim, é necessária a implementação de estratégias de prevenção, com vista a melhorar os processos de produção e eliminação de resíduos, assim como cumprir os planos nacionais de prevenção de resíduos, publicados nos diferentes diplomas legais. [11].

2.1.2.1 Classificação de resíduos perigosos

A Lista Europeia de Resíduos (LER) classifica os resíduos de acordo com a sua proveniência e com a actividade industrial que os origina (tabela 2.1).

Ao longo do tempo, esta lista tem sofrido alterações devido ao aumento das exigências no encaminhamento e tratamento dos resíduos, que levam os

produtores a evitar os produtos ou processos que os originam e assim se evitem impactos ambientais e encargos acrescidos com o seu tratamento [13]. Para tornar tudo mais claro, recorreu-se a uma listagem dos locais para onde este tipo de resíduos deve seguir, de acordo com a qual as operações de gestão de resíduos são discriminadas. Estas operações dividem-se em operações de valorização de resíduos (código R) e operações de eliminação de resíduos (código D). As operações de valorização são numeradas de R1 a R13 e as de eliminação são numeradas de D1 a D15 [14]. Esta lista encontra-se estruturada em 20 capítulos, os quais são constituídos por 2 dígitos, que por sua vez estão subdivididos em sub-capítulos que contêm 4 dígitos. Os códigos LER têm 6 dígitos e são estes que devem ser usados. Os resíduos perigosos são identificados com o símbolo (*), a seguir aos dígitos do código LER [11].

Tabela 2.1: Resíduos de processos químicos orgânicos - excerto do capítulo 7 da Lista Europeia de Resíduos [15].

7	Resíduos de Processos Químicos Orgânicos
(...)	(...)
07 05	Resíduos do FFDU de Produtos Farmacêuticos
07 05 01 (*)	Líquidos de lavagem e lícores mãe aquosos
07 05 03 (*)	Solventes, líquidos de lavagem e lícores mãe orgânicos halogenados
07 05 04 (*)	Outros solventes, líquidos de lavagem e lícores mãe orgânicos
07 05 07 (*)	Resíduos de destilação e resíduos de reação halogenados
07 05 08 (*)	Outros resíduos de destilação e resíduos de reação
07 05 09 (*)	Absorventes usados e bolos de filtração halogenados
07 05 10 (*)	Outros absorventes usados e bolos de filtração
07 05 11(*)	Lamas do tratamento local de efluentes contendo substâncias perigosas
07 05 12	Lamas de tratamento local de efluentes não abrangidos em 07 05 11
07 05 13 (*)	Resíduos sólidos contendo substâncias perigosas
07 05 14	Resíduos sólidos não abrangidos em 07 05 13
07 05 99	Outros resíduos não anteriormente especificados

2.1.3 Operadores licenciados de gestão de resíduos industriais

Todos os resíduos, depois de devidamente separados, são encaminhados através de operadores licenciados, para serem submetidos aos processos de tratamento adequado: valorização, reciclagem, incineração e/ou aterros [11]. O armazenamento dos resíduos deverá ser efectuado de modo a não provocar danos para o ambiente ou para a saúde humana e de forma a evitar a possibilidade de derrame, incêndio ou explosão, devendo ser respeitadas as condições de segurança relativas às características que conferem perigosidade aos resíduos. Além disso, os resíduos devem estar armazenados temporariamente, devem ter em consideração a sua classificação de acordo com o código LER, as suas características físico-químicas e as que conferem perigosidade. O armazenamento temporário de resíduos previamente à sua valorização ou à sua eliminação deverá ser efectuado dentro do tempo limite estipulado [14].

2.1.3.1 Preenchimento de guias de acompanhamento de resíduos

O procedimento de gestão de resíduos implementado assegura o cumprimento dos requisitos legais aplicáveis, nomeadamente através do preenchimento das Guias de Acompanhamento de Resíduos (GAR), por tipologia de resíduos e suas características (figura 2.1). Concretizando, as GAR são preenchidas no campo do produtor e transportador, ficando o original na empresa; o triplicado é enviado posteriormente ao destinatário [11].

No seguimento de tornar mais fiável e simplificado o procedimento de registo e controlo relativo do sistema de acompanhamento de transporte de resíduos, introduziu-se a Guia de Acompanhamento de Resíduos eletrónica (e-GAR). O transporte de resíduos estará assim sujeito a registo electrónico a efectuar pelos produtores, detentores, transportadores e destinatários dos resíduos através de uma e-GAR, disponível na página da Autoridade Nacional de Resíduos (ANR) na internet [14].

S. R.
 MINISTÉRIO DO AMBIENTE, DO ORDENAMENTO DO TERRITÓRIO E DO DESENVOLVIMENTO REGIONAL
 Modelo A – GUIA DE ACOMPANHAMENTO DE RESÍDUOS N.º 6 0 9 2 9 4 3
Não aplicável a resíduos hospitalares

Dados do Produtor de Resíduos.

Designação dos Resíduos e do respectivo Código LER.

Dados do Transportador dos Resíduos.

Condições de transporte dos Resíduos (sacos de plásticos, a granel,...).

Dados do receptor e preenchimento somente por este no local de destino dos resíduos. É também efectuada a pesagem e se estiver tudo conforme a guia é assinada e entregue o duplicado ao produtor de resíduos.

1- PRODUTOR / DETENTOR	
Nome e endereço: _____	
Telefone: _____ Fax: _____ Telex: _____	
Pessoa a contactar: _____	
Designação do resíduo	Destino do resíduo
Indique o código correspondente (1)	Quantidade _____ kg litros
Assinale com um X qual o estado que melhor descreve o resíduo: Líquido <input type="checkbox"/> Pastoso <input type="checkbox"/> Sólido <input type="checkbox"/>	
<small>(1) Utilize a lista de resíduos em anexo</small>	
Declaração: certifico a exactidão das declarações prestadas e que o destinatário está devidamente autorizado a receber este resíduo.	
Data: _____	Assinatura do responsável produtor
<small>(Assinatura)</small>	
2- TRANSPORTADOR	
Nome e endereço: _____	
Telefone: _____ Fax: _____ Telex: _____	
Pessoa a contactar: _____	
Matrícula da viatura transportadora	
Identificação do meio de transporte	
Condições de acondicionamento do resíduo	
TIPO	MATERIAL
<input type="checkbox"/> Tambor	<input type="checkbox"/> Aço
<input type="checkbox"/> Barrica de madeira	<input type="checkbox"/> Alumínio
<input type="checkbox"/> Jerrycane	<input type="checkbox"/> Madeira
<input type="checkbox"/> Caixa	<input type="checkbox"/> Matéria plástica
<input type="checkbox"/> Saco	<input type="checkbox"/> Vidro, porcelana ou grés
<input type="checkbox"/> Embalagem composta	<input type="checkbox"/> Outro (indique qual)
N.º DE EMBALAGENS OU RECIPIENTES	
Data: _____	
Assinatura do responsável transportador	
<small>(Assinatura do transportador)</small>	
3- DESTINATÁRIO	
Nome e endereço: _____	
Telefone: _____ Fax: _____ Telex: _____	
Pessoa a contactar: _____	
Data da recepção do resíduo: ____/____/____ Identificação do meio de transporte	
Recepção aceite	Recepção recusada
Quantidade: Peso dos Resíduos _____ kg litros	Motivo: _____
Data: _____	Assinatura do responsável destinatário
<small>(Assinatura)</small>	

Infraestrutura onde são colocados os resíduos (Aterro Sanitário, Estação de Transferência, Estação de triagem ou Ecocentro).

Os resíduos podem ser pesados na origem, mas o valor oficial é pesado na báscula da infraestrutura que recebe os resíduos.

Modelo n.º 1428 (Actualizado 03/04/01)

Figura 2.1: Guia de acompanhamento de resíduos. Modificada [ERSUC, <http://ersuc.pt/www/index.php/componentes/componente-contacto/guia-de-acompanhamento-de-residuos-gar-como-preencher>].

2.1.4 Técnicas para tratamento de resíduos perigosos

No que diz respeito a técnicas para tratar os resíduos, existe uma hierarquia a seguir: em primeiro lugar é necessário reduzir ou evitar a produção de resíduos, logo a sua nocividade para o meio ambiente e saúde humana; em segundo lugar, deve averiguar-se se não existe uma forma de reutilizar o produto e/ou rentabilizar as suas propriedades e em terceiro lugar, caso as técnicas descritas anteriormente não sejam viáveis, há que tentar aproveitar os materiais constituintes dos resíduos para serem transformados em novos produtos, ou seja, reciclados.

Estes três princípios consubstanciam a sigla dos 3 R's: Reduzir, Reutilizar e Reciclar, que nem sempre é devidamente respeitada.

A indústria gera uma grande quantidade de resíduos perigosos, sendo uma, senão a principal causadora do seu aparecimento. Assim, antes de mais, é necessário avaliar se as indústrias estão a potenciar as suas capacidades para reduzir ao máximo a produção deste tipo de resíduos, evitando assim os problemas associados ao seu tratamento [16].

De acordo com o Regulamento dos Centros Integrados de Recuperação, Valorização e Eliminação de Resíduos Perigosos, anexo à Portaria nº 172/2009, de 17 de Fevereiro, é possível considerar quatro grupos de resíduos no que toca ao seu tratamento:

1. Resíduos perigosos que necessitam de tratamento devido à sua fracção inorgânica,
2. Resíduos perigosos passíveis de serem sujeitos a tratamento devido à sua fracção orgânica, com o objectivo de eliminar a sua perigosidade ou permitir a sua regeneração/reciclagem,
3. Resíduos perigosos que necessitam de tratamento em virtude da sua fracção orgânica, tendo como finalidade a valorização energética,
4. Embalagens contaminadas com resíduos perigosos.

Os resíduos referidos no grupo 1 não devem ser reciclados. Maioritariamente este tipo de resíduos passa por um tratamento físico-químico de estabilização ou inertização, antes de ser depositado em aterros. Este tipo de resíduos inclui-se nos resíduos inorgânicos da Lista Europeia de Resíduos (LER) e refere-se aos resíduos com os código 01 (excepção de 01 05 05), código 06 e código 09 (com excepção de 09 01 11).

Os resíduos descritos nos itens 2 e 3, que carecem de tratamento devido à sua fracção orgânica, devem ser sujeitos a um tratamento físico-químico para posterior reciclagem ou valorização energética. No entanto, existem resíduos que tornam a reciclagem ou a valorização energética um processo difícil. Nestes casos, pondera-se recorrer à incineração ou à co-incineração, minimizando os custos e dando-lhes o tratamento mais adequado.

Ainda relativamente aos resíduos do grupo 2, terminadas as operações de tratamento físico-químico, deverá ser feita a caracterização dos materiais residuais para se garantir que as emissões do processo respeitam a legislação.

Os resíduos inseridos no grupo 3 devem ser sujeitos a uma análise dos constituintes orgânicos e de materiais pesados, cuja concentração está sujeita a limites máximos. Este tipo de resíduos engloba os códigos 05, com excepção de 05 07

01; 07, com excepção de resíduos fortemente halogenados e 08, excepto quando contenham metais pesados. Para este tipo de resíduos é necessário implementar uma metodologia que garanta que o combustível de substituição não contém substâncias perigosas na sua manipulação, transporte e queima.

O destino a dar ao último grupo de resíduos passa pela seguinte ordem preferencial: reutilização, reciclagem e valorização energética. As embalagens destinadas à valorização material adoptam o tratamento dos resíduos do item 2; as embalagens destinadas à valorização energética carecem do tratamento dos resíduos do grupo 3 [13].

2.1.4.1 Tratamento físico-químico

Os métodos físicos não modificam a natureza química dos contaminantes, enquanto que os métodos químicos transformam os contaminantes em substâncias não prejudiciais para o ambiente ou para a saúde pública.

Com o tratamento químico pretende-se remover substâncias nocivas presentes nos resíduos, estando o resultado destes métodos dependente da natureza química das substâncias em questão.

As reacções químicas predominantes neste tratamento são processos de ácido-base, oxidação-redução, assim como reacções de precipitação. Para diminuir a toxicidade de uma espécie, utiliza-se o processo de redução do ião crómio hexavalente (Cr^{6+} , representado por Cr(VI)) a ião trivalente (Cr^{3+} , representado por Cr(III)) seguida da reacção de precipitação [13].

Processo precipitação e redução

Este processo é indicado para separar os metais pesados em soluções aquosas. Cada metal possui um intervalo de pH óptimo, de acordo com o qual o ião a ele associado vai provocar a precipitação sob a forma de um sal insolúvel.

O ião crómio hexavalente (Cr(VI)) é mais solúvel e mais perigoso que o ião trivalente (Cr(III)). Por isso, os cromatos têm que ser reduzidos antes de se remover o crómio trivalente por precipitação. Após a redução da forma hexavalente, é necessário efectuar um ajuste de pH e a consequente remoção dos metais. A adição de sulfureto de sódio ou o bissulfureto de sódio (Na_2S) forma sais metálicos menos solúveis¹.

A adição destes sais deve ser devidamente controlada, pois pode levar à formação de gases tóxicos (provocar a libertação de sulfureto de hidrogénio, H_2S), o que implica um perigo para a saúde e segurança pública.

Quando os iões de alumínio (Al^{3+}) e de cálcio (Ca^{2+}) estão presentes na mistura, os metais pesados presentes nas soluções aquosas são facilmente separados. O controlo de pH é um dos factores determinantes no tipo de tratamento a efectuar. Uma das formas mais utilizadas para alterar o pH consiste em adicionar cal ou soda cáustica. Devido ao baixo custo da cal, esta é a melhor escolha para as empresas [16, 17].

¹A adição de sulfureto provoca a precipitação de uma grande quantidade de metais pesados perigosos tais como o chumbo, mercúrio, cádmio, níquel, crómio, etc.

2.1.4.2 Unidades de estabilização

Este processo pode ser feito de diversas formas, como por exemplo, estabilização química por neutralização, absorção e adsorção. A estabilização química com reagentes é o processo usado com mais frequência, desde que os parâmetros de lixiviabilidade estipulados do Decreto-Lei nº183/2009 sejam cumpridos [18].

Neutralização

A neutralização é o nome que se dá ao ajuste de pH. Quando um efluente possui uma elevada concentração de iões metálicos, é possível através deste processo removê-los por precipitação química. Tal como no processo de precipitação, os metais precipitam numa gama de pH dependendo do ião metálico.

Neste processo, existem diferentes factores que são importantes: o reagente, a solubilidade do ião metálico e a temperatura. A solubilidade depende do produto de solubilidade do metal a remover e a temperatura afecta a solubilidade deste.

Um sistema de neutralização química com controlo de pH é constituído por um tanque e por um medidor de pH.

Nesta unidade são estabilizadas lamas de tratamento de efluentes gasosos e, por exemplo, resíduos de valorização de solventes usados [16, 17].

Adsorção

É um processo de remoção de contaminantes cuja eficiência está relacionada com a área superficial disponível para atrair as moléculas. A descontaminação é feita pelo uso de colunas utilizando carvão activado. Este remove compostos orgânicos, azoto, sulfuretos e metais pesados [16].

2.1.4.3 Unidades de tratamento de resíduos orgânicos

Neste processo empregam-se métodos físicos para a separação de sólidos/líquidos, tais como sedimentação, filtração e floculação. O critério do processo é a valorização da fracção orgânica contida nos resíduos, sendo a fracção inorgânica não recuperável e os contaminantes orgânicos de elevado peso molecular depositados em aterros [13].

Floculação

Este processo é usado para separar os sólidos dos líquidos, quando as suas velocidades de sedimentação são muito pequenas, não sendo possível uma clarificação efectiva.

A floculação consiste na aglomeração de partículas coloidais através de um tratamento de coagulação. Assim, a velocidade de coagulação tem de ser reduzida para não haver a degradação do tamanho dos flocos, assim como para facilitar o contacto dos flocos uns com os outros [17].

Coagulação

Consiste em adicionar e misturar um coagulante com uma velocidade elevada para neutralizar as cargas, o que diminui a repulsão entre elas, facilitando, deste

modo, a união entre partículas de maiores dimensões [17].

Filtração

Consiste em separar um sólido de um líquido, retendo o sólido e deixando passar o fluido[17].

2.1.4.4 Incineração e Co-Incineração

O tratamento de resíduos sujeitos a incineração ou a co-incineração tem como objectivo prevenir ou reduzir ao mínimo os seus efeitos negativos sobre o meio ambiente e a saúde pública.

Incineração

A incineração é uma forma de tratamento de resíduos maioritariamente orgânicos, que utiliza a combustão a altas temperaturas por reacção destes compostos com o oxigénio do ar.

Este processo apenas acontece quando não é possível a reutilização ou a reciclagem. Uma vez efectuado, espera-se que haja o máximo de aproveitamento energético.

Pretende-se com o processo de incineração reduzir o volume e a massa de resíduos através da combustão, destruindo os seus componentes orgânicos e originando a produção de cinzas e emissões gasosas inócuas.

Neste tipo de processo, o oxigénio é fornecido em excesso para permitir uma combustão completa. A capacidade de destruição das moléculas orgânicas depende da temperatura a atingir e do tempo de permanência a alta temperatura.

Tratando-se de um processo de incineração, as substâncias a utilizar devem possuir um poder calorífico significativo.

Existem dois tipos de incineradores: os que se situam dentro do complexo industrial, tratando resíduos próprios da sua actividade e os que se situam no exterior das unidades industriais, cujo objectivo é unicamente lucrar, ao tratar vários tipos de resíduos vindos de outros locais.

Um sistema de incineração exterior das unidades industriais é constituído por quatro unidades:

1. Unidade de pré-tratamento e armazenagem de resíduos,
2. Unidade de combustão,
3. Unidade de recuperação e conversão de energia,
4. Unidade de depuração e controlo de efluentes.

A primeira unidade é aquela que tem por objectivo distinguir os diferentes tipos de resíduos armazenados e fazer a sua análise química.

Na segunda unidade, são utilizados fornos rotativos devido à sua capacidade de tratar resíduos na forma sólida e líquida. Este tipo de fornos é constituído por duas partes: o forno rotativo e a câmara de pós-combustão. Os resíduos são introduzidos na parte mais elevada do forno rotativo. A câmara encontra-se no final do forno rotativo para receber os gases e as cinzas que dele provêm. Nesta câmara completa-se a queima dos gases.

Na parte superior do forno rotativo encontra-se um injector de combustível auxiliar, que provoca o arranque da unidade, a fim de ser alcançada a temperatura ideal para a introdução e queima dos resíduos. Este injector entra também em funcionamento durante o processo de incineração, sempre que o poder calorífico dos resíduos não seja suficiente para manter a temperatura nos níveis desejados.

Os fornos rotativos funcionam em co-corrente, sendo o ar comburente adicionado ao forno pelo mesmo local por onde são introduzidos os resíduos.

Os resíduos começam a ser aquecidos na parte superior do forno rotativo e, enquanto que a fracção volátil é gaseificada, a parte sólida ainda em combustão desloca-se para a parte inferior do forno, por força da sua inclinação e do movimento de rotação exercido. As cinzas e outras substâncias sólidas são depois deslocadas para um cinzeiro, devido à gravidade, onde arrefecem. Os gases remanescentes são extraídos para a câmara de pós-combustão, onde fica concluída a reacção de oxidação.

Na terceira unidade, a recuperação de energia térmica é feita numa caldeira de recuperação de calor, onde é produzido vapor que pode ser aproveitado directamente em indústrias para aquecimento central ou para geração de energia eléctrica.

Com a recuperação e a conversão energéticas, é possível obter uma poupança significativa, não só ao nível dos recursos energéticos, como também dos meios financeiros e ainda diminuir a emissão de gases de estufa e poluentes.

Na última unidade, os elementos vestigiais não destruídos, assim como as substâncias resultantes do enxofre, do cloro e do flúor têm de sair com os efluentes.

Os gases resultantes da incineração têm de sofrer um tratamento posterior, uma vez que podem estar contaminados com concentrações elevadas de metais tóxicos ou conter poluentes resultantes da queima incompleta. Estes gases podem ser compostos por substâncias tóxicas como chumbo, cádmio, mercúrio, crómio e/ou outros metais pesados, como ácido clorídrico, óxidos de azoto e dióxidos de enxofre, dioxinas e furanos.

Este processo possui alguns inconvenientes, tais como, elevados custos de investimento e de manutenção, emissão de substâncias como dioxinas e furanos e cinzas produzidas pela combustão do processo com elevado teor em metais pesados.

2.1.4.5 Aterros

A deposição de resíduos industriais perigosos em aterros exigirá especial cuidado e, pelo número de transformações que os resíduos podem sofrer com o tempo, esta deverá ser a última alternativa a considerar, tal como foi definido na hierarquia de gestão de resíduos.

Os aterros são classificados de acordo com os resíduos que lá são depositados: aterros para resíduos inertes; aterros para resíduos perigosos e aterros para resíduos não perigosos. Um aterro é geralmente uma cavidade efectuada no solo, em que o *input* são os resíduos depositados a grande profundidade e o *output* são gases e lixiviados. O material armazenado nos aterros inclui matéria orgânica parcialmente biodegradável e materiais inorgânicos. Os gases emitidos a partir dos aterros são utilizados para combustão controlada, a fim de se disseminarem os constituintes perigosos na atmosfera.

Os resíduos que se encontram associados a líquidos e que constituem pastas ou lamas, de composição muito variada, terão de ser inertizados antes de serem colocados em aterros. Os resíduos devem ser colocados em células, sendo estas ocupadas por curtos períodos de tempo.

Para a protecção do solo, é feito um revestimento do fundo com telas impermeáveis. A cobertura deve ser tal, que minimize a infiltração dos produtos no solo, por acção da precipitação.

O revestimento do fundo deve ser constituído por argila compactada e geomembrana, que servem como uma barreira composta ao movimento dos gases e líquidos. De seguida, é depositada uma câmara de areia ou gravilha, que tem como função a retenção e a drenagem de lixiviados que se formam no aterro. Imediatamente acima, é introduzida uma membrana geotêxtil, que é usada para minimizar as misturas do solo e areia. Na superfície, antes da introdução dos resíduos, é depositada uma camada de solo que permite a fixação da vegetação. [16, 19, 20]

De seguida, na (figura 2.2), será apresentado um esquema geral que representa o tratamento e o destino final a dar aos resíduos perigosos :

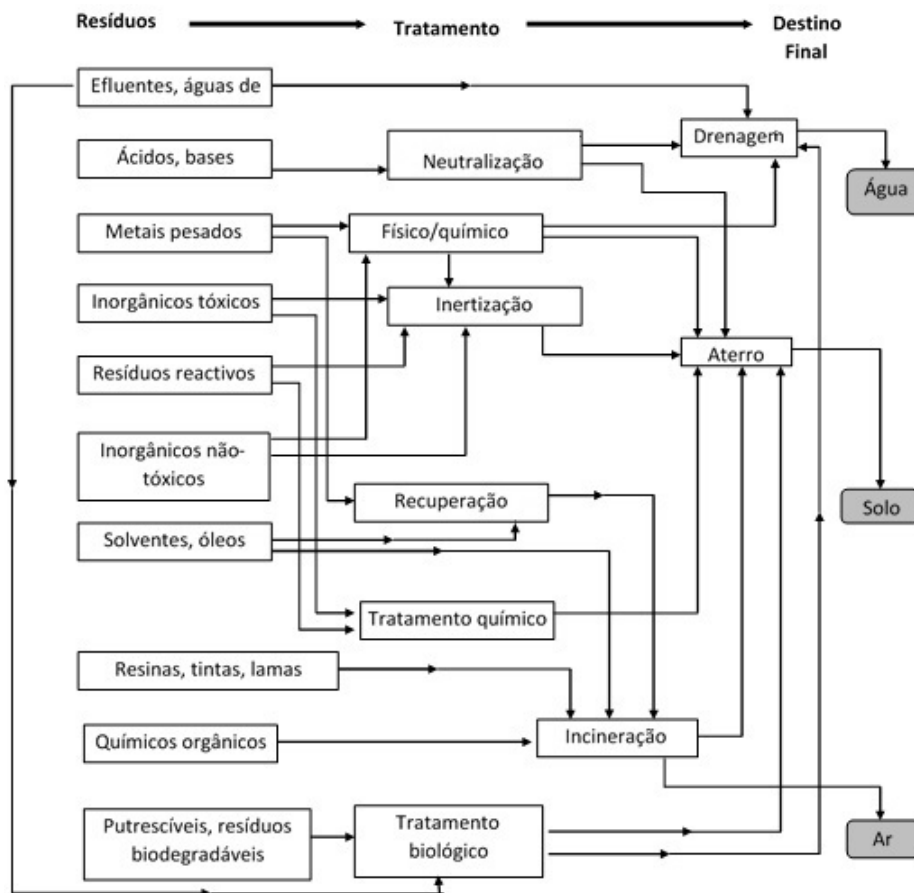


Figura 2.2: Esquema geral para o tratamento e desposição final de resíduos perigosos [21].

2.1.5 Resíduos em Indústrias Farmacêuticas

As indústrias farmacêuticas têm como produtos finais medicamentos destinados a diagnosticar, prevenir, curar ou aliviar doenças [22].

O processo de fabrico de comprimidos, até chegarem ao consumidor, passa por diferentes etapas, sendo que em todas elas são produzidos resíduos. As indústrias farmacêuticas são geradoras de uma grande quantidade de resíduos, devido à quantidade de comprimidos rejeitados, às perdas inerentes aos processos que possuem pó ou granulados, aos materiais de embalagem contaminados, aos resíduos considerados recicláveis e aos resíduos considerados urbanos [23].

A maioria dos resíduos produzidos na indústria farmacêutica é considerada perigosa, principalmente os reagentes e as substâncias activas utilizadas no fabrico de comprimidos. Assim sendo, deve ser-lhes dado um tratamento adequado, de modo a minimizar os danos ambientais e os da saúde pública.

Para uma gestão de resíduos eficiente, devem identificar-se e registar-se todos os resíduos gerados na empresa. Uma vez identificados e caracterizados, estes devem ser separados e acondicionados de forma independente. Enquanto os resíduos não são enviados para o destino final, devem ser armazenados de forma correcta e no devido local, nas instalações da empresa. Assim que exista um volume de resíduos considerável armazenado, o operador licenciado transporta os resíduos para os locais de tratamento adequado [11].

É de salientar que esta gestão interna é da responsabilidade do industrial. Assim sendo, e de acordo com a legislação, o industrial é responsável por qualquer contaminação que ocorra, assim como, por danos ambientais. Deste modo, é necessário que uma empresa implemente um ou mais procedimentos do controlo operacional para a gestão interna de resíduos, com regras de manuseamento, segregação, acondicionamento, transporte interno e armazenamento temporário dos resíduos produzidos. É também importante sensibilizar todos os colaboradores da empresa para efectuarem devidamente a separação dos diferentes tipos de resíduos. Contudo, é necessário não esquecer que os colaboradores devem estar informados sobre como proceder em caso de acidente ou derrame.

Como não poderia deixar de ser, na indústria farmacêutica, é essencial determinar qualitativa e quantitativamente os resíduos gerados, pois tal possibilita a implementação de novos métodos e procedimentos, tendo como finalidade diminuir ou minimizar os desperdícios e as perdas geradas no processo produtivo [14].

As empresas farmacêuticas procuram um Sistema de Gestão Ambiental que lhes permita prevenir e controlar os efeitos ambientais, de modo a operar de forma mais sustentável, tendo como objectivo tornarem-se mais competitivas no mercado [24].

Além disso, existem empresas que possuem já o certificado de Boas Práticas de Fabrico (GMP), tendo como objectivo diferenciar-se no mercado, pois respondem a um conjunto de recomendações que obriga as organizações a tomar medidas que garantam a segurança dos consumidores [25].

2.2 Sistema de Gestão Ambiental

Nos últimos anos tem havido uma intensa competitividade no mercado mundial, nomeadamente em termos de progresso tecnológico, nas exigências cada vez

maiores do mercado e nos recursos naturais cada vez mais escassos. Assim, é necessária uma constante mudança no funcionamento das empresas de modo a que se adaptem a novas condições, tendo como objectivo a redução dos impactos ambientais [8].

O Sistema de Gestão Ambiental ajuda as organizações a cumprir os regulamentos ambientais, a identificar os benefícios e a garantir que as políticas ambientais sejam adoptadas, numa óptica de redução do material desperdiçado e dos danos ambientais.

As empresas vêm adoptando a norma ISO 14001 como padrão para estabelecer o seu SGA.

Esta norma tem como objectivo prescrever e implementar metas, políticas e responsabilidades ambientais, assim como auditorias regulares com a finalidade de melhorar continuamente o desempenho das empresas, facilitando a competitividade no mercado mundial, alcançando uma confiança acrescida por parte dos clientes e da sociedade. Além disso, o SGA pode auxiliar o processo de decisão através de indicadores adequados que sustentem o programa de monitorização na organização. Deste modo, é necessário que a organização reveja periodicamente o seu Sistema de Gestão Ambiental, para identificar aspectos que possam facilitar a sua implementação.

Um SGA é aplicável a qualquer tipo de empresa, seja ela de grandes ou pequenas dimensões e refere-se à forma como esta gere e encara os impactos ambientais.

A norma tem ainda como objectivo fornecer directrizes através das quais as empresas projectam e implementam o SGA. Estas directrizes passam por três fases: a sua política ambiental, aspectos ambientais relacionados com as suas actividades e programas de gestão ambiental.

A política ambiental deve definir os princípios do desempenho ambiental de modo a que haja uma melhoria contínua do SGA, adoptar técnicas de trabalho que visam a redução da poluição gerada pelas diferentes fases do processo, cumprir os requisitos legais e os aprovados pela empresa.

Esta política é única para cada organização e, como tal, deve ser desenvolvida tendo em conta os impactos ambientais gerados pelas suas actividades e produtos.

A organização deve estabelecer, implementar e manter um procedimento para identificar os aspectos decorrentes da sua actividade que podem ter impacto no meio ambiente.

No que diz respeito ao programa de gestão ambiental, a organização deve mencionar como é que este programa está em concordância com a sua política, metas e objectivos ambientais, assim como os prazos, os meios e os responsáveis pela sua implementação. Espera-se que a organização reveja o seu programa periodicamente de modo a que seja actualizado e que sejam efectuadas revisões sempre que necessárias.

A norma ISO 14001 baseia-se no Ciclo de Deming (PDCA), com quatro etapas. A etapa *planear* consiste em estabelecer os objectivos e os processos necessários para atingir as metas de gestão da organização; a etapa *executar* consiste em implementar os processos decididos na etapa anterior; a etapa *verificar* confere se os objectivos foram alcançados e relata os resultados; a etapa *actuar* põe em prática acções para melhorar continuamente o desempenho do Sistema de Gestão Ambiental.

Para uma indústria farmacêutica, a implementação de um Sistema de Gestão

Ambiental constitui uma série de vantagens, conforme mencionadas a seguir: redução de custos devido à redução de matéria-prima, água e energia, pois existe uma melhoria da eficiência dos processos nas organizações, minimização do tratamento de resíduos e efluentes, redução de emissões e acidentes; vantagem competitiva no mercado mundial; maior qualidade dos processos tecnológicos [26].

2.2.1 Resíduos produzidos na Bluepharma

A Bluepharma é uma Indústria Farmacêutica, localizada em Coimbra, fundada a 1 de Fevereiro de 2001 por um grupo de investigadores portugueses.

A empresa tem autorização para fabricar/importar medicamentos na forma de cápsulas duras, pós, granulados e comprimidos.

Actualmente, a produção das formas sólidas satisfaz os requisitos associados às Boas Práticas de Fabrico (GMP), pelo que a Bluepharma concentra a sua produção nas formas farmacêuticas sólidas. A empresa trabalha em três principais áreas: desenvolvimento e registo próprio e para terceiros de novos medicamentos; produção e comercialização de medicamentos genéricos; e investigação e desenvolvimento

A empresa é reconhecida pela qualidade dos seus produtos, sendo certificada pela GMP e o modelo de boas práticas de gestão é apoiado no Sistema de Gestão Integrado (SGI) segundo as Normas da Qualidade (ISO 9001), Ambiente (14001), Segurança e Saúde Ocupacional (OHSAS 18001) e pelo Sistema Comunitário de Eco-Gestão e Auditoria (EMAS).

2.2.1.1 Principais resíduos produzidos na Bluepharma

A separação dos resíduos e a respectiva identificação é feita com base nos seus códigos, definidos na LER e de acordo com o determinado na legislação em vigor.

Os principais resíduos produzidos na Bluepharma são: resíduos farmacêuticos, resíduos provenientes do laboratório, resíduos de embalagem não contaminados (alumínio, PVC, PVDC)², resíduos de material de embalagem impresso (caixas e literatura), óleos e absorventes com óleo, areias com carvão activado, resíduos para reciclar (papel/cartão-incluindo documentação técnica, plástico, metal, vidro e embalagem), agulhas, material eléctrico e electrónico (ex: lâmpadas, baterias e acumuladores, etc) e resíduos pontuais.

2.2.1.2 Resíduos farmacêuticos

Os resíduos farmacêuticos constituem todo o material que resulta do processo produtivo que teve contacto com o produto e não é aproveitado (ex. blisters com comprimidos; blisters sem comprimidos, mas contaminados; comprimidos; cápsulas e pós) e ainda todos os consumíveis contaminados (luvas e/ou toalhas descartáveis).

Englobam-se, assim, nesta categoria:

²Actualmente o alumínio possui o número 179 da Lista Prioritária de Substâncias Perigosas e o cloreto de vinilo o número 4 [ATSDR, <http://www.atsdr.cdc.gov/sp1/>]. O cloreto de vinilo possui o número 88 da Lista de Poluentes Prioritários [EPA, <http://water.epa.gov/scitech/methods/cwa/pollutants.cfm>].

- Amostras em stock,
- Materiais de embalagem contaminada,
- Matéria-prima, produto semi-acabado e acabado rejeitado e matéria-prima obsoleta,
- Granulados e comprimidos/cápsulas rejeitadas,
- Reclamações/devoluções, assim como, misturas e restos de material de embalagem não utilizáveis, contaminados com produto,
- Luvas e toalhas descartáveis contaminadas com produto,
- Poeiras e outras partículas sólidas retidas em filtros e filtros de ar.

Os recipientes onde estão acondicionados os resíduos farmacêuticos estão perfeitamente identificados. Assim, todos os sacos, barricas, bidões ou contentores que contenham este tipo de resíduos são sempre identificados com uma etiqueta com linha vermelha.

2.2.1.3 Resíduos líquidos e sólidos provenientes dos laboratórios

Os resíduos líquidos provenientes dos laboratórios de Controlo de Qualidade (CQ) e Investigação e Desenvolvimento Analítico e Galénico (IDAG) são separados em jerricans adequados e devidamente identificados:

- Incineráveis sem halogéneos, fósforo e enxofre ³,
- Incineráveis com halogéneo (excepto flúor), fósforo e enxofre,
- Não incineráveis e resultantes dos ensaios de dissolução.

Os resíduos provenientes dos ensaios de dissolução cujo meio de dissolução é a água podem ser descarregados na Estação de Tratamento de Águas Residuais Industriais (ETARI), com excepção dos antibióticos.

Os jerricans que contêm os resíduos líquidos são encaminhados para o armazém, onde são vertidos para cubas. As cubas estão localizadas numa zona específica do armazém, onde existe uma infraestrutura - ventilação-forçada - concebida para proteger ao máximo os colaboradores que efectuem a trasfega dos resíduos líquidos dos recipientes para as cubas; apesar desta medida de segurança, os colaboradores utilizam ainda equipamentos de protecção individual: luvas de protecção e máscara panorâmica com filtros adequados para vapores e gases.

Todos os recipientes e cubas são reutilizados até deixarem de ser considerados como estando em boas condições, sendo então encaminhados como embalagens contaminadas.

No laboratório de microbiologia é ainda produzido um outro tipo de resíduo líquido, separado para jerrican devidamente identificado, proveniente dos ensaios realizados que contém:

- Meios de cultura (solução tamponizada),

³Actualmente o fósforo possui o número 19 da Lista Prioritária de Substâncias Perigosas [EPA, <http://www.atsdr.cdc.gov/spl/>].

- Pequenas quantidades de resíduos farmacêuticos (antibióticos) que resultam dos ensaios.

É adicionado no jerrican um desinfectante à base de cloro para evitar a proliferação de microrganismos⁴.

A identificação dos resíduos líquidos de microbiologia é feita com uma etiqueta com linha castanha.

Relativamente aos resíduos sólidos, no laboratório de CQ existem contentores de recolha, devidamente identificados, para a correcta separação de:

- PVC e alumínio,
- Embalagens compósitas,
- Material reciclável, tais como plástico, vidro, papel, cartão, caixas e literaturas/material de embalagem impresso,
- Resíduos domésticos.

Todos estes materiais, no caso de estarem contaminados com produto, são considerados resíduos farmacêuticos e, como tal, devem ser colocados nos respectivos recipientes.

2.2.1.4 Resíduos de processo de embalagem não contaminados

Durante o processo de montagem e acerto das máquinas de embalagem e mesmo durante o processo de embalamento são gerados resíduos não contaminados, isto porque não entram em contacto com o produto. Estes resíduos não contaminados são:

- Folhas de polímeros com ou sem alvéolas resultantes de acertos de máquinas (PVC, PVDC),
- Alumínio e blisters sem produto, resultantes do acerto das máquinas,
- Restos de polímeros selados com alumínio.

Estes resíduos são separados na embalagem em contentores próprios, com a designação PVC/alumínio e são identificados com uma etiqueta com linha azul claro.

2.2.1.5 Resíduos de material de embalagem impresso

Existe um conjunto de resíduos (caixas de medicamentos, literatura, etiquetas) que resulta de excedentes ou defeitos de fabrico.

A descaracterização efectiva deste material é um requisito legal, pelo que o destinatário deste resíduo tem de garantir a sua descaracterização e posterior reciclagem. A sua identificação é feita com uma etiqueta com linha azul-turquesa.

⁴O cloro possui o número 104 da Lista Prioritária de Substâncias Perigosas [ATSDR, <http://www.atsdr.cdc.gov/spl/>].

2.2.1.6 Óleos e absorventes com óleos

Na Bluepharma são produzidos óleos hidráulicos e óleos de lubrificação, que resultam da manutenção das máquinas. Estes óleos usados são colocados em bidões ou barricas existentes na Oficina de Manutenção e identificados com uma etiqueta com linha cinza claro.

Os resíduos de absorventes com óleos são depositados em pequenos recipientes azuis, estando os mesmos devidamente identificados com a respectiva etiqueta com linha de cinza escura. Estes resíduos resultam dos processos de desmontagem, montagem, manutenção e limpeza de equipamentos.

2.2.1.7 Areias com carvão activado

Estes resíduos gerados pela Estação de Tratamento de Águas Residuais Industriais são acondicionados dentro de barricas. Este tipo de resíduo é identificado com uma etiqueta com linha laranja claro.

2.2.1.8 Resíduos para reciclar

Existem pequenos recipientes para efectuar a separação dos resíduos que são posteriormente reciclados como papel, cartão, plástico e vidro.

Quando os sacos estão cheios, são levados para um local devidamente identificado para armazenamento temporário, situado no exterior do armazém, num contentor metálico alugado ao operador licenciado, que fará o devido encaminhamento os resíduos. Este contentor apenas contempla resíduos recicláveis. Neste sentido, o mesmo é identificado com as etiquetas adequadas e definidas para cada tipo de resíduo. Para resíduos de papel e cartão, a etiqueta utilizada possui uma linha azul escura; para resíduos de vidro utiliza-se uma etiqueta com linha verde alface e resíduos de plástico utiliza-se uma etiqueta com linha amarela.

Sempre que os resíduos de papel e cartão sejam provenientes de documentação técnica, e tendo em conta as características e a legislação da Indústria Farmacêutica, a descaracterização efetiva deste material, antes da sua reciclagem, é um requisito legal, pelo que o destinatário deste resíduo garante a sua descaracterização e posterior reciclagem. Após a destruição deste resíduo, o destinatário declara a sua descaracterização, garantindo assim a sua destruição. A etiqueta utilizada para este caso possui uma linha azul-marinho.

2.2.1.9 Materiais eléctricos e electrónicos

O material eléctrico e electrónico avariado ou fora de uso é colocado num contentor metálico. A sua identificação é feita com uma etiqueta com linha rosa.

2.2.1.10 Resíduos pontuais

Quando surgem resíduos pontuais é efectuada a sua identificação com base nos códigos definidos na LER.

Nesta etiqueta deve ser registada a descrição do resíduo e deve ser identificada com linha preta.

2.2.1.11 Agulhas

Estão disponíveis nos Laboratórios de CQ e IDAG recipientes destinados à colocação de agulhas usadas. As agulhas são colocadas num recipiente plástico rígido devidamente identificado. As etiquetas deste resíduo possuem uma linha roxa.

2.2.1.12 Resíduos de meios de dissolução

Após um estudo efectuado ao efluente líquido e a autorização por parte da empresa Águas de Coimbra, a separação e encaminhamento deste resíduo foi alterada. Sempre que o meio de dissolução não contenha antibióticos, ele é separado de acordo com a seguinte classificação:

- Meio de dissolução-fosfatos,
- Meio de dissolução-cloreto,
- Meio de dissolução-água descarregada directamente da ETARI.

Estes resíduos de meios de dissolução são encaminhados para a ETARI da empresa.

2.2.1.13 Separação e encaminhamento de resíduos

Antes de os colaboradores depositarem os resíduos em recipientes apropriados, e de modo a não colocar em causa a saúde e segurança de quem os manipula “*a posteriori*”, os mesmos devem ser devidamente acondicionados. Assim, a parte exterior do armazém é composta por dois pontos de armazenamento principais: a zona de estantes, barricas, paletes, cubas,... e o contentor metálico.

A zona de estantes, barricas, paletes, cubas,... contém:

- Resíduos farmacêuticos,
- Resíduos líquidos dos laboratórios (excepto os não incineráveis),
- Material eléctrico e electrónico, incluindo toners e tinteiros,
- Areia com carvão activado,
- Absorventes com óleo,
- Outros resíduos pontuais.

Todos os resíduos recicláveis são armazenados temporariamente no contentor metálico:

- Plástico, vidro, metal, papel e cartão, incluindo material de embalagem impresso e documentação técnica,
- Embalagens de plástico, de vidro, de metal, de papel e cartão e material de embalagem impresso,
- Resíduos de embalagem não contaminados.

Todos os resíduos na Bluepharma são devidamente acondicionados temporariamente, garantindo que não se ultrapassa o prazo legal de um ano, sendo os mesmos enviados através de operadores licenciados de gestão de resíduos para um destino final, onde são submetidos aos processos de tratamento adequado: valorização, reciclagem, incineração ou deposição em aterro [27].

2.2.2 Sistema de Gestão da Qualidade

O Sistema de Gestão da Qualidade consiste numa estrutura organizacional para a melhoria contínua, com o objectivo de aumentar a satisfação de clientes ou partes interessadas. Tem também como objectivo melhorar a imagem de uma organização perante o cliente e assim aceder a um mercado cada vez mais exigente.

O SGQ deve apoiar-se num Manual da Qualidade, que deve incluir procedimentos do sistema em questão, normas, especificações, registos e resultados do processo de trabalho [28].

2.2.2.1 ISO 9000

A série de normas ISO 9000 fornece a estrutura para as organizações instalarem um SGQ, seguindo certos procedimentos e processos que satisfazem as exigências dos clientes.

As normas da família ISO 9000 que se encontram actualmente em vigor são:

- ISO 9000: Sistema de Gestão da Qualidade - Fundamentos e vocabulário,
- ISO 9001: Sistema de Gestão da Qualidade - Requisitos,
- ISO 9004: Sistema de Gestão da Qualidade - Linhas de orientação para a melhoria do desempenho.

Para que a norma ISO 9000 seja implementada numa empresa com sucesso, foram identificados oito princípios de Gestão da Qualidade, de modo a que a mesma possua um melhor desempenho: 1) focalização no cliente, 2) liderança, 3) envolvimento das pessoas, 4) abordagem por processos, 5) abordagem à gestão através de um sistema, 6) melhoria contínua, 7) abordagem à tomada de decisão baseada em factos, 8) relação mutuamente benéfica com os fornecedores. De seguida será apresentada uma breve descrição de cada princípio.

- 1) A sobrevivência das organizações está dependente dos seus cliente. Assim, estas procuram compreender as expectativas e necessidades actuais e futuras dos clientes, visando atingir os seus requisitos e exceder as suas expectativas, considerando os interesses das partes interessadas.
- 2) Os líderes estabelecem a finalidade e a orientação da empresa com o intuito de definir objectivos e metas a cumprir, formalizar uma visão prospectiva do mercado, melhorar a confiança dos clientes e fornecer aos colaboradores formação necessária para que possam desempenhar as suas funções.
- 3) Reconhecimento da importância das pessoas da organização no seu sucesso.
- 4) *“Um resultado desejado é atingido de forma mais eficiente quando as actividades e os recursos associados são geridos com um processo”* (ISO

9000), pelo que qualquer organização deve ter como objectivo verificar resultados através da satisfação dos clientes, gerir as interfaces com os clientes/fornecedores internos e focalizar-se nos processos e não na função.

- 5) O sistema visa melhorar a eficácia e a eficiência da organização através da identificação, compreensão e gestão de processos inter-relacionados.
- 6) De forma a melhorar o desempenho das empresas, há que diagnosticar e intervir continuamente nos diferentes processos .



Figura 2.3: Ciclo de Deming (PDCA): *Plan* (planear) - estabelecer objectivos e metodologias com vista à satisfação do cliente; *Do* (executar) - implementar as metodologias aplicadas anteriormente; *Check* (verificar) - monitorizar e medir processos ou produtos e relatar os dados; *Act* (actuar) - exercer acções de melhoria contínua.

- 7) “*As decisões eficazes são baseadas em análises de dados e de informações*” (ISO 9000). As tomadas de decisão devem ser fundamentadas numa análise de informação comprovada pela experiência e apoiada pela empresa.
- 8) Uma organização e os seus fornecedores são interdependentes e uma relação de benefício mútuo potencia a aptidão de ambas as partes para criar valor [29].

2.2.2.2 ISO 9001:2008

A norma ISO 9001:2008 especifica os requisitos de um Sistema da Gestão da Qualidade e está organizada segundo a metodologia PDCA - tem como finalidade aumentar a satisfação do cliente e melhorar continuamente o desempenho das organizações, indo ao encontro dos seus requisitos.

Um processo é definido como uma actividade ou um conjunto de actividades que transformam as entradas em saídas (figura 2.4).



Figura 2.4: Representação da actividade de uma empresa sob a forma de segmentos de um processo [30].

O processo é iniciado por entradas que vêm do fornecedor e como resultado deverá produzir uma saída que vai para o cliente. A saída é o resultado de um processo, podendo revestir a forma de resíduos ou poluição.

As entradas podem ser internas ou externas à organização. Esta fase inclui factores tais como: a competência das pessoas, documentação, capacidade e monitorização do equipamento e saúde, segurança e ambiente de trabalho. As entradas são definidas e registadas como suporte de base aos requisitos a serem utilizados na verificação e validação de saídas.

As saídas devem ser verificadas de forma a terem em consideração as necessidades e expectativas dos clientes. Estas saídas devem ser documentadas face aos requisitos de entrada. A verificação do produto deve ser efectuada durante o processo, para identificar variações e para averiguar se é necessário adoptar acções preventivas e de melhoria da eficiência e eficácia do processo. As saídas incluem os dados de entrada e saída, especificações do processo e produto, assim como, especificações de materiais.

Frequentemente, a saída de um processo constitui directamente a entrada do próximo.

De uma maneira geral, os requisitos mencionados anteriormente realçam os seguintes aspectos: melhoria contínua do desempenho de organizações, aumento da satisfação dos colaboradores, redução dos custos devido à diminuição de desperdícios, reclamações e rejeições, melhor imagem da empresa contribuindo para atrair a confiança dos seus clientes, aumento da satisfação dos clientes, acesso a mercados e a clientes cada vez mais exigentes e adopção das mais actuais ferramentas de gestão.

2.2.2.3 ISO 9004

Esta norma fornece linhas de orientação para a melhoria do desempenho do Sistema de Gestão da Qualidade.

Esta norma complementa a norma ISO 9001, mas ambas podem ser utilizadas de forma independente. Esta norma não é destinada para certificação.

A norma ISO 9004 visa dar um contributo significativo ao SGQ no sentido de melhorar o desempenho, a eficiência e a eficácia de uma organização [30].

2.2.3 Sistema Comunitário de Ecogestão e Auditoria

O Sistema Comunitário de Ecogestão e Auditoria (EMAS) é aplicado a todas as organizações que pretendem melhorar o seu desempenho ambiental mediante o

estabelecimento e a implementação do SGA, bem como uma avaliação sistemática, objectiva e periódica do desempenho desses mesmos sistemas. O EMAS tem também como objectivo prestar informações relevantes sobre o desempenho ambiental ao público e a outras partes interessadas. Neste sentido, a organização deve identificar e quantificar os aspectos ambientais relacionados com os seus processos que podem ser controlados e aqueles sobre os quais tem apenas influência. Os aspectos ambientais que estão relacionados com os seus processos e que se devem ser tidos em conta na revisão são, por exemplo: emissões para a atmosfera e água, produção e gestão de resíduos, uso e contaminação de solos, uso de recursos naturais, energia e matéria-prima, risco de acidentes, incidentes e derrames e questões locais (ruído, vibrações, poeiras...).

Os aspectos ambientais sobre os quais a organização apenas tem influência são: novos mercados, questões relacionadas com o ciclo de vida de um produto, questões ambientais e investimento de capital.

Os benefícios que advêm da participação das organizações no EMAS são: melhor imagem da organização no mercado, maior confiança por parte dos clientes, Gestão Ambiental de Qualidade e aumento da credibilidade, em geral [31].

2.2.3.1 ISO 14001 e EMAS

O regulamento EMAS é semelhante à norma ISO 14001 relativamente aos requisitos adoptados para a implementação do Sistema de Gestão Ambiental. No entanto, esta norma e este regulamento possuem algumas diferenças, nomeadamente, o EMAS é mais rigoroso que a ISO 14001 na análise dos aspectos ambientais e exige que as organizações emitam uma declaração ambiental que inclua a descrição da organização e as suas actividades, os produtos, a sua política ambiental, uma breve descrição do SGA e os resultados do programa ambiental com o objectivo de fornecer informações validadas ao público [32].

2.3 Sistema de Gestão de Segurança e Saúde no Trabalho

Os progressos tecnológicos constantes e a intensa competitividade trouxeram uma alteração das condições de trabalho e processos na organização. Assim, o desenvolvimento de um Sistema de Gestão de Segurança e Saúde no Trabalho eficiente é necessário numa empresa para informar e levar os colaboradores a proceder de forma prudente e saudável, de modo a que esta melhore as suas condições de trabalho.

A família de normas OHSAS é aplicável à gestão de segurança e saúde no trabalho, isto é, à forma como a organização gere os riscos operacionais para com os trabalhadores e outras partes interessadas.

A família norma ISO 18000 é composta por:

- Norma ISO 18001:2007 - Sistema de Gestão da Segurança e Saúde no Trabalho - Requisitos,
- Norma ISO 18002:2008 - Sistema de Gestão da Segurança e Saúde no Trabalho - Orientações para a implementação da OHSAS 18001:2007.

Esta família de normas é aplicável a qualquer organização e requer que esta estabeleça um Sistema de Gestão Integrado.

2.3.1 OHSAS 18001:2007

A OHSAS 18001:2007 tem como objectivo estabelecer os componentes para a construção de um Sistema de Gestão de Saúde e Segurança no Trabalho eficaz que visa minimizar os riscos de acidentes e garantir a protecção dos recursos humanos. Neste sentido, esta norma garante que as empresas atendam aos seus requisitos legais, sociais e financeiros associados à saúde e segurança ocupacional. Semelhante à ISO 14001, a OHSAS 18001 também é baseada no ciclo de PDCA e são muitos os elementos comuns aos Sistemas de Gestão, de modo a que todos os perigos da Segurança e Saúde no Trabalho (SST) sejam identificados continuamente, os respectivos riscos avaliados e definidas estratégias para os minimizar.

A política da SST de uma organização deve ter em conta os riscos associados aos processos, produtos ou outras actividades desenvolvidas. A política deve ser coerente com os riscos, legislação e deve ser comunicada a toda a empresa.

A organização deve redigir os procedimentos de identificação e avaliação de riscos, bem como a forma de os prevenir, minimizar e se possível, eliminar. Assim, a organização deve estabelecer procedimentos de modo a que os perigos sejam identificados continuamente e os respectivos riscos sejam avaliados, tendo como objectivo definir estratégias para os minimizar.

Finda esta etapa, os processos são implementados. Após o planeamento e a implementação, é necessário monitorizar e medir os processos face às políticas da SST, tais como o seu desempenho. Por fim, a gestão de topo deve rever periodicamente o OHSAS para verificar a implementação dos objectivos estabelecidos. Esta revisão serve também para a empresa avaliar se a política continua a ser apropriada e estabelecer novas metas para melhorar continuamente o desempenho e a eficácia deste sistema de gestão.

2.3.2 OHSAS 18002:2008

A OHSAS 18002:2008 tem como finalidade apoiar as organizações que pretendem implementar ou melhorar o desempenho de Saúde e Segurança do Trabalho.

As principais vantagens que uma empresa retira da certificação de OHSAS são: redução de riscos de acidente, redução de custos relativamente a prejuízos resultantes de acidentes, melhoria da satisfação e motivação dos colaboradores, garantia de um ambiente de trabalho seguro e saudável, melhoria da imagem da organização, levando à sua aceitação no mercado e na sociedade, tendo como resultado uma vantagem competitiva, confiança acrescida por parte dos clientes e melhoria contínua da gestão e do desempenho do SST [33].

2.4 Tratamento de dados

De seguida vamos resumir algumas questões fundamentais referentes ao tratamento de resultados, suporte estatístico para a tomada de decisão e inferência de factores por via da modelação.

Os testes estatísticos são cruciais para permitir a verificação de algumas condições preliminares à estimativa e inferência estatística de modo a garantir que *a)* os resultados foram bem estimados, *b)* as conclusões foram bem tiradas, *c)* independentemente da avaliação heurística de cada interveniente na análise de dados.

De acordo com o objectivo principal deste trabalho - o de identificar causas mais relevantes na produção de resíduos no processo - foi necessário desenvolver estratégias para conseguir estabelecer relações do tipo causa-efeito. Deste modo, as variáveis processuais avaliadas foram classificadas em:

- causas (factores ou predictores) - sempre que se trata de condições estabelecidas ou impostas ao processo de fabrico e,
- efeitos (respostas) - às variáveis cujo resultado depende do processo, nomeadamente no que se refere à formação de resíduos.

A análise de factores à escala industrial não seria praticável dada a impossibilidade de controlar todas as variáveis do processo e impor condições rígidas de teste/ensaio que poderiam interferir e comprometer a normal laboração da empresa. Assim, na impossibilidade de intervir activamente no processo foi necessário desenvolver uma estratégia não invasiva de modo a obter informação sem interferir no sistema.

Optou-se então por apenas acompanhá-lo, auscultando cuidadosamente as variáveis do sistema, a fim de preservar a identidade de cada evento e a rastreabilidade do mesmo⁵. Deste modo, cada evento registado possui dois tipos de variáveis:

- causas (factores ou predictores) - são variáveis independentes que podem ser impostas ou directamente controladas e que vão descrever as condições em que o processo se desenvolve;
- respostas - são variáveis que resultam das condições processuais impostas, nas quais se dá especial ênfase à quantidade e tipo de resíduo formado.

Para evidenciar variáveis mais relevantes optou-se pela análise de Pareto, impondo a ordenação de causas/respostas por ordem decrescente de relevância de modo a preservar o máximo de efeito nas primeiras variáveis evidenciadas e, deste modo, concentrar a atenção nas mais relevantes.

No sentido de evidenciar o impacto de cada variável *per-si* sobre determinada resposta, foi realizada uma abordagem inicial por correlação embora se saiba que, nos sistemas multivariados, a correlação directa causa-efeito surge de um emaranhado complexo de interdependências e relações internas que muitas vezes obscurecem a verdadeira causa.

Para melhor avaliar esta interdependência multivariada, optou-se por utilizar um método de modelação implícita já que os modelos lineares utilizados, embora sejam rudimentares, permitem evidenciar as interacções simultâneas de efeitos sobre a resposta sem haver necessidade de impor qualquer modelo.

⁵A identidade de cada evento está relacionada com a sua identificação no tempo e espaço enquanto que a rastreabilidade se refere à coordenação de toda a informação relativa à respectiva identidade, previamente assinalada.

Já para averiguar a veracidade da interacção evidenciada por modelação implícita foi desenvolvido um procedimento de simplificação progressiva do modelo que descreve o sistema multivariado, com suporte estatístico, de forma a obter um modelo parcimonioso que melhor descreve o sistema e, deste modo, evidenciar as causas mais relevantes para cada uma das respostas estudadas.

2.4.1 Ferramentas da Qualidade

Segundo Bunney e Dale [3], a introdução da Gestão Total da Qualidade (TQM) ocorre em três fases: 1) diagnóstico e preparação; 2) foco da gestão e compromisso; 3) melhoria contínua. Na fase do diagnóstico e preparação são apurados os factos de um problema da empresa. A segunda fase utiliza as ferramentas de análise de dados com o objectivo de identificar os problemas, quantificar os seus efeitos e equacionar soluções. Na última fase, são utilizadas ferramentas mais complexas que ajudam na resolução de problemas, melhorando o desempenho da empresa. As ferramentas utilizadas na melhoria de processos são: fluxogramas, diagrama causa-efeito, diagrama de Pareto, gráfico de controlo, histograma, lista de verificação e diagrama de dispersão.

Consoante a natureza do problema, é feita a escolha da ferramenta mais adequada, logo, para um problema não é necessário utilizar todas as ferramentas mencionadas anteriormente. Deste modo, neste caso em particular, apenas se irá recorrer ao diagrama de Pareto [3].

2.4.1.1 Análise de Pareto

A origem deste tipo de abordagem deve-se a um italiano, Vilfredo Pareto [1848-1923], engenheiro civil de formação mas que desempenhou cargos de economista e desenvolveu ainda trabalhos relacionados com a sociologia e a política do início do século XX [34].

Pareto observou que, quando se está a avaliar o impacto de causas sobre efeitos em muitas situações reais, um número muito restrito de causas era em grande parte responsável pelo efeito global final - este avaliou que cerca de 20% das causas são frequentemente responsáveis por cerca de 80% do efeito na resposta estudada. Esta regra conhecida como 20:80 (cerca de 1/5 das causas justificam cerca de 4/5 do efeito) [35].

Na figura (2.5) vemos uma representação de um exemplo de diagrama de Pareto.

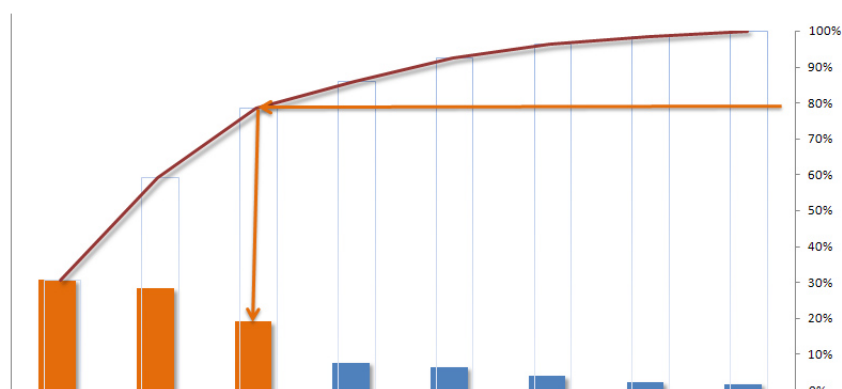


Figura 2.5: Diagrama exemplificativo do tipo de análise introduzida por Pareto; as setas evidenciam a detecção do grupo restrito de causas que são responsáveis pela maioria do efeito global.

Da figura anterior verifica-se que, ao fazer a ordenação das causas por ordem decrescente de importância, representadas por colunas, e estimando a contribuição cumulativa percentual respectiva, representada por uma linha sobreposta com o gráfico anterior, o conjunto de causas mais relevantes é responsável por cerca de 80% do fenómeno observado na totalidade da sua extensão.

O desenvolvimento desta análise de Pareto por via estatística revelou mais à frente que este facto estava relacionado com uma distribuição estatística baseada numa função potência.

Este tipo de análise é hoje em dia muito utilizado em diversos domínios científico-tecnológicos distintos como, por exemplo, na avaliação da qualidade, para evidenciar as maiores contribuições de factores sobre uma determinada resposta.

2.4.2 Testes estatísticos

Toda a análise estatística de resultados parte de um certo conjunto de pressupostos essenciais que condicionam drasticamente os resultados e convém por isso enunciar: *a)* ausência de erro sistemático, *b)* erro aleatório com distribuição normal e *c)* erro não correlacionado.

2.4.2.1 Nível de confiança e nível de significância

Ainda que a estimativa central esteja correcta⁶, devido à componente aleatória intrinsecamente envolvida na quantificação, há sempre uma certa margem de possibilidade de se cometer um erro de juízo na avaliação estatística de hipóteses já que se trata de um evento binomial onde existe uma hipótese inicial (H_0) e uma hipótese alternativa (H_1) que lhe é complementar,

$$p[H_0] + p[H_1] = 1 \quad (2.1)$$

Dado que a hipótese nula (H_0) é concebida no sentido da igualdade, esta será aceite por defeito, se não for demonstrada a sua falsidade.

⁶Estimativa concordante com o valor correcto.

Assim sendo, três situações distintas podem ocorrer no que diz respeito à decisão estatística sobre a validade das hipóteses em estudo, ver tabela 2.2.

Tabela 2.2: Teste estatístico de hipóteses: tipos de erros que podem ocorrer na tomada de decisão.

realidade decisão tomada	$H_0 = V$ ($H_1 = F$)	$H_0 = F$ ($H_1 = V$)
aceitar H_0	\checkmark $(1 - \alpha)$	erro do tipo II (β) (aceitação abusiva)
aceitar H_1	erro do tipo I (α) (rejeição abusiva)	\checkmark $(1 - \beta)$

Desta tabela conclui-se que existem duas situações correctas e duas anómalas.

Nas situações correctas a decisão tomada está em conformidade com a realidade enquanto que nas situações de erro a decisão tomada é antagónica.

Assim, os erros estatísticos cometidos podem ser de dois tipos:

a) erro por excesso - refere-se a uma falsa rejeição (rejeição abusiva da hipótese quando esta deveria ter sido aceite). Este erro chama-se de erro do tipo I (α) - a hipótese nula estava correcta e foi abusivamente rejeitada por ter sido considerada falsa;

b) erro por defeito - refere-se a uma falsa aceitação (aceitação abusiva da hipótese que estava incorrecta). Este erro designa-se por erro do tipo II (β) - a hipótese nula estava errada e foi abusivamente aceite como verdadeira.

Por uma questão de segurança estatística, é preferível cometer um erro por rejeição abusiva (erro do tipo I) do que um erro devido a uma aceitação abusiva (erro do tipo II).

A probabilidade máxima com que se pretende proceder à rejeição abusiva (erro do tipo I) é designada de nível de significância do teste e é designada de α (geralmente indicado em termos percentuais por $100\alpha\%$). A probabilidade de aceitação da hipótese correcta designa-se de nível de confiança e corresponde à probabilidade de $(1 - \alpha)$ (expressa em termos percentuais por $100(1 - \alpha)\%$).

Chama-se valor de prova à probabilidade de aceitação da hipótese nula, $p[H_0]$.

Na formulação de hipóteses estatísticas, a hipótese nula (inicial, H_0) vai no sentido de não haver diferença significativa: no sentido de pertencer ao grosso da distribuição também designado de $(1 - \alpha)$; a hipótese alternativa (H_1) visa essencialmente a diferença significativa (α): o complemento da hipótese nula.

2.4.2.2 Procedimento

Os testes estatísticos servem para, com critérios estatísticos, auxiliar a tomada de decisões na interpretação dos resultados. Os testes estatísticos devem ser efectuados com base numa certa sequência lógica de procedimento.

Formulação do problema

Em primeiro lugar deve efectuar-se uma análise do evento por forma a raci-

analisar a questão e poder testar. Procura-se deste modo traduzir a questão experimental para a lógica estatística: o que é que se pretende provar?

Escolha do teste

Com base no que se pretende verificar, o teste é escolhido atendendo à distribuição estatística que melhor se adequa a este tipo de teste. As distribuições estatísticas mais comuns para efectuar testes estatísticos são a normal, t -student, F -Fisher e qui-quadrado.

Estabelecimento do nível de significância

Antes de mais, tem que ser estabelecido o nível de significância com que se pretende tirar conclusões - prever qual o erro máximo admissível para se tirarem conclusões erradas por rejeição abusiva (α).

Regra geral o nível de significância refere-se a $\alpha = 0.05$ podendo também ser baixado para $\alpha = 0.01$ para serem tiradas conclusões mais definitivas, como no caso de testes que impliquem decisões críticas, tais como escolher o melhor modelo de calibração ou retirar pontos da curva de calibração, o nível de significância de referência é o valor $\alpha = 0.01$.

Hipóteses de trabalho

As hipóteses colocadas devem ser antagónicas e disjuntas (não sobreponíveis) por forma a abranger o universo do evento.

A hipótese nula (H_0) deve ser formulada no sentido de não haver diferença (está tudo correcto); a hipótese alternativa (H_1) incide sobre a diferença significativa (procura demonstrar a não conformidade) e é estabelecida de forma a abranger o universo complementar da hipótese nula.

Simetria do teste

Um vez que a hipótese nula vai sempre no sentido da igualdade, a simetria do teste a efectuar depende da hipótese alternativa formulada. Se o que se pretende é um teste de desigualdade ($H_1 : x_i > x_j$) apenas se está interessado em comparar a sobreposição unilateral de caudas (valores extremos) das duas distribuições o que corresponde a um teste unilateral. Se o que se pretende é um teste de igualdade ($H_0 : x_1 \simeq x_2$), está-se interessado em comparar a sobreposição central do corpo das distribuições (valores interiores, intervalo de confiança) o que irá corresponder a um teste bilateral (as áreas de rejeição situam-se nas duas zonas extremas à distribuição).

Cálculo do teste

O valor experimental da função discriminante, TV (parâmetro estatístico) é calculado com base na expressão da distribuição estatística subjacente. Em geral os testes estatísticos baseiam-se nas distribuições t -student, F de Fisher e χ^2 .

Comparação com valores críticos

Os valores críticos estão tabelados de acordo com o nível de significância e com o número de graus de liberdade. Estes valores permitem definir as regiões de aceitação e de rejeição das hipóteses formuladas.

Sempre que possível deve estimar-se o respectivo valor de prova, $p[H_0]$, no sentido de confirmar a conclusão anteriormente assumida.

Conclusão

No caso do valor experimental exceder os limites tabelados (entrar na região de rejeição) diz-se que ao nível de confiança $100(1 - \alpha)\%$ há diferença significativa e a hipótese inicial deve ser rejeitada em detrimento da hipótese alternativa (complementar); caso contrário, não há evidência estatística significativa para rejeitar a hipótese nula.

As normas ISO referentes à qualidade recomendam testes de significância aos níveis de 5% ($\alpha = 0.05$, probabilidade de efectuar 1 insucesso em cada 20 decisões) e 1% ($\alpha = 0.01$, probabilidade de taxa de insucesso de 1/100).

Assim, três situações podem ser obtidas,

$$\left\{ \begin{array}{ll} TV \leq X_{0.05(\nu)} & \implies \text{H0 aceite} \\ X_{0.05(\nu)} < TV \leq X_{0.01(\nu)} & \implies \text{H0 dúbia} \\ TV > X_{0.01(\nu)} & \implies \text{H1 aceite} \end{array} \right. \quad (2.2)$$

2.4.2.3 Testes baseados nas distribuições estatísticas

As distribuições estatísticas (t -student, F e χ^2) estão muito relacionadas com o teste de hipóteses já que estas, devido a serem bem caracterizadas, podem ser utilizadas como termo de comparação em diversas situações reais, providenciando os valores críticos necessários.

Distribuição t -student

Quando está em causa a comparação de estimativas de posição, assumindo que se trata de distribuições normais e independentes, calcula-se o valor teste

$$TV = \frac{|\bar{x} - \mu|}{\sigma_{\bar{x}}} \leq t_{\alpha(n-1)} \quad (2.3)$$

e compara-se este resultado com o valor crítico da distribuição t -student ao nível de confiança de $100(1 - \alpha)\%$. Se a hipótese nula assumir a igualdade de valores a hipótese alternativa assume a desigualdade,

$$\left\{ \begin{array}{l} H_0 : \bar{x} = \mu \\ H_1 : \bar{x} \neq \mu \end{array} \right. \iff \left\{ \begin{array}{l} H_0 : \bar{x} = \mu \\ H_1 : \bar{x} < \mu \vee \bar{x} > \mu \end{array} \right. \quad (2.4)$$

e consequentemente o valor crítico refere-se à distribuição cumulativa bilateral (hipótese centrada assumindo eventual diferença nos extremos). Caso se procure demonstrar que o valor estimado é superior ao valor correcto, a hipótese nula assume a situação da igualdade (valor inferior ou igual a) enquanto que a hipótese alternativa dá conta desta diferença,

$$\left\{ \begin{array}{ll} H_0 : \bar{x} \leq \mu & p[H_0] \sim (1 - \alpha) \\ H_1 : \bar{x} > \mu & p[H_1] \sim (\alpha) \end{array} \right. \quad (2.5)$$

Os valores críticos correspondem à distribuição cumulativa unilateral (hipótese marginal).

Distribuição F de Fisher

Quando está em causa a comparação de variâncias de distribuições normais aleatórias e independentes, calcula-se o valor teste

$$TV = \frac{\sigma_1^2}{\sigma_2^2} \quad (2.6)$$

que, no caso de distribuições normais e independentes, deve seguir uma distribuição de Fisher. Como hipóteses assume-se como hipótese inicial (H_0) que não há diferença significativa entre as variâncias (diferença puramente aleatória),

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \iff \begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \vee \sigma_1^2 > \sigma_2^2 \end{cases} \quad (2.7)$$

o que corresponde a efectuar um teste bilateral para o quociente das variâncias. A hipótese nula (H_0) assume que não há diferença significativa entre variâncias e a hipótese alternativa (H_1) assume que se houver diferença significativa, a variância do numerador excede a do denominador.

O teste F também pode ser efectuado de forma diferencial, também conhecido como teste de Mandel [36]. Este processo consiste em comparar a alteração da variabilidade ($\Delta\sigma^2 = \Delta SS/\Delta\nu$), o aumento da variabilidade resultante de uma diminuição do número de graus de liberdade, ($\nu_1 - \nu_2 > 0$) com uma estimativa puramente aleatória da dispersão ($\sigma_{pe}^2 = \sigma_2^2$)

$$TV = \frac{\Delta\sigma^2}{\sigma_{pe}^2} = \frac{\Delta SS/\Delta\nu}{\sigma_{pe}^2} = \frac{(\nu_1 \times \sigma_1^2 - \nu_2 \times \sigma_2^2)/(\nu_1 - \nu_2)}{\sigma_2^2} \quad (2.8)$$

Como hipóteses considera-se H_0 : os objectos/parâmetros removidos não afectam significativamente a variância da população⁷ caso contrário H_1 : estes afectam significativamente a variabilidade final⁸.

Distribuição qui-quadrado

Neste caso o valor teste a calcular pode assumir uma das três expressões seguintes, que estão relacionados com a distribuição qui-quadrado e que facultam diferentes tipos de abordagens,

$$TV = \begin{cases} \sum_i^N z_i^2 = \sum_i^N \left(\frac{x_i - \mu}{\sigma}\right)^2 & \leftrightarrow \text{normalidade (qualidade do ajuste)} \\ \frac{(n-1) \times s^2}{\sigma^2} & \leftrightarrow \text{intervalos de confiança (variância)} \\ \sum_i^N \frac{(O_i - E_i)^2}{E_i} & \leftrightarrow \text{comparação directa (frequências)} \end{cases} \quad (2.9)$$

sendo a primeira opção utilizada na avaliação da normalidade das distribuições.

⁷Se se tratar de objectos, estes pertencem à distribuição (não são outliers), se se tratar de parâmetros do modelo, estes são supérfluos (não melhoram significativamente a qualidade do ajuste).

⁸No caso de objectos, estes podem ser outliers uma vez que a sua remoção diminui drasticamente a variabilidade final; no caso de parâmetros, estes são considerados importantes dado que melhoram significativamente a qualidade do ajuste.

2.4.2.4 Outros testes

Há diversos testes que não se baseiam nas distribuições estatísticas anteriores mas nem por isso deixam de ser relevantes para a avaliação estatística e tomada de decisão ao nível do controlo da qualidade.

Teste de Grubbs

Dado que as estimativas paramétricas são sensíveis a valores “contaminados”, estes valores “outliers” produzem em geral erros de estimativa quer na posição (enviesamento) quer na dispersão (inflação da imprecisão) afectando deste modo tanto a exactidão como a dispersão.

O teste de Grubbs [37, 38] é recomendado pela IUPAC, ISO 17025 e ASTM E-178. Um determinado valor pode ser considerado discrepante se a sua distância à estimativa central exceder um valor múltiplo do desvio padrão global. Como H_0 assume-se que o valor em causa ($x_?$) não difere significativamente da estimativa central. O valor a calcular é,

$$G = \frac{\max |x_i - \bar{x}|}{s_x} = \frac{|x_? - \bar{x}|}{s_x} \quad (2.10)$$

em que \bar{x} é a média e s_x o desvio padrão dos N valores experimentais (incluindo o valor em causa, $x_?$).

Este valor só é excluído se exceder o valor crítico tabelado referente a 95% de confiança e para n valores experimentais, tabela A.1.

Teste de simetria e curtose

Uma vez que a estatística paramétrica convencional está baseada na presunção de que os valores se aproximam de uma distribuição normal, é de todo conveniente avaliar esta hipótese já que pode comprometer a representatividade da estimativa em termos de posição do valor central (mais provável) bem como do respectivo intervalo de valores (assimetria da distribuição).

O perfil da distribuição normal é testado calculando o valor

$$TV(SC) = \left(\frac{a_3}{\sigma(a_3)} \right)^2 + \left(\frac{a_4}{\sigma(a_4)} \right)^2 \quad (2.11)$$

onde a_3 e a_4 representam as estimativas de simetria e de curtose da distribuição e

$$\sigma^2(a_3) \simeq \frac{6(N-2)}{(N+1)(N+3)} \quad (2.12)$$

$$\sigma^2(a_4) \simeq \frac{4N(N-3) \times \sigma^2(a_3)}{(N+1)(N+5)} \quad (2.13)$$

representam as respectivas variâncias.

A amostra em causa terá um perfil normal se o valor obtido ($TV(SC)$) não exceder o valor crítico previsto pela distribuição qui-quadrado ($\chi_{\alpha(2)}^2$), com dois graus de liberdade, referente ao nível de confiança de $100(1 - \alpha)$.

2.4.3 Análise de interacção

A covariância e a correlação são estimativas de interdependência directa entre variáveis, duas-a-duas.

Se assumirmos que se pretende avaliar a interdependência directa entre duas variáveis x e y , a covariância é obtida através de

$$CoV[x, y] = \sigma_{xy}^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{(N - 2)} \quad (2.14)$$

em que \bar{x} e \bar{y} representam as estimativas centrais dos vectores x e y que contêm N elementos.

Esta estimativa de associação pode apresentar valores negativos (interdependência decrescente) ou positivos (interdependência crescente), ver figura (2.6)

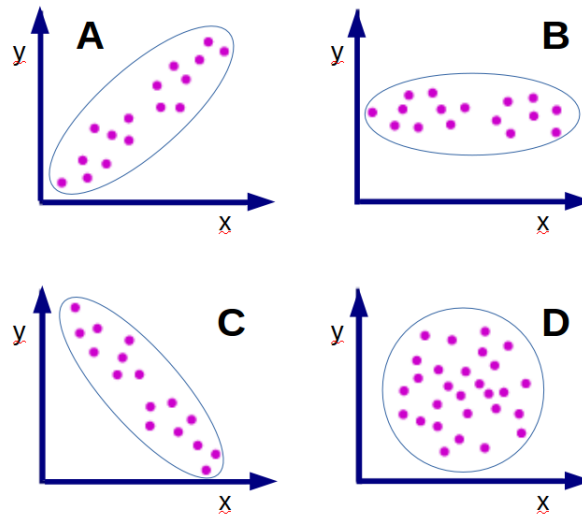


Figura 2.6: Avaliação da interdependência entre as variáveis x e y através da representação mútua assinalando a azul a área da respectiva influência. Em (A) e (C) ambas variáveis apresentam boa interdependência embora no segundo caso esta seja decrescente (covariância negativa, $CoV[x, y] < 0$); já em (B) e (D) a interdependência é considerada nula uma vez que não existe uma relação unívoca entre as variáveis representadas sendo o valor de covariância próximo de zero.

Uma vez que o contradomínio da função covariância se situa em todo o conjunto R ($CoV[x, y] \in] - \infty, +\infty[$), este estimador não permite facilmente evidenciar o grau de interdependência entre variáveis.

Já a correlação difere da covariância neste ponto - sendo ela uma medida de associação relativa que corresponde a uma covariância escalada com a dispersão das variáveis em estudo (σ_x e σ_y)

$$Corr[x, y] = \rho_{xy} = \frac{\sigma_{xy}^2}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2.15)$$

Assim sendo, os valores de correlação situam-se no intervalo $\rho \in [-1, +1]$ o que facilita a sua interpretação em termos de interdependência, ver figura ..

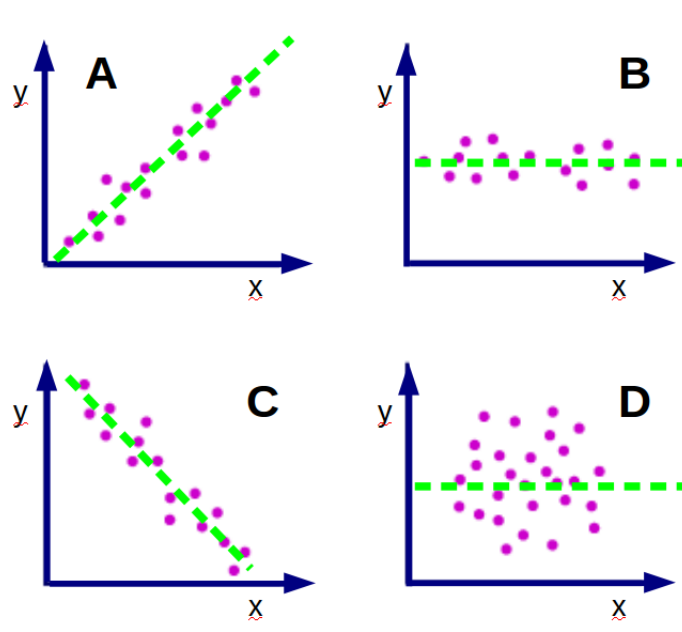


Figura 2.7: Avaliação da interdependência entre as variáveis x e y através da representação mútua com indicação da linha de tendência da interacção. Em (A) e (C) ambas variáveis apresentam boa interdependência com $\rho \simeq +1$ e $\rho \simeq -1$, dando conta da forte interdependência crescente e decrescente, respectivamente; já em (B) e (D) a interdependência é considerada nula ($\rho \simeq 0$) uma vez que não existe uma relação bem definida entre as variáveis.

Sendo a correlação uma medida de interacção normalizada esta pode facilmente ser testada por via estatística.

Há dois tipos de abordagens possíveis - procurar demonstrar que *a)* não existe correlação ($H_0 : |\rho| = 0$ e $H_1 : |\rho| \neq 0$) ou *b)* a correlação é muito significativa ($H_0 : |\rho| = 1$ e $H_1 : |\rho| \neq 1$).

No primeiro caso o valor a calcular será

$$TV = \frac{|\rho| \cdot \sqrt{N-2}}{\sqrt{1-\rho^2}} \quad (2.16)$$

enquanto que para o segundo caso o valor de teste será dado por

$$TV = \frac{|\rho - 1| \cdot \sqrt{N-2}}{\sqrt{1-\rho^2}} \quad (2.17)$$

A correlação é por isso um meio de estimar a interdependência mútua de duas variáveis. Contudo, em casos multivariados, em que a dimensão do sistema atinge valores muito altos ($M \gg 10$), não é praticável utilizar a equação (2.15) para estimar esse parâmetro sendo preferível optar por uma abordagem matricial.

Se assumirmos que se pretende analisar a interdependência de M variáveis de uma matriz $X_{(N \times M)}$ onde cada elemento dessa matriz (x_{ij}) está representado por um valor situado na linha i (índice do objecto) e coluna j (índice da variável), é necessário fazer uma normalização prévia de cada uma das M variáveis através de

$$z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_j} \quad (2.18)$$

onde $(x_{ij} - \bar{x}_j)$ corresponde à centragem na média (\bar{x}_j) e ao escalamento (s_j) com as respectivas estimativas referentes à variável j .

A matriz normalizada ($Z_{(N \times M)}$) pode agora ser utilizada para estimar a matriz de correlação ($R_{(M \times M)}$) através da operação matricial

$$R_{(M \times M)} = Z_{(M \times N)}^T \cdot Z_{(N \times M)} / (N - 1) \quad (2.19)$$

que corresponderá a uma matriz com diagonal unitária já que a correlação de uma variável consigo mesma dá 1.

2.4.4 Modelação Implícita

A modelação implícita [Soft Modeling] é um tipo de abordagem multidimensional para a análise de sistemas multivariados que envolvem diversos factores que contribuem em simultâneo para diversas respostas do mesmo sistema.

A técnica de Mínimos Quadrados Parciais (PLS⁹) foi desenvolvida na década de 1960 pelo estatístico sueco Herman Wold [39] como uma técnica econométrica. Mais tarde, o seu filho, Herman Svante, juntamente com outros cientistas defenderam o seu uso na química.

O algoritmo PLS permite efectuar uma modelação multivariada de sistemas através do método de mínimos quadrados parciais. Este tipo de algoritmo, através de “Soft Modeling”¹⁰ explora um sistema através da maximização da correlação entre os predictores (sub-espço das causas) e o sub-espço das respostas.

Neste caso cada objecto é um resultado de ensaio que resulta da imposição M de condições ao sistema (variáveis controladas ou factores) e avaliação de L respostas simultâneas, características dessas condições.

Assim, os valores a processar são compostos por dois sub-espços: o sub-espço dos predictores $X_{(N \times M)}$ (composto por N objectos avaliados através de M variáveis independentes) e o sub-espço das respostas $Y_{(N \times L)}$ (que avalia os mesmos N objectos através de L variáveis dependentes que caracterizam a resposta do sistema).

Este algoritmo procura inicialmente analisar cada sub-espço, predictor ($X_{(N \times M)}$) e resposta ($Y_{(N \times L)}$), no sentido de encontrar combinações lineares de variáveis, designadas de factores latentes, que se encontram correlacionadas e permitem descrever ao máximo a variabilidade dos resultados de cada sub-grupo de valores.

De seguida vamos rever, de um modo simples, os passos fundamentais do algoritmo PLS.

⁹Do inglês, PLS significa Partial Least Squares ou Projection on Latent Structure.

¹⁰“Soft Modeling” significa efectuar uma modelação implícita não imposta pelo operador.

O conjunto inicial de dados consiste numa tabela de valores onde cada ensaio (objecto) é avaliado sob um conjunto de variáveis ($M + L$). Este conjunto de valores constitui dois sub-espacos - o dos predictores, $X_{(N \times M)}$

$$X_{(N \times M)} = \left\{ \begin{array}{ccc} x_{11} & \cdots & x_{1m} \\ \vdots & x_{ij} & \vdots \\ x_{n1} & \cdots & x_{nm} \end{array} \right\} \quad (2.20)$$

e as respostas, $Y_{(N \times L)}$

$$Y_{(N \times L)} = \left\{ \begin{array}{ccc} y_{11} & \cdots & y_{1l} \\ \vdots & y_{ij} & \vdots \\ y_{n1} & \cdots & y_{nl} \end{array} \right\} \quad (2.21)$$

sendo N o número de objectos, M o número de variáveis do sub-espaco predictor e L o número de variáveis do sub-espaco resposta.

O sub-espaco dos predictores apresenta os objectos representados sob o hiperespaco constituído pelas variáveis independentes que originam (ou se pensa originar) as respostas do sistema. Estes são os valores de entrada no sistema.

Já o sub-espaco das respostas apresenta os mesmos objectos representados sobre as respectivas variáveis dependentes que resultam das condições impostas ao sistema em estudo. Estas variáveis dependem, através de uma função previamente desconhecida, dos estímulos fornecidos ao sistema.

Chama-se de factor latente (FL) a combinação linear de variáveis que consegue descrever uma fracção da variabilidade de cada sub-espaco de valores.

O algoritmo PLS permite representar os sub-espacos dos predictores ($X_{(N \times M)}$) e respostas ($Y_{(N \times L)}$) através de uma decomposição matricial ortogonalizada¹¹ com f factores latentes, ver figura (2.8).

¹¹Representação da informação num novo espaco em que as suas componentes são mutuamente ortogonais (independentes).

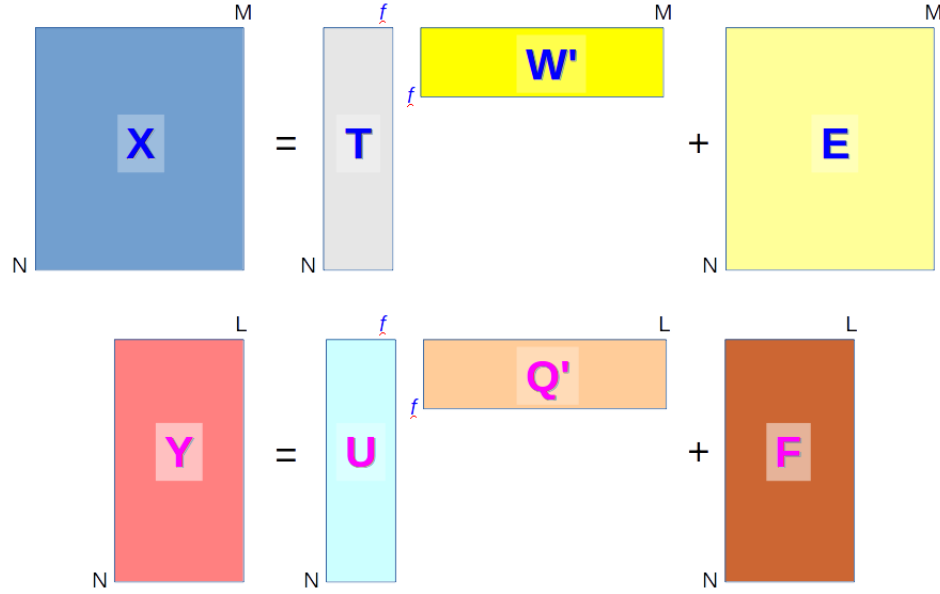


Figura 2.8: Decomposição ortogonalizada dos sub-espços predictor ($X_{(N \times M)}$) e resposta ($Y_{(N \times L)}$) sob a forma de produto matricial com dimensão de f factores latentes sendo $E_{(N \times M)}$ e $F_{(N \times L)}$ as matrizes residuais que contêm a informação que não foi utilizada na modelação PLS. As matrizes $T_{(N \times f)}$ e $U_{(N \times f)}$ correspondem à representação dos N objectos em f factores latentes (scores) enquanto que as matrizes $W_{(M \times f)}$ e $Q_{(L \times f)}$ expressam o impacto das M e L variáveis de cada sub-espço sobre os f factores latentes (loads).

A figura anterior expressa a decomposição da informação contida em cada sub-espço numa combinação linear de “scores” e “loads”.

O sub-espço dos predictores ($X_{(N \times M)}$) que contém a informação de N objectos representada sobre M variáveis pode ser representado sob a forma de um produto matricial

$$\hat{X}_{(N \times M)} = T_{(N \times f)} \cdot W'_{(f \times M)} \quad (2.22)$$

onde $W_{(M \times f)}$ representa a nova base ortonormada de dimensão f (loads) e $T_{(N \times f)}$ contém a informação relativa à representação dos N objectos (scores) no sub-espço dos f factores latentes previamente considerados ao estabelecer a representação de $X_{(N \times M)}$ sob a forma de $\hat{X}_{(N \times M)}$.

A informação residual de X não considerada na representação com f factores latentes encontra-se na matriz

$$E_{(N \times M)} = X_{(N \times M)} - T_{(N \times f)} \cdot W'_{(f \times M)} \quad (2.23)$$

De modo similar, considerando o mesmo número de f factores latentes, a informação contida no sub-espço da resposta, $Y_{(N \times L)}$, pode ser representada sob a forma

$$\hat{Y}_{(N \times L)} = U_{(N \times f)} \cdot Q'_{(f \times L)} \quad (2.24)$$

sendo a respectiva informação residual dada por

$$F_{(N \times L)} = Y_{(N \times L)} - U_{(N \times f)} \cdot Q'_{(f \times L)} \quad (2.25)$$

As relações estabelecidas pelas equações (2.22) a (2.25) expressam as relações externas [outer relations] de PLS, ver figura

O diagrama apresenta duas equações de decomposição PLS, cada uma dentro de um retângulo colorido. A primeira equação, dentro de um retângulo azul, mostra a decomposição da matriz de preditores X (dimensão $N \times M$) em um produto da matriz de escores T (dimensão $N \times f$) multiplicada pela matriz de pesos W' (dimensão $f \times M$), mais a matriz de erro E (dimensão $N \times M$). A segunda equação, dentro de um retângulo magenta, mostra a decomposição da matriz de respostas Y (dimensão $N \times L$) em um produto da matriz de escores U (dimensão $N \times f$) multiplicada pela matriz de pesos Q' (dimensão $f \times L$), mais a matriz de erro F (dimensão $N \times L$). As dimensões das matrizes são indicadas por letras minúsculas em cantos das matrizes.

$$X = T * W' + E$$

$$Y = U * Q' + F$$

Figura 2.9: Decomposição PLS dos sub-espaços dos preditores ($X_{(N \times M)}$) e respostas ($Y_{(N \times L)}$) com base nas relações externas.

O algoritmo passa essencialmente por sete fases distintas: verificação inicial, pré-acondicionamento, estimativa das matrizes de covariância combinada (de X com Y e de Y com X), estimativa da variabilidade (SS_0), estimativa do factor latente, remoção do efeito do factor latente (nos sub espaços X e Y), verificação da convergência (SS_i).

1. Verificação inicial

Sendo este tipo de análise de factores efectuada através de operações matriciais, é necessário garantir previamente que não vão surgir singularidades de cálculo com valores mal acondicionados nem vão existir incompatibilidades nessas operações matriciais. Assim é necessário verificar, objecto a objecto, se todos os valores foram correctamente inseridos sobre as respectivas variáveis e se as linhas estão completas.

2. Pré-acondicionamento

Na modelação PLS assume-se que os modelos são lineares em relação aos factores latentes - apenas é assumido o primeiro grau. Assim, o termo constante previsto para um modelo do tipo polinomial, deve ser suprimido. Esta simplificação do modelo é conseguida através da centragem de todas as variáveis.

Por outro lado, uma segunda dificuldade experimental pode inviabilizar este tipo de tratamento. No caso de as variáveis apresentarem escalas muito diferentes, a sua influência sobre os parâmetros estimados do modelo seria nítida

obscurecendo o efeito das restantes variáveis. Assim, é usual efectuar uma normalização prévia das variáveis.

Este escalamento prévio torna possível verificar o efeito de todas as variáveis, independentemente da sua variabilidade relativa. Para evitar singularidades no auto-escalamento das variáveis, é necessário garantir que cada variável não é constante.

3. Covariância combinada

Sendo necessário maximizar a correlação entre os sub-espacos X e Y , calculam-se as matrizes de covariância de $X_{(N \times M)}$ em $Y_{(N \times L)}$ (Cxy)

$$(N-1).Cxy_{(M \times M)} = X'_{(M \times N)} \cdot Y_{(N \times L)} \cdot Y'_{(L \times N)} \cdot X_{(N \times M)} = |Y' \cdot X|^2 \quad (2.26)$$

e de covariância de $Y_{(N \times L)}$ em $X_{(N \times M)}$ (Cyx)

$$(N-1).Cyx_{(L \times L)} = Y'_{(L \times N)} \cdot X_{(N \times M)} \cdot X'_{(M \times N)} \cdot Y_{(N \times L)} = |X' \cdot Y|^2 \quad (2.27)$$

4. Variabilidade do sistema

Chama-se de traço da matriz ao produto dos termos da diagonal principal ($Trace(C)$). É forçoso que o valor do traço obtido em ambas matrizes de covariância, Cxy e Cyx , seja igual. Este valor corresponde à variância global do sistema.

Em qualquer instante do processo iterativo, a soma de quadrados de cada sub-espaco pode ser estimada com base em

$$SSx = \sum_i^N \sum_j^M x_{ij}^2 \quad (2.28)$$

$$SSy = \sum_i^N \sum_k^L y_{ik}^2 \quad (2.29)$$

5. Estimativa dos factores latentes

Chama-se de Factor Latente ao vector de Loads $w_f(m \times 1)$ e $q_f(m \times 1)$

1. Ao processar a matriz de Covariância de X em Y (Cxy) através da decomposição em valores próprios, as variáveis independentes são decompostas nas matrizes dos Scores ($T_{(N \times K)}$) e Loads ($W_{(K \times M)}$),

$$X_{(N \times M)} = T_{(N \times K)} \cdot W'_{(K \times M)} \quad (2.30)$$

sendo as colunas da matriz dos Loads ($W_{(M \times K)}$) as componentes dos K factores latentes (vectores latentes) no sub-espaco dos M predictores.

2. Ao processar a matriz de Covariância de Y em X (Cyx) as variáveis dependentes são decompostas nas matrizes dos Scores ($U_{(N \times K)}$) e Loads ($Q_{(K \times M)}$),

$$Y_{(N \times M)} = U_{(N \times K)} \cdot Q'_{(K \times M)} \quad (2.31)$$

onde as colunas da matriz dos Loads ($Q_{(M \times K)}$) são as componentes dos K factores latentes (vectores latentes) no sub-espaço das M respostas.

Dado que o processo é iterativo, as equações (2.30) e (2.31) devem antes ser escritas sob a forma,

$$\hat{X}_{(N \times M)} = t_{(N \times 1)} \cdot w'_{(1 \times M)} \quad (2.32)$$

$$\hat{Y}_{(N \times M)} = u_{(N \times 1)} \cdot q'_{(1 \times M)} \quad (2.33)$$

em que \hat{X} e \hat{Y} representam os valores estimados para os predictores e para as respostas.

3. No espaço dos factores latentes (f), a representação dos N objectos do sub-espaço predictor faz-se através dos respectivos $T_{(N \times f)}$ scores

$$T_{(N \times f)} = X_{(N \times f)} \cdot W_{(f \times M)} \iff t_{(N \times 1)} = X_{(N \times M)} \cdot w_{(M \times 1)} \quad (2.34)$$

4. No espaço dos mesmos f factores latentes, a representação dos N objectos do sub-espaço resposta faz-se através dos respectivos $U_{(M \times f)}$ scores

$$U_{(N \times f)} = Y_{(N \times f)} \cdot Q_{(f \times L)} \iff u_{(N \times 1)} = Y_{(N \times L)} \cdot q_{(L \times 1)} \quad (2.35)$$

5. A modelação entre os scores $T_{(N \times f)}$ e $U_{(M \times f)}$ faz-se através de f combinações lineares sucessivas

$$U_{(N \times f)} = T_{(N \times f)} \cdot B_{(f \times f)} \iff u_{(N \times 1)} = t_{(N \times 1)} \cdot b_{(1 \times 1)} \quad (2.36)$$

onde a matriz $B_{(f \times f)}$ é uma matriz diagonal que contém os f parâmetros do modelo (b_k).

6. Atendendo à equação (2.36) o parâmetro do modelo é estimado por

$$b_{(1 \times 1)} = t'_{(1 \times N)} \cdot u_{(N \times 1)} \quad (2.37)$$

As equações (2.34) a (2.37) dão conta da relação interna entre os objectos representados ao nível dos predictores e os objectos representados ao nível das respostas, ver figura (2.10).

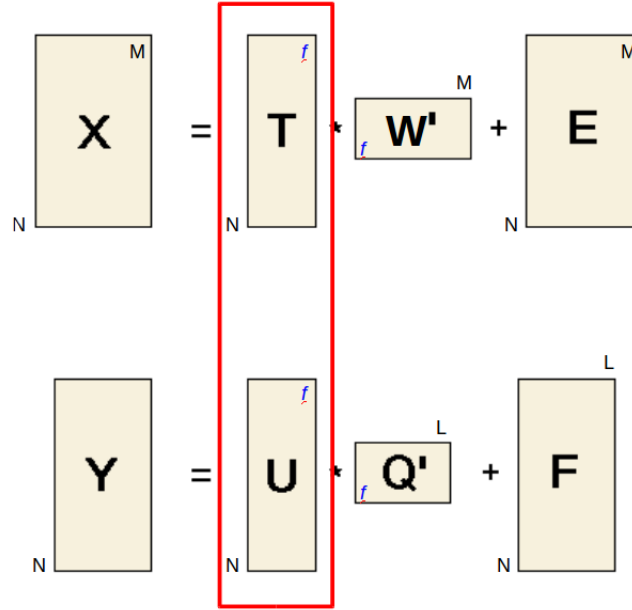


Figura 2.10: Representação da relação interna entre a informação contida nos objectos do espaço predictor e os objectos do espaço resposta.

6. Efeito do factor latente

Atendendo à equação 2.32, o efeito do factor latente no espaço dos Loads é dado por

$$p_{(M \times 1)} = X'_{(M \times N)} \cdot t_{(N \times 1)} \quad (2.38)$$

sendo os valores previstos por este factor ao nível do sub espaço dos predictores dado por

$$\hat{X}_{(N \times M)} = t_{(N \times M)} \cdot p'_{(1 \times M)} \quad (2.39)$$

O efeito do factor latente ao nível do sub-espaço das respostas é dado por

$$\hat{Y}_{(N \times L)} = b_{(1 \times 1)} \cdot t_{(1 \times M)} \cdot q'_{(M \times L)} \quad (2.40)$$

7. Remoção do efeito do factor

Em cada passo da iteração os factores latentes (mais relevantes) vão sendo sucessivamente removidos dos sub-espaços respectivos

$$E_{(N \times M)} = X_{(N \times M)} - \hat{X}_{(N \times M)} \quad (2.41)$$

$$F_{(N \times L)} = Y_{(N \times L)} - \hat{Y}_{(N \times L)} \quad (2.42)$$

removendo a informação pertinente que relaciona os dois sub-espaços e as matrizes diferença vão tendendo para matrizes nulas (erro puramente aleatório).

Há diversos pormenores que convém atender em PLS:

1. Sendo as variáveis previamente centradas e escaladas, a soma de quadrados de X e de Y é dada pela soma dos termos de X e Y ao quadrado já que estas matrizes se encontram normalizadas.

2. O Primeiro valor próprio de $Cxy_{(M \times M)}$, eq.(2.26), surge do valor próprio mais significativo, com maior valor assim como o primeiro valor próprio de $Cyx_{(L \times L)}$, eq.(2.27) também será o maior valor próprio dessa matriz de covariância.
3. Estes valores próprios estão intimamente relacionados já que a sua correlação foi maximizada através da covariância cruzada entre os sub-espacos predictor-resposta.
4. Esta decomposição matricial através dos seus valores próprios faz com que surja agora um novo sistema de vectores directores ortonormalizado definido através dos respectivos vectores próprios $(W_{(M \times f)})$ e $(Q_{(L \times f)})$.
5. Para um determinado factor latente f , escurtinado por via decrescente de importância, vai existir uma relação estabelecida entre os sub-espacos X e Y que está explicitamente definida através dos respectivos pesos. Assim, o vector próprio w_f do sub-espaco predictor e q_f do sub espaco resposta são constituídos por componentes, w_{fj} e q_{fk} , que traduzem o impacto das variáveis originais sobre esse factor latente e, deste modo indicam quais as inter-relações entre variáveis de ambos sub-espacos.
6. Se existirem p respostas, é previsível que o número de factores a considerar para descrever esse sistema seja próximo deste valor. Este algoritmo visa essencialmente simplificar o problema multidimensional inicial. Assim, importa estabelecer critérios para recuperar uma fracção significativa da variabilidade inicial do sub-espaco resposta. É portanto necessário estabelecer critérios de aceitação e de rejeição de factores latentes no sentido de estabelecer qual o número mínimo de factores latentes que devem ser considerados para descrever o sistema em causa.
Um dos critérios mais utilizados consiste em tentar reproduzir cerca de 80% da resposta original. Sendo este critério bastante significativo em termos de reprodução do sistema, é geralmente difícil obter índices tão elevados de desempenho com sistemas estocásticos onde as inter-relações das variáveis são obscurecidas através de causas desconhecidas.
Outro critério também utilizado consiste em representar o valor paramétrico obtido na modelação ou a variância incremental descrita por esse factor através de um gráfico de desempenho (scree plot).
Os factores latentes significativos são também reconhecidos por estarem anormalmente acima da contribuição basal, relacionada com a contribuição aleatória. Esta alternativa é, regra geral, bastante robusta para evidenciar os factores latentes mais relevantes.
7. Estes factores latentes eleitos podem agora ser utilizados para evidenciar informação preciosa acerca da interacção entre variáveis do sistema em causa através dos respectivos Loads mais significativos. Para se reconhecer as contribuições mais relevantes estabelecidas por determinado factor latente que cruza os sub-espacos causa-efeito, é necessário efectuar uma inspecção minuciosa aos respectivos loads, catalogar os loads de cada sub-espaco por ordem decrescente de importância e garantir que uma fracção significativa do módulo do vector latente é recuperada. De igual modo têm sido sugeridos desempenhos da ordem de 80% contudo, em sistemas

mais complexos é aceitável obter recuperações da ordem de apenas 50% para evitar o sobre dimensionamento do problema.

2.4.5 Modelação Explícita

A modelação explícita consiste em representar os valores experimentais (variável dependente) através de uma função matemática pré definida.

No caso de se tratar de um sistema multivariado, a função do modelo relaciona diversas variáveis independentes com a respectiva variável dependente¹².

2.4.5.1 Pressupostos

Para que as estimativas paramétricas conduzam a estimativas correctas tem que se garantir que *a*) os valores experimentais estão isentos de erro sistemático, *b*) apenas a variável dependente (que se pretende modelar) apresenta erro aleatório aditivo com distribuição normal e independente, *c*) que se conhece previamente o modelo a utilizar e que *d*) não existem valores discrepantes (outliers).

2.4.5.2 Tipos de estimativas

As estimativas estatísticas de posição e dispersão podem ser classificadas em estimativas não paramétricas e estimativas paramétricas.

As estimativas não paramétricas geralmente recorrem a operações simples tais como a ordenação dos valores e avaliação dos percentis. São exemplos destas estimativas a mediana e os intervalos de confiança com base em percentis¹³. Regra geral, estes estimadores são robustos dado serem muito insensíveis a valores discrepantes o que os torna satisfatórios para servirem com indicadores de valores discrepantes. Estes estimadores são tanto mais correctos quanto maior a dimensão da amostra de valores considerados. Contudo este tipo de abordagem pode-se tornar algo complicado como por exemplo a estimativa da imprecisão dos parâmetros da calibração só pode ser obtida através de técnicas de reamostragem ou de simulação.

As estimativas paramétricas em geral estão relacionadas com métodos de modelação sendo mais frequentemente utilizados os métodos de ajuste por mínimos quadrados. Estes métodos procedem à estimativa paramétrica com base na minimização da função de erro do modelo¹⁴.

Se o modelo utilizado for uma função linear¹⁵, como por exemplo os polinómios, e os pressupostos forem válidos, este método de estimativa apresenta valores correctos para as estimativas paramétricas incluindo a sua imprecisão.

Dado que estes estimadores baseiam-se na minimização do erro do modelo em relação aos valores experimentais, os métodos paramétricos são vulneráveis a valores discrepantes (outliers).

¹²No caso de uma calibração a variável independente é uma concentração e a variável dependente o respectivo sinal instrumental.

¹³A estimativa mediana baseia-se na ordenação dos valores e estimativa central com base no valor percentil 50%.

¹⁴Esta função de erro é muitas vezes expressa sob a forma da função qui-quadrado e geralmente designa-se de função de mérito do ajuste.

¹⁵Entenda-se por função linear toda a função cujas derivadas parciais em relação aos seus parâmetros são independentes destes.

Vamos admitir que os n valores experimentais y_i podem ser descritos através do modelo $\eta_i(\beta)$ que depende de p parâmetros (β).

O erro do modelo no ponto i (e_i) é dado pela diferença do modelo ao respectivo valor experimental,

$$e_i = y_i - \eta_i(\beta)$$

A variância do ajuste (σ_{fit}^2) traduz a variância do erro do modelo,

$$\sigma_{fit}^2 = \frac{\sum_{i=1}^n f_i \cdot (y_i - \eta_i(\beta))^2}{(\sum f_i - p)} = \frac{SS}{(n - p)} \quad (2.43)$$

A variância do ajuste é uma grandeza estritamente crescente e dá origem a uma parábola multidimensional ($dim = p$) em que o melhor conjunto de parâmetros do modelo correspondem ao mínimo global desta hipersuperfície.

Dado que o número de graus de liberdade é constante ($n - p$), a soma de quadrados (SS) é frequentemente utilizada como figura de mérito do ajuste. Deste modo, o conjunto de parâmetros do modelo que minimiza a soma de quadrados constitui a melhor estimativa paramétrica para esse modelo.

Sendo o valor mínimo da hipersuperfície de erro (SS) dado pelo ponto onde as diversas derivadas parciais se anulam simultaneamente,

$$\frac{\partial SS}{\partial \beta_j} = 0 \quad , \quad \forall j \in \{1, \dots, p\} \quad (2.44)$$

os p parâmetros do modelo são determinados através da resolução simultânea de um sistema de p equações normais para garantir que se trata do mínimo global da função de erro do ajuste (σ_{fit}^2).

Modelos lineares

No caso de o gradiente paramétrico, equação (2.44), ser uma expressão que não depende de qualquer dos parâmetros do modelo, o modelo diz-se linear e a estimativa paramétrica faz-se directamente usando essa expressão.

Vejam os por exemplo um modelo polinomial do tipo

$$\eta_i = b_0 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} \quad (2.45)$$

as derivadas parciais em ordem aos parâmetros b_0 , b_1 e b_2 dão 1, x_{1i} , x_{2i} , respectivamente, não sendo nenhuma delas dependente dos parâmetros sendo por isto apelidado de modelo linear.

Notação matricial

Os n valores experimentais, de coordenadas x_i (var. independente) e y_i (var. dependente) são representados sob a forma de vectores coluna ($n \times 1$).

No caso de modelos lineares, a variável dependente vai ser descrita através de um modelo contendo p parâmetros sendo o sistema de equações representado sobre a forma de produto matricial,

$$X_{(n \times p)} \times B_{(p \times 1)} = Y_{(n \times 1)} \quad (2.46)$$

onde B representa o vector coluna que contém os p parâmetros, Y o vector coluna dos termos independentes e X a matriz referente à variável independente. A resolução desta equação matricial passa por:

a) multiplicar ambos os membros da equação pela matriz transposta ($X_{(p \times n)}^T$),

$$A_{(p \times p)} \times B_{(p \times 1)} = X_{(p \times n)}^T \times Y_{(n \times 1)} \quad (2.47)$$

$$A_{(p \times p)} = \left(X_{(p \times n)}^T \times X_{(n \times p)} \right) \quad (2.48)$$

sendo A conhecida por matriz de projecto [**Design Matrix**].

b) produto pela matriz de covariância ($C_{(p \times p)} = A_{(p \times p)}^{-1}$)

$$B_{(p \times 1)} = C_{(p \times p)} \times \left(X_{(p \times n)}^T \times Y_{(n \times 1)} \right) \quad (2.49)$$

$$C_{(p \times p)} = \left(X_{(p \times n)}^T \times X_{(n \times p)} \right)^{-1} \quad (2.50)$$

Representatividade

Importa definir previamente o factor de peso de cada valor experimental (w_i) sobre o conjunto global dos pontos descritos pelo modelo. No sentido de obter estimadores eficientes (de máxima precisão) convém utilizar a dispersão dos valores experimentais (s_i) como factores de ponderação. Assim, os pesos estatísticos são construídos através de,

$$w_i = \frac{1}{s_i^2}$$

Por sua vez, estes podem ser convenientemente convertidos em frequências (m_i) no sentido de facilitar as estimativas paramétricas,

$$w_i^* = w_i \times \frac{N}{\sum_{i=1}^n w_i} = w_i \times \frac{N}{W} = \frac{w_i}{\bar{W}}$$

Estes factores de frequência ($w_i^* \approx f_i$) devem ser tomados em consideração na expressão da soma de quadrados (SS), eq. (2.43).

Deste modo, a estimativa obtida corresponde ao método ponderado também conhecido como “Weighted Least Squares” (WLS).

No caso WLS a matriz de projecto e de covariância paramétrica assumem a forma,

$$A_{w(p \times p)} = X_{(p \times n)}^T \times W_{(n \times n)} \times X_{(n \times p)} \quad (2.51)$$

$$C_{w(p \times p)} = \left(X_{(p \times n)}^T \times W_{(n \times n)} \times X_{(n \times p)} \right)^{-1} \quad (2.52)$$

onde $W_{(n \times n)}$ representa a matriz “peso estatístico” que contabiliza o erro experimental associado à variável dependente¹⁶. A matriz $W_{(n \times n)}$ contém informação referente à covariância do erro experimental. Se o erro for não correlacionado esta matriz apresenta-se sob a forma de matriz diagonal principal.

No caso de os valores experimentais serem equiprováveis (caso homocedástico) o estimador da variância do ajuste vem simplificado,

$$\sigma_{fit}^2 = \frac{\sum_{i=1}^n (y_i - \eta_i(\Theta))^2}{(n - p)} = \frac{SS}{(n - p)} \quad (2.53)$$

e corresponde à aproximação não ponderada, também conhecida como “Ordinary Least Squares” (OLS).

¹⁶Corresponde à matriz inversa da covariância da variável dependente.

Estimativa da dispersão

A imprecisão dos parâmetros estimados por mínimos quadrados ponderados (WLS) pode ser obtida directamente através da matriz de covariância paramétrica ($C_{w(p \times p)}$), eq. (2.52) em que os elementos da diagonal principal (c_{ii}) estão relacionados com a variância do respectivo parâmetro ($\sigma_{\theta_i}^2$) e os restantes elementos (c_{ij}) relacionados com as covariâncias inter-paramétricas ($\sigma_{\theta_i \theta_j}^2$),

$$\sigma_{\theta_i} = \sqrt{c_{ii}} \quad \sigma_{\theta_i \theta_j} = \sqrt{c_{ij}} \quad (2.54)$$

No caso de uma estimativa não ponderada (OLS), a dispersão prevista para o parâmetro é calculada com base no desvio padrão do ajuste (σ_{fit} , equação (2.43)),

$$\sigma(\beta_j) = \sqrt{c_{jj}} \times \sigma_{fit} \quad (2.55)$$

$$\sigma(\beta_j, \beta_k) = \sqrt{c_{jk}} \times \sigma_{fit} \quad (2.56)$$

onde c_{jj} e c_{jk} são os elementos da diagonal principal e extra diagonal principal, da matriz de covariância paramétrica, na página anterior, respectivamente.

As estimativas obtidas por mínimos quadrados correspondem a estimativas “centrais” e a respectiva estimativa de dispersão refere-se à imprecisão da estimativa central correspondendo a um erro padrão.

Para se ter a noção do grau de incerteza associado à estimativa paramétrica respectiva é útil calcular a precisão relativa sob a forma de erro padrão relativo ($\%RSE$)

$$\%RSE = 100 \times \frac{\sigma(\beta_i)}{\beta_i} \quad (2.57)$$

usando os valores estimados em (2.55) e (2.49).

Capítulo 3

Parte experimental

O desenvolvimento deste trabalho baseou-se em averiguar a quantidade de resíduos gerados nas operações de produção e embalagem, utilizando os métodos de PLS e MOLS com o intuito de analisar os parâmetros que influenciam a produção de resíduos a fim de os minimizar. Foi também utilizada a Ferramenta do Diagrama de Pareto com o objectivo de analisar graficamente os grupos de resíduos com maior impacto, ou seja, os que eram gerados em maior quantidade.

A seguir, descrevem-se as metodologias utilizadas para desenvolver este tópico.

3.1 Descrição das operações de produção e seus resíduos

De seguida são descritas as operações realizadas no processamento de um lote de medicamentos e os resíduos gerados em cada uma delas.

3.1.1 Pesagem da matéria-prima

Considera-se esta etapa como sendo uma das operações mais importantes do processo produtivo.

A produção de comprimidos e cápsulas inicia-se com a pesagem e conferência das matérias-primas que são utilizadas para formar o produto desejado, etapa da qual resultam já resíduos, provenientes dos materiais de acondicionamento.

A fim de evitar problemas de contaminação cruzada, os operadores deste processo vestem uniformes que são colocados e retirados à entrada da “zona limpa”. Uma matéria-prima entra na sala quando a anterior é retirada, tal como, os operadores retiram as luvas quando terminam de manusear uma matéria-prima e colocam novo par assim que iniciam a pesagem de outra. É de salientar que a caixa das luvas está num local fechado de modo a que não fiquem contaminadas com o pó gerado nesta operação.

Após as pesagens da matéria-prima advém a necessidade de pulverizar, ou pelo menos, tamisar todos os componentes que irão dar origem aos comprimidos sólidos.

De seguida, segue-se a granulação.

Os comprimidos podem ser obtidos por três vias diferentes: granulação húmida, granulação seca ou compressão directa.

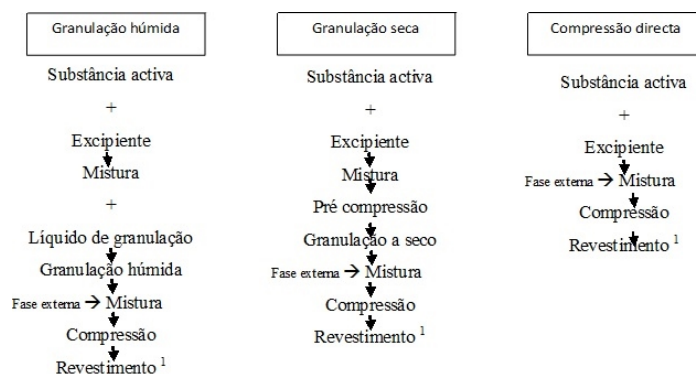


Figura 3.1: Diagrama representativo das diferentes fases processuais na produção de comprimidos por granulação seca, granulação húmida e compressão directa.

¹ A fase do revestimento não se aplica a todos os produtos.

3.1.2 Processo de granulação

A granulação tem como finalidade modificar as características do complexo farmacêutico, transformando o pó em agregados maiores, sólidos, uniformes e mais resistentes. Pretende-se a consolidação dos grânulos, aumentar o contacto entre as partículas e aumentar a área de superfície para facilitar a secagem. São utilizados dois tipos de métodos de granulação: granulação húmida e granulação seca.

3.1.2.1 Granulação húmida

A granulação húmida inicia-se com a adição de um fluido de granulação, geralmente água, que funciona como líquido de granulação na mistura de pós, convertendo a massa húmida em agregados maiores. À medida que se adiciona o líquido, as partículas do pó aproximam-se, formando pontes líquidas e diminuem os espaços vazios dentro dos grânulos, aumentando a resistência entre as partículas. Estas ligações são conseguidas devido à tensão superficial e à pressão capilar, que originam as forças responsáveis pela dureza e pela formação inicial de grânulos. No entanto, se a adição de líquido for superior à estipulada, as partículas permanecem juntas devido à tensão superficial, mas originam grânulos pouco resistentes devido às forças intragranulares fracas. De seguida, a massa obtida é passada por um crivo, tendo como finalidade a obtenção de partículas com a mesma dimensão. Esta passa depois pelo processo de secagem para remover a maior parte da humidade do produto. É desejável que o granulado possua uma quantidade residual de humidade para otimizar a força entre as partículas. Durante esta etapa há formação de pontes sólidas do grânulo, o que se deve à cristalização dos materiais dissolvidos, à fusão das pontes nas áreas de contacto entre as partículas, assim como à presença de agentes aglutinantes. Nesta fase do processo, o granulado é introduzido num depósito onde o ar aspirado do exterior é aquecido à temperatura requerida. Este circula dentro do

depósito de baixo para cima, tendo como finalidade “sacudir” o granulado. No extremo superior do depósito, existe uma válvula que mede o ar de exaustão. No meio do depósito existem ainda umas mantas de filtro que impedem a saída das partículas a secar. Depois da secagem as partículas ficam ligadas entre si por forças electrostáticas.

Uma vez seco, o granulado é obrigado a passar por outro crivo com um determinado diâmetro, com o objectivo de fragmentar alguns grãos que se tenham aglomerado. O granulado que atravessou esse crivo é aproveitado, granulando-se novamente o grão ao tamanho desejado. Depois deste processo, as partículas ficam ligadas entre si por forças moleculares.

Grânulos com resistência abaixo da resistência inferior tendem a quebrar na fase da mistura, manuseamento e pré-compressão, impossibilitando um enchimento uniforme da matriz na compressão; grânulos duros mantêm a sua estrutura quando submetidos a forças elevadas, mas a sua consolidação pode ser difícil, originando comprimidos pouco resistentes.

Os resíduos gerados na granulação húmida são: as amostras do granulado após a realização dos testes para determinação da humidade e do teor do produto, a amostra que é enviada para o laboratório de Controlo de Qualidade para a realização de testes, o pó retirado dos crivos e o possível granulado derramado[40, 41].

3.1.2.2 Granulação a seco

Como o nome indica, esta etapa é realizada a seco.

Numa primeira fase, a partir da mistura dos pós, obtêm-se comprimidos irregulares e com peso diferente. De seguida, com a ajuda de crivos, os comprimidos irregulares são fragmentados, originando granulados também eles irregulares que vão ser usados para a obtenção de comprimidos. Nesta fase é necessário utilizar um lubrificante, tendo como objectivo reduzir o atrito dos pós aos punções, matrizes e peças metálicas. Normalmente, as máquinas utilizadas na pré-compressão possuem punções com maior diâmetro e potência, o que permite a aplicação de forças mais elevadas. Este processo envolve a compactação controlada de pós em máquinas de comprimir, originando comprimidos com peso semelhante, o que o torna mais vantajoso do que a granulação húmida, na medida em que o tempo para preparar comprimidos é menor, pois permite dispensar a secagem, etapa demorada na granulação húmida.

Numa granulação a seco, são particularmente as forças Van der Waals as responsáveis pela ligação entre as partículas [22].

3.1.2.3 Compressão directa

Este método de produção tem aplicação em formulações com fármacos directamente compressíveis ou com fármacos potentes em que propriedades de escoamento e compressibilidade advêm do diluente. O diluente é um produto inerte, com boa fluidez e que pode ser compactado facilmente.

Este fenómeno só é conseguido quando os pós possuem densidade suficiente para escoar livremente do distribuidor da máquina e encher regularmente a matriz sem provocar atritos na compressora e se apresentem num estado cristalino tal que permita a fácil coesão dos cristais entre eles. Em geral originam comprimidos regulares, apenas é necessário padronizar o tamanho dos cristais e

submetê-los a uma ligeira secagem.

A compressão directa apresenta algumas vantagens quando comparada com o processo de granulação, principalmente no que se refere às etapas de produção figura(3.1) . No entanto, apresenta limitações, tais como, a selecção de excipientes que não é adequada a todo o tipo de produtos, as diferenças no tamanho das partículas que podem levar a uma não uniformidade de conteúdo do fármaco e o facto de ser limitada a compostos de baixas dosagens.

Os resíduos gerados neste processo são uma pequena quantidade de granulado que é enviada para o laboratório do Controlo de Qualidade, assim como a amostra de resíduos resultantes para a determinação do teor do produto.

Para concluir, seja qual for a via destinada à formação de comprimidos, obtêm-se sempre resíduos de pó.

As técnicas descritas anteriormente recorrem às etapas que serão de seguida apresentadas [22].

3.1.3 Mistura

Nesta etapa, é necessário introduzir um lubrificante a fim de diminuir, na compressão, o atrito dos pós aos punções, matrizes e peças metálicas, assim como facilitar o deslizamento para a matriz.

A mistura de pós visa a melhor distribuição possível das partículas de dois ou mais componentes. Este processo pode ocorrer entre sistemas sólidos ou sistemas sólidos/líquidos e tem como objectivos garantir a homogeneidade do produto final e a preparação para as fases posteriores (compressão e revestimento). As características físicas das partículas, tais como a densidade, o tamanho, a forma, a elasticidade e a rugosidade superficial exercem a sua influência sobre as propriedades dos pós. Da mesma maneira, os aspectos relacionados com o equipamento são importantes e requerem a devida atenção.

A mistura efectua-se normalmente em misturadores, de acordo com a quantidade a preparar. Os misturadores são máquinas capazes de proporcionar misturas eficazes de pós secos ou granulados. O mecanismo de funcionamento ideal é o movimento mais suave e lento possível, pois, desta forma, não se modificam as propriedades das partículas, evitam-se as aglomerações e previne-se a formação de calor. Existem dois tipos de misturadores: recipientes fixos com agitação e recipientes móveis, podendo ter intensificadores de agitação, sendo os misturadores móveis mais utilizados na indústria por possuírem menos potência e por serem de fácil limpeza.

Nesta etapa podem ocorrer dois tipos de forças: 1) forças produzidas pelo movimento translacional e rotacional das partículas resultantes do contacto com as paredes dos misturadores e com outras partículas; 2) forças elevadas de adesão/coesão entre partículas que normalmente são difíceis de misturar devido à aglomeração. Quando tal ocorre, os agregados de partículas podem ser quebrados por misturadores de corte, caso contrário sujeitam o pó a forças de impacto. As forças Van der Waals actuam sobre todas as partículas, levando a que estas se agreguem quando a repulsão electrostática é suficientemente reduzida ao ponto de anular as cargas à superfície das partículas. Nesta etapa os resíduos gerados são os resultantes da determinação do teor do produto e as amostras do granulado que são enviadas para o laboratório de Controlo de Qualidade, o granulado retido nas misturadoras e o possível derrame [42, 43].

3.1.4 Compressão

A compressão traduz-se por uma redução do volume do granel, uma vez que as partículas do pó ou a superfície dos grânulos se aproximam entre si, por acção de forças mecânicas. O aumento da compressão provoca uma deformação nessas partículas, podendo levar à quebra e formação de novas ligações, o que leva ao aumento da força mecânica entre estas, à medida que as suas superfícies são forçadas a unir-se. A principal consequência do aumento do contacto entre as partículas é a diminuição da energia livre de superfície do pó resultante da criação de ligações sólido/sólido. Neste caso, a adesão e/ou coesão das partículas é devida às forças Van der Waals.

Os comprimidos são fabricados através da compressão do pó quando se exerce uma pressão entre dois punções (superior e inferior) no interior de uma matriz.

A quantidade de pó admitida na matriz é regulada pela posição desta e pela posição do punção inferior. As propriedades físicas dos comprimidos dependem da escolha das matérias-primas utilizadas e dos detalhes do processo aplicado ao equipamento de compressão.

As máquinas de comprimir são constituídas por: tremonha, alimentada pelo contentor que contém o granulado; matrizes, que regulam a forma e o tamanho do comprimido; punções, para comprimir o granulado dentro da matriz; calhas, que orientam o movimento dos punções; e um mecanismo de alimentação que conduz o granulado da tremonha para a matriz.

Deste modo, como se descreve a seguir, as máquinas de compressão podem ser classificadas de compressora rotativa ou compressora excêntrica [44].

3.1.4.1 Compressora rotativa

Este tipo de compressoras é utilizado na produção de grandes lotes. Neste caso, o mecanismo de alimentação de pó é constituído por uma tremonha ligada à estrutura de alimentação. A estrutura de alimentação geralmente é composta por um certo número de rodas de pás que transferem o pó da saída da tremonha para a matriz. O pó flui para a matriz sob o efeito da gravidade e é auxiliado pela roda de pás. O comportamento do escoamento depende do tamanho e forma das partículas, bem como das condições ambientais. Nesta etapa, os punções superiores estão elevados e os punções inferiores estão mais baixos, permitindo assim que a matriz seja preenchida com grânulo em excesso. Terminada esta etapa, os punções inferiores são empurrados em direcção à matriz removendo a quantidade de grânulo em excesso. De seguida, inicia-se a fase da pré-compressão, que tem como objectivo reduzir a porosidade e o ar ainda existente entre os granulados, através do movimento dos punções superiores e inferiores. Já na fase da compressão principal, o punção superior desce e o punção inferior sobe realizando assim a compressão do granulado, finalizando a formação do comprimido. Por fim, os comprimidos são expulsos da compressora dando lugar a um novo ciclo.

3.1.4.2 Compressora excêntrica

Esta compressora, ao contrário da compressora rotativa, é utilizada para pequenos lotes. Nesta compressora, o punção superior inicia o movimento de descida e comprime a mistura que está na matriz, voltando depois à sua posição inicial, estando o punção inferior imóvel. De seguida, o punção inferior sobe, conduzindo

o comprimido ao nível superior da matriz, por onde vai ser expulso. O punção inferior retoma a sua posição inicial, encetando um novo ciclo de operações.

Nesta fase do processo são feitos ajustes da máquina antes do início e durante a operação. Estes ajustes são os testes do peso médio e peso individual, assim como a determinação da friabilidade, dureza e desintegração dos comprimidos. As amostras de comprimidos utilizadas nos testes para ajustes da compressora são constituídas por resíduos. Também durante a compressão são enviadas amostras de comprimidos para o laboratório de Controlo de Qualidade para se proceder ao teste de dissolução. Ainda como resíduos, são considerados o pó de granulado aspirado da compressora, sobras no alimentador, os comprimidos rejeitados pelas compressoras e os que caem ao chão[43].

3.1.5 Revestimento

A aplicação de revestimento aumenta o custo do produto, pelo que só serão revestidos os comprimidos que dele necessitam. Assim, há que ter em atenção três aspectos no que diz respeito ao revestimento: propriedades dos comprimidos, processo de revestimento e sua composição.

Este processo tem como objectivo modificar uma ou mais características do produto, como por exemplo, a cor, o sabor, a textura, controlar a libertação do fármaco no comprimido e conferir o aspecto, numa perspectiva de marketing pela utilização de determinadas cores.

Geralmente uma mistura de revestimento (*opadry*) é dissolvida num solvente orgânico ou em água para formar uma solução. Esta solução tem que ser compatível com o fármaco, a fim de melhorar as suas propriedades, assim como a resistência mecânica e a sua adesão ao comprimido. Esta solução é preparada numa panela e é homogeneizada com a ajuda de uma bomba até deixar de apresentar grânulos. Após esta solução estar homogeneizada, a bacia do equipamento de revestimento é carregada com os comprimidos obtidos na fase da compressão. De seguida, com a ajuda de uma bomba peristáltica, a solução é sugada para as pistolas que se encontram dentro da bacia.

As pistolas de pulverização são colocadas sobre um braço que se encontra no interior do equipamento e são dirigidas para os comprimidos. As pistolas movem-se sempre à mesma distância, que é estabelecida de acordo com o produto, sem nunca se intersectarem, de modo a que o revestimento seja homogéneo.

Assim que o sistema esteja preparado, dá-se início ao revestimento. A bacia, que contém quatro pás fixas para que a queda dos comprimidos seja amortizada e para que não sejam todos pulverizados ao mesmo tempo, gira sobre o seu eixo numa câmara fechada com velocidade apropriada. Assim sendo, à medida que a bacia gira, os comprimidos são pulverizados com a suspensão de revestimento até que estejam totalmente revestidos. Ao serem pulverizados, a sua superfície fica húmida e a película aderente tem de secar antes da próxima quantidade de solução ser aplicada.

O revestimento é dado como concluído quando os comprimidos apresentarem a massa desejada.

Nesta etapa, à semelhança da anterior, são realizados testes do peso médio e do peso individual dos comprimidos. Assim, os resíduos gerados durante esta etapa são as amostras utilizadas na realização dos testes do peso dos comprimidos, sobras da solução e a amostra que é enviada para o laboratório de Controlo de Qualidade para posterior realização de testes.

No final deste longo processo, os comprimidos prosseguem para o embalagem. Esta etapa consiste em acondicionar o produto, de forma a manter as suas características químicas e físicas, até ao momento em que chega ao consumidor.

O embalagem primário, que é também o mais crítico devido à quantidade de agentes envolvidos, refere-se ao embalagem dos medicamentos nos respectivos blisters. O embalagem secundário diz respeito às caixas de cartão que contêm os blisters e os medicamentos e às literaturas que os acompanham. Por fim, o embalagem terciário refere-se ao embalagem das caixas de medicamentos em caixas de cartão, para posteriormente serem entregues ao cliente.

No embalagem primário, o PVC e o alumínio são colocados em rolos na máquina. Ao formar os alvéolos por deformação plástica a quente, o rolo vai diminuir a sua dimensão inicial. Assim, o plástico deformado é arrefecido para que se possa manusear em segurança o alvéolo que contém os comprimidos. Os comprimidos são colocados no saco de tremonha e caem para uma calha linear, graças a uma vibração, chegando a um prato em espiral. De seguida, estes passam para o canal de dosagem, que tem como objectivo distribuir os comprimidos pelos vários alvéolos. Ainda na fase do embalagem primário, ocorre a selagem do alumínio com o PVC. Este processo é continuamente repetido, rejeitando os produtos não conformes para caixas, onde serão verificados posteriormente (contentores). No início do embalagem secundário, a encartonadora recebe as caixas de comprimidos fechadas, abrindo-as com recurso a um sistema de vácuo por ventosa, sendo feita a impressão do lote, da validade nos blisters e a colocação da literatura dentro de cada caixa. As caixas rejeitadas neste processo são enviadas para os respectivos contentores. Fechadas as caixas que contêm os medicamentos, estas passam por uma balança, onde são pesadas e etiquetadas. Se não tiverem o peso desejado são também rejeitadas e colocadas nos respectivos contentores. Por fim, ocorre o embalagem terciário que diz respeito à cintagem das caixas maiores que contêm as dos medicamentos.

Resta apenas salientar que as caixas ou blisters que sejam rejeitados durante o processo são posteriormente verificados um a um, manualmente, evitando qualquer possibilidade de ocorrência de erros.

Durante toda esta operação são realizados testes de controlo de processo: variação de peso, dureza e friabilidade. Da mesma forma que nas operações descritas acima, são enviadas amostras de comprimidos ao laboratório de Controlo de Qualidade para se proceder ao teste de dissolução e determinação do teor no produto acabado. A quantidade de comprimidos amostrada para o teste, os que são utilizados para os testes de controlo de processo, os comprimidos rejeitados pela máquina de embalagem, os comprimidos com defeito, bem como o meio de dissolução são considerados resíduos. Como resíduos recicláveis desta operação, poderão resultar as caixas e literaturas danificadas ou destruídas, além do material usado no ajuste da encartonadora e do PVC [41, 43, 22].

3.2 Tratamento de dados

Para o tratamento de dados, utilizaram-se como métodos de análise o Diagrama de Pareto, a técnica de Mínimos Quadrados Parciais (PLS) e a técnica dos

Mínimos Quadrados Multivariados (MOLS). Desta forma, os dados recolhidos em cada processo foram tratados através do Diagrama de Pareto, a fim de identificar as causas que estão na base da produção de resíduos, enquanto que os detalhes de cada processo foram tratados através de PLS e MOLS para retirar conclusões acerca da produção dos ditos resíduos.

3.3 Critérios

A base de dados resultou de um acompanhamento sistemático da produção durante um período de três meses (recolha de dados efectuada entre Janeiro e Abril de 2014). Teve-se especial cuidado em considerar o máximo de condições variáveis do processo e, sempre que possível, contabilizar de forma individualizada cada tipo de resíduo.

Matriz de valores

Esta matriz original continha $N = 241$ condições de fabrico (objectos referenciados como linhas) em função de 134 variáveis das quais $M = 119$ se referem a condições de fabrico (predictores) e $L = 13$ respostas. Assim sendo, originalmente a matriz de predictores possuía as dimensões $X_{(241 \times 119)}$ enquanto que a matriz de respostas as dimensões $Y_{(241 \times 13)}$.

No sentido de extrair informação numericamente tratável da base de dados original do sistema em estudo sob a forma de uma sub-matriz numérica e contínua, para cada caso de resposta em estudo foram feitas diversas eliminações:

- para maximizar a retenção de objectos foi feita a supressão de variáveis que apresentavam graves omissões de valores,
- para maximizar a retenção de variáveis foi feita a supressão de alguns objectos.

Critérios de aceitação e rejeição

Ao nível da qualidade é frequente ter-se como referência os níveis de significância de $\alpha = 0.05$ e $\alpha = 0.01$ para estabelecer a tomada de decisão ao nível de confiança de 95 e 99%, respectivamente. Assim, se o valor de prova da hipótese nula for $p[H_0] \geq 0.05$ estamos perante a situação em que a hipótese nula é directamente validada; caso esse valor seja $p[H_0] < 0.01$ a hipótese nula terá de ser liminarmente rejeitada.

Contudo, na região de probabilidade situada entre 0.01 e 0.05, há uma dúvida estatística que é difícil de ser ultrapassada.

Como critério de decisão poderemos então acrescentar que, sendo 0.03 o valor intermédio entre 0.01 e 0.05, este poderá ser utilizado como limite de decisão. Assim, se se obtiver um valor de prova $\alpha \geq 0.03$ isto significará aprovação da hipótese nula e em caso de se atingirem valores de prova $\alpha < 0.03$, este implicará a rejeição da hipótese nula com a subsequente aceitação da hipótese alternativa (H_1).

Modelação

No sentido de evitar os efeitos de escala das variáveis e ter de utilizar termos

independentes, no caso da modelação explícita, as variáveis foram previamente normalizadas.

Impacto das variáveis

No sentido de antever o impacto de cada variável representada sobre cada factor latente relevante, foi considerado o módulo do valor do respectivo peso. Dado que os factores latentes constituem uma nova base vectorial ortonormal, os seus vectores próprios além de ortogonais apresentam norma unitária

$$\sqrt{\sum_{k=1}^M w_{jk}^2} = 1 \quad (3.1)$$

se a dimensão do sub-espço predictor for $j = M$ então as contribuições passam a ser superiores à média quando cada peso excede o valor médio

$$\bar{w} = \frac{1}{\sqrt{M}} \quad (3.2)$$

estimado com base na dimensão M do sub-espço predictor.

Modelação explícita

Os modelos polinomiais multivariados apresentavam no início a dependência de grau um em relação a todas as variáveis a utilizar.

Estando as variáveis normalizadas o modelo a utilizar (η_i) apenas deverá considerar a dependência directa relativa a cada uma das M variáveis a avaliar

$$\eta_i = \sum_{j=1}^M b_j \cdot x_{ij} \quad (3.3)$$

onde os parâmetros b_j traduzem a dependência linear relativa à variável j .

Capítulo 4

Resultados e discussão

Neste capítulo serão avaliados e discutidos os dados obtidos na empresa durante os meses de Janeiro a Abril de 2014 quanto à produção de resíduos gerados na fabricação e embalagem. Para isso, recorrer-se-á aos métodos de análise descritos no capítulo 2.

É de mencionar que esta metodologia foi utilizada para avaliar a quantidade de resíduos gerados, a fim de reduzir a sua produção.

4.1 Análise preliminar

Os resíduos industriais podem ser avaliados segundo diferentes perspectivas, sendo mais comum a avaliação do seu impacto em termos de quantidade, perigosidade, persistência, custos, etc.

No sentido de tentar identificar o impacto dos resíduos gerados na empresa, nesta secção do trabalho vamos avaliar a quantidade de resíduo gerado segundo duas abordagens - considerando a quantidade gerada (impacto material) e o seu custo (impacto financeiro).

4.1.1 Impacto material

Foi realizada uma análise do impacto material do resíduo gerado, segundo a abordagem de Pareto.

Uma vez que no sistema em análise existem treze respostas para descrever por via multivariada, convém fazer previamente uma avaliação do impacto de cada uma destas respostas (resíduos industriais) no seu cômputo global, quer em termos de massa (volume de resíduos gerados), quer em termos de custos (dispêndio do processo ao nível do tratamento subsequente). Faz sentido que a primeira análise a realizar seja em termos de massa. Escolhemos a abordagem do Diagrama de Pareto para mais facilmente localizarmos as respostas mais relevantes neste caso. Na tabela seguinte, tabela (4.1), apresenta-se o cômputo global de massa de resíduo gerado durante o período de Janeiro a Abril de 2014.

Tabela 4.1: Quantidade de resíduos gerados e seu impacto individualizado segundo a abordagem de Pareto.

Resposta	Quantidade (kg)	Quantidade relativa (%)	Quantidade cumulativa (%)
y13	643.50	18	18
y7	507.61	14	32
y1	492.69	14	46
y12	402.00	11	57
y9	311.40	9	66
y4	259.70	7	73
y5	252.42	7	80
y8	197.60	6	86
y3	192.82	6	92
y6	111.80	3	95
y11	90.88	3	98
y10	78.10	2	100
y2	2.94	0	100

Para facilitar a análise destes resultados, a quantidade de resíduos foi representada em diagrama circular figura (4.1) e num gráfico de Pareto (4.2).

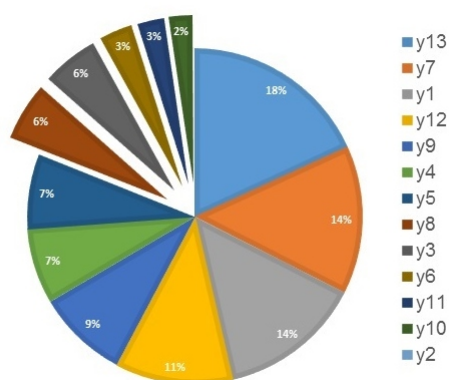


Figura 4.1: Representação do impacto material de cada um dos resíduos avaliados: as secções separadas do círculo central representam contribuições menos relevantes de acordo com a análise de Pareto.

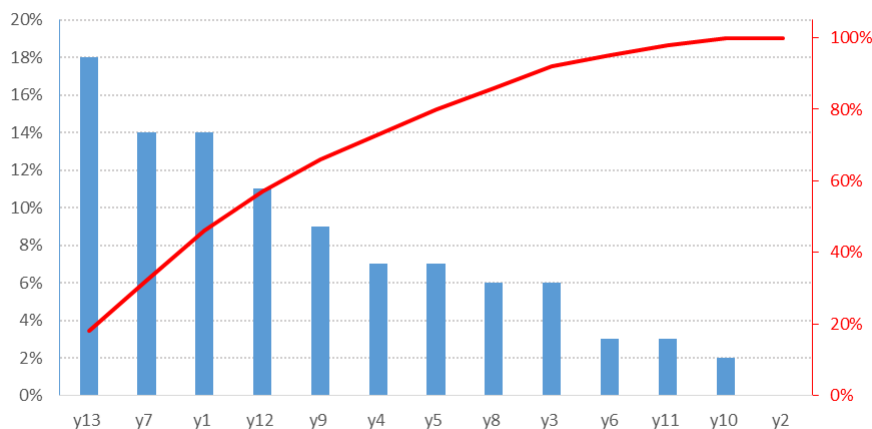


Figura 4.2: Diagrama de Pareto para a avaliação do impacto do material dos resíduos estudados.

De acordo com a tabela (4.1), existem sete resíduos industriais que devem ser tomados em conta de modo a justificar cerca de 80 % do resíduo total gerado em massa. De entre estas variáveis assinaladas, y13 é aquela que apresenta maior impacto seguida de y7, y1, y12 e y9 e por fim com duas contribuições mais ligeiras de y4 e y5.

Da figura (4.1), verifica-se que as contribuições de y13, y7 e y1 representam 50 % do resíduo total estudado. Se a esta contribuição forem adicionados os resíduos y12, y9 e y4, passaremos então a estar a considerar cerca de 3/4 do resíduo total.

Olhando o Diagrama de Pareto, figura (4.2), nota-se que as primeiras cinco contribuições (y13, y7, y1, y12, y9) apresentam uma sequência nitidamente decrescente enquanto que a partir de y4 até ao final estas são muito similares - isto revela um maior impacto nas primeiras cinco variáveis em relação às restantes.

Como observado anteriormente, no impacto de massa de resíduo produzido, foram identificadas as respostas y13, y7, y1, y12, y9, y4 e y5 como sendo as mais relevantes, embora apenas as cinco primeiras apresentassem maior impacto.

4.1.2 Impacto financeiro

Passando agora à análise do impacto em termos de custos do tratamento do resíduo, vamos tentar (de modo similar) identificar as respostas mais relevantes.

Na tabela seguinte, tabela 4.2, está representado o custo para os diferentes tratamentos de resíduos gerados durante o período em estudo.

Tabela 4.2: Custo dos resíduos gerados e seu impacto individualizado segundo o Diagrama de Pareto

Resposta	Quantidade (kg)	Custo (€)	Custo relativo (%)	Custo total (%)
y7	507.61	395.9	17.3	17.3
y1	492.69	384.3	16.8	34.0
y12	402.00	313.6	13.7	47.7
y9	311.40	242.9	10.6	58.3
y4	259.70	202.6	8.8	67.1
y5	252.42	196.9	8.6	75.7
y8	197.60	154.1	6.7	82.4
y3	192.82	150.4	6.6	89.0
y6	111.80	87.2	3.8	92.8
y11	90.88	70.9	3.1	95.8
y10	78.10	60.9	2.7	98.5
y13	643.50	32.2	1.4	99.9
y2	2.94	2.3	0.1	100.0

Para facilitar a análise destes resultados, o impacto financeiro para o tratamento dos resíduos foi representado em diagrama circular (figura 4.3) e num gráfico de Pareto (figura 4.4).

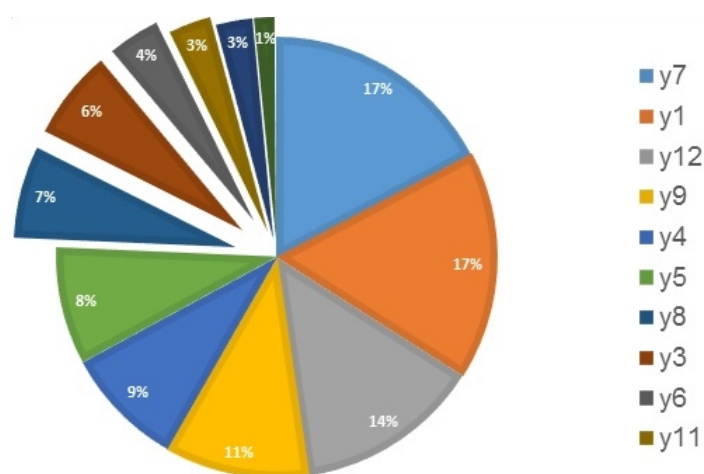


Figura 4.3: Representação do impacto financeiro de cada resíduo gerado: as secções separadas do círculo central representam contribuições pouco relevantes de acordo com a análise de Pareto.

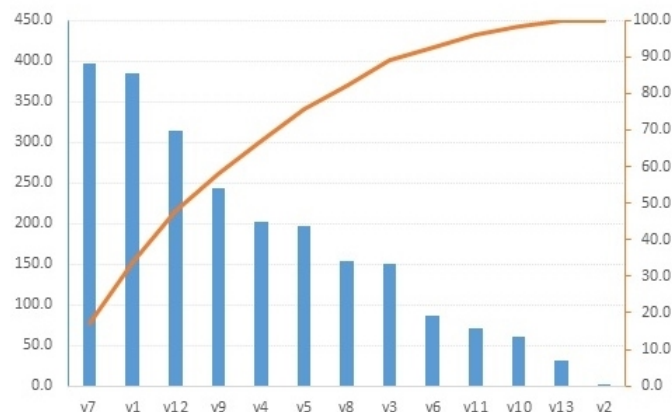


Figura 4.4: Diagrama de Pareto para a avaliação do impacto financeiro dos resíduos gerados.

De acordo com a tabela 4.2 existem seis resíduos que devem ser considerados de modo a justificar cerca de 80 % do resíduo total gerado em custo. De entre estas variáveis assinaladas, y7 é aquela que apresenta maior impacto seguida de y1, y12, y9, y4 e y5 e por fim, com contribuições menos relevantes estão inseridas as restantes variáveis.

Da figura 4.3 verifica-se que as contribuições de y7, y1 e y12 representam 50% do resíduo total estudado. Se a esta contribuição forem adicionados os resíduos y9, y4 e y5, passaremos então a estar a considerar cerca de 3/4 do resíduo total.

No que diz respeito ao Diagrama de Pareto, figura 4.4, observa-se que as primeiras cinco contribuições (y7, y1, y12 e y9) apresentam uma sequência nitidamente decrescente enquanto que a partir de y4 até ao final estas são muito similares, ou seja, o decréscimo é menos acentuado.

4.2 Análise detalhada

Na análise preliminar do impacto material, secção (4.1.1), foram evidenciados os resíduos y13, y7, y1, y12, y9 e y4 como sendo os mais relevantes, por ordem decrescente de importância.

Já na análise do impacto financeiro, secção (4.1.2), foram identificadas as causas maioritárias como sendo y7, y1, y12, y9, y4 e y5.

Comparando a primeira situação com a segunda verifica-se que o resíduo y13 (maior massa identificada), por possuir menores custos de processamento deixa de ser relevante na análise de custos. Os restantes resíduos identificados (y7, y1, y12, y9 e y4) preservam suas posições respectivas de relevância atendendo a que os custos de processamento são similares sendo a massa gerada a componente mais relevante nesta sequência identificada. Por fim, o resíduo identificado como y5, ausente do conjunto maioritário na análise de impacto material surge na análise de impacto de custo como consequência da perda de relevância de y13 devido ao custo de tratamento ser muito inferior aos restantes.

Se se atentar nos resultados da tabela, (4.1), na página 60, a variável y_5 surge ainda dentro dos 80% do impacto material global.

No sentido de tentar agora compreender as causas dos resíduos gerados na empresa, e poder sugerir modo de redução dos seus impactos material e financeiro, nesta secção vamos estudar em detalhe as causas associadas a cada um destes resíduos previamente evidenciados.

4.2.1 Resposta y_{13}

Neste caso em concreto a matriz dos predictores é muito exígua - contém apenas 20 objectos ($N = 20$) representados sobre 36 dimensões ($M = 36$, variáveis x_{55} , x_{56} , x_{57} , x_{58} , x_{59} , x_{60} , x_{61} , x_{62} , x_{63} , x_{64} , x_{65} , x_{66} , x_{67} , x_{68} , x_{69} , x_{70} , x_{75} , x_{76} , x_{77} , x_{78} , x_{79} , x_{80} , x_{81} , x_{82} , x_{84} , x_{85} , x_{86} , x_{87} , x_{91} , x_{92} , x_{93} , x_{101} , x_{102} , x_{103} , x_{107} , x_{108}) apresentando as dimensões $X_{(20 \times 36)}$.

4.2.1.1 Análise de correlação

Com $N = 20$, a correlação será nula ($H_0 : |\rho| = 0$), se valores da correlação forem, (ver secção (2.4.3) na página 46) em valor absoluto, inferiores a 0.444 ($\alpha \geq 0.050$ enquanto que valores superiores a 0.561 já não podem ser considerados nulos ($\alpha < 0.01$)).

Para avaliar se a correlação pode ser considerada unitária ($H_0 : |\rho| = 1$), valores superiores a 0.561 são considerados unitários ($\alpha \geq 0.05$) enquanto que valores inferiores a 0.444 já são considerados distintos da unidade ($\alpha < 0.01$).

Se se atender aos valores limites obtidos, a sobreposição destes intervalos não facilita esta interpretação.

Estipulando o critério $\alpha = 0.03$ para tentar resolver esta situação, ver secção (3.3) na página 57, fica-se então com os limites 0.485 e 0.530 para afirmar que $|\rho| = 0$, $|\rho| \neq 0$ e $|\rho| = 1$, respectivamente, o que se traduz numa excelente simplificação. Assim sendo, módulos de correlação inferiores a 0.485 são considerados nulos (indicando que não existe correlação significativa) enquanto que valores de correlação superiores a 0.530 indicam uma elevada correlação. Já os valores intermédios indicam que apenas que estes não são nulos.

Estudando a correlação interna ao sub-espaco predictor verificou-se a igualdade ($\rho = 1$) entre os predictores $x_{59} = x_{58}$, $x_{65} = x_{64}$ e $x_{103} = x_{102}$. Foi também identificada uma forte interdependência entre x_{70} - x_{66} [-0.999] e x_{101} - x_{60} [0.996].

Estas variáveis provavelmente terão de ser omitidas no caso de ajuste polinomial explícito (MOLS) (ver secção (2.4.5), na página 46).

Atendendo ao valor limite inferior para correlação não nula ($|\rho| \geq 0.485$) detectaram-se as seguintes interdependências, x_{84} , x_{60} , x_{101} , x_{58} e x_{59} .

Atendendo aos valores de aceitação e rejeição encontrados conclui-se que x_{84} e x_{60} possuem uma correlação elevada com a resposta y_{13} enquanto que os restantes três (x_{101} , x_{58} e x_{59}) apresentam alguma interdependência, não nula.

Na figura (4.5) apresentam-se representadas as interdependências de y_{13} com os predictores que evidenciaram maior correlação directa.

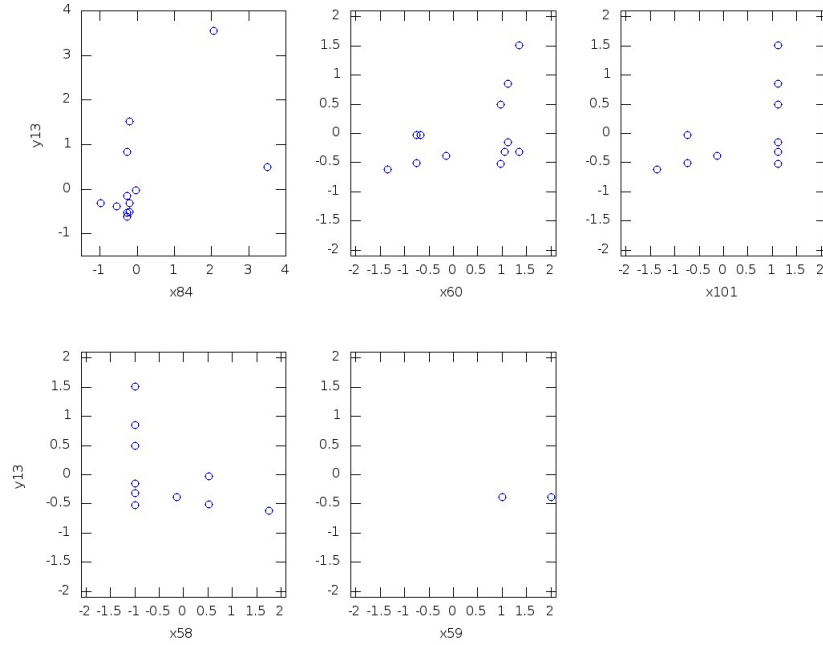


Figura 4.5: Representação das interdependências de y13 com as variáveis mais correlacionadas x84 [0.557], x60 [0.526], x101 [0.452], x58 [-0.490] e x59 [-0.490].

Da figura anterior verifica-se que a interdependência detectada como significativa (x84, x60) ou não nula (x58, x59 e x101) apresenta uma variação mútua com y13 onde é difícil de conceber esta interdependência.

4.2.1.2 Abordagem PLS

Como foi anteriormente mencionado, na secção (2.4.4), página 38, o algoritmo PLS procura descrever o sub-espço da resposta com base na informação contida no sub-espço dos predictores. A melhor forma de avaliar a progressão desta análise é normalizar, em termos percentuais, a variabilidade contida nos sub-espços predictor (SS_x) e resposta (SS_y), sob a forma de somas de quadrados, em variabilidade residual dos respectivos sub-espços ($\%R_x$ e $\%R_y$). Em cada iteração realizada vai-se utilizando informação contida no sub-espço predictor ($\%$) para justificar a resposta ($\%J_y$), sendo esta informação correlacionada através da estimativa de um parâmetro do modelo (β_i). Assim, em cada iteração pode ser estimada a eficiência do factor latente (η_i) em causa através de

$$\eta_i = \frac{\%J_y}{\%U_x} \quad (4.1)$$

Como critérios de decisão do número de factores latentes (f) necessários à modelação da resposta usa-se a descrição cumulativa da resposta ($\sum(\%J_y)$) até atingir os 80% da recuperação da informação e/ou factores latentes com eficiências acima do normal ($\eta_i > 1$).

Na tabela (4.3) encontram-se as estimativas obtidas da análise PLS em relação à resposta y_{13} .

Tabela 4.3: Resultados obtidos na abordagem da resposta y_{13} através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum(\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum(\%J_y)$	β_i	η_i
0	684.00	100.0	—	—	19.00	100.0	—	—	—	—
1	570.82	83.5	16.5	16.5	7.60	40.0	60.0	10.0	3.376	3.63
2	518.46	75.8	7.7	24.2	2.58	13.6	26.4	86.4	2.241	3.45
3	376.48	55.0	20.8	45.2	1.63	8.6	5.0	91.4	0.976	0.24
4	310.57	45.4	9.6	54.6	1.15	6.1	2.5	93.9	0.689	0.26
5	272.29	39.8	5.6	60.2	0.66	3.5	2.6	96.5	0.701	0.46
6	215.00	31.4	8.4	68.6	0.31	1.6	1.8	98.4	0.592	0.22
7	154.99	22.7	8.8	77.3	0.10	0.5	1.1	99.5	0.463	0.13
8	78.96	11.5	11.1	88.5	0.05	0.3	0.2	99.7	0.2152	0.02
9	37.75	5.5	6.0	94.5	0.00	0.0	0.2	100.0	0.217	0.04
10	21.22	3.1	2.4	96.9	0.00	0.0	0.0	100.0	0.044	0.00

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos predictores e das respostas); $\%R_x$ e $\%R_y$ informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justificar a resposta em y ; $\sum(\%U_x)$ e $\sum(\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y/\%U_x$).

Da tabela anterior verifica-se que com apenas dois factores latentes ($f = 2$) consegue-se descrever 86.4% da resposta em causa.

Em termos de impacto na justificação da resposta vejamos como se comportam os factores latentes estimados:

1. o primeiro factor latente ($FL1$) utiliza cerca de 16.5% da informação contida no sub-espaco dos predictores para descrever cerca de 60% da informação contida nesta resposta o que corresponde a uma eficiência de $\eta_1 = 3.63$, muito superior a 1;
2. o segundo factor latente ($FL2$) requer 7.7% da informação contida nos predictores para descrever cerca de 26.4% da resposta possuindo também uma eficiência bastante elevada ($\eta_2 = 3.45$);
3. o terceiro factor latente utiliza cerca de 20.8% da informação residual dos predictores para apenas justificar 5.0% da resposta apresentando uma elevada ineficiência sendo prenúncio de não ser relevante a informação veiculada.
4. os restantes factores latentes apenas conseguem descrever cerca de 9% da informação residual da resposta.

As observações anteriores evidenciam que apenas os dois primeiros factores latentes veiculam informação pertinente, capaz de descrever a interdependência da resposta com o sub espaco dos predictores.

Na figura (4.6) encontram-se representados os scree plots¹ referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

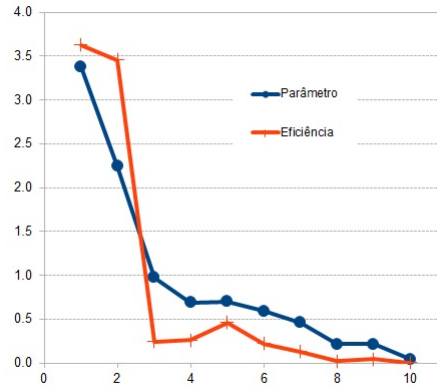


Figura 4.6: Scree plot dos parâmetros estimados pelo modelo e da eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Da figura anterior constata-se facilmente que as variáveis representadas (valores dos parâmetros do modelo e eficiência do factor latente) são concordantes e peremptórias em mostrar que apenas os dois primeiros factores latentes, $FL1$ e $FL2$, é que são relevantes.

Importa agora avaliar a informação contida em $FL1$ e $FL2$ em termos de representatividade dos predictores sobre a resposta.

Na tabela (4.4) apenas se apresentam os pesos mais significativos sobre a resposta $y13$, atendendo à condição expressa pela equação (3.2) na página 58.

¹Embora estejamos aqui a utilizar o termo Inglês scree plot o que não é totalmente correcto, a correspondente tradução para Português, sugerida por pesquisa em rede, é “gráfico de escarpa” o que ainda nos deixa mais insatisfeitos com esta sugestão, razão pela qual optou-se pelo termo original.

Tabela 4.4: Avaliação do impacto dos predictores mais relevantes sobre y13 através da análise das respectivas contribuições.

Predictor	FL1	FL2
x55	-0.263	
x58	0.275	
x59	0.275	
x60	-0.296	
x62		-0.420
x63		0.212
x68	-0.170	0.539
x69		-0.171
x78	-0.216	
x79	-0.243	
x80	-0.240	
x81	-0.258	
x82	0.208	
x84	-0.313	
x86		0.231
x87		0.233
x101	-0.294	

FL1 - primeiro factor latente; FL2 - segundo factor latente.

Atendendo a que ambos factores apresentam eficiências similares, da tabela (4.4) estamos a prever a dependência de y13 com 17 predictores, designadamente: x55, x58, x59, x60, x62, x63, x68, x69, x78, x79, x80, x81, x82, x84, x86, x87 e x101.

Estas dependências serão avaliadas com base no ajuste explícito com polinómios de grau unitário.

4.2.1.3 Modelação explícita

A modelação explícita, realizada através da imposição directa de um modelo polinomial multivariado de primeiro grau, (equação (3.3), na página 58), foi realizada através da perspectiva matricial desenvolvida na secção 2.4.5, página 46. Segundo este método, para conseguir obter as estimativas paramétricas é necessário que a matriz de projecto ($A_{(p \times p)}$), equação 2.48, na página 48, seja invertível o que corresponde a dizer que o seu determinante não pode ser nulo.

Neste caso em concreto a matriz de predictores possui as dimensões $X_{(20 \times 36)}$ evidenciando que o número de objectos ($N = 20$) é notoriamente inferior ao número de variáveis ($M = 36$). Como consequência o sistema de equações que permitem obter a estimativa paramétrica diz-se indeterminado - existirá uma infinidade de valores para as estimativas paramétricas que satisfazem o ajuste do modelo. Neste caso em concreto a matriz de projecto deverá então possuir um determinante próximo de zero o que dificultará a estimativa paramétrica.

Teve-se que enfrentar a decisão de suprimir alguma da informação original para permitir a continuação do estudo por via explícita.

A primeira abordagem testada consistiu em segmentar as 36 variáveis em dois sub-conjuntos (variáveis de 1 a 18 e variáveis de 19 a 36) e realizar a abordagem MOLS iterativa para tentar apurar as variáveis relevantes em cada

sub-conjunto e, no final, reunir as variáveis relevantes para um estudo derradeiro. Contudo esta abordagem não teve sucesso dado que a matriz de projecto não era invertível.

Olhando para os resultados obtidos na avaliação da correlação directa entre predictores e resposta, (2.4.3) na página 36, e a modelação implícita, 2.4.4, página na página 38., verificou-se que as correlações mais relevantes foram também detectadas ao nível dos factores latentes relevantes. Sendo este conjunto de 17 predictores evidenciados por PLS (x55, x58, x59, x60, x62, x63, x68, x69, x78, x79, x80, x81, x82, x84, x86, x87 e x101) aquele em que acreditamos estar definidos os termos de interacção predictor-resposta mais relevantes, foi utilizado como matriz de entrada para a modelação explícita. Contudo, verificou-se que estes não podiam ser utilizados em simultâneo neste estudo já que o cálculo matricial conduzia a matrizes singulares, não invertíveis.

Foi averiguada a causa desta anomalia numérica e chegou-se à conclusão de que a matriz dos predictores estava mal acondicionada já que ainda continha três colunas fortemente dependentes (x59, x81 e x101) de (x58, x78 e x60), respectivamente.

Após a remoção destes já foi possível realizar o estudo do ajuste explícito com modelo polinomial usando apenas 14 parâmetros.

Segundo a estratégia da remoção sucessiva da variável cujo parâmetro possuía menor significado estatístico, foram realizadas 14 iterações - começando com um modelo completo ($p = 14$) convergiu-se para uma única variável ($p = 1$). Os resultados encontram-se sistematizados na tabela (4.5).

Tabela 4.5: Convergência iterativa do modelo polinomial obtido por MOLS; inicia-se o estudo com todas as variáveis (número de parâmetros $p = 14$) e vai-se rejeitando sucessivamente a variável que apresenta parâmetro com menor significado estatístico (que possui maior erro relativo). Os valores 1 e 0 associados às variáveis em estudo indicam a presença ou ausência do respectivo parâmetro no modelo polinomial.

Iteração	1	3	4	5	6	7	8	9	10	14
x55	1	0	0	0	0	0	0	0	0	0
x58	1	1	1	1	1	1	1	1	1	0
x60	1	1	1	1	1	1	0	0	0	0
x62	1	1	1	0	0	0	0	0	0	0
x63	1	1	1	1	1	0	0	0	0	0
x68	1	1	1	1	1	1	1	1	1	0
x69	1	1	1	1	1	1	1	1	0	0
x78	1	1	1	1	1	1	1	1	1	1
x79	1	1	1	1	0	0	0	0	0	0
x80	1	1	1	1	1	1	1	1	1	0
x82	1	1	1	1	1	1	1	0	0	0
x84	1	1	1	1	1	1	1	1	1	0
x86	1	0	0	0	0	0	0	0	0	0
x87	1	1	0	0	0	0	0	0	0	0
Nm.	M14	M12	M11	M10	M09	M08	M07	M06	M05	M01
p	14	12	11	10	9	8	7	6	5	1
SS	124.56	0.00	0.12	0.71	0.80	0.93	1.41	1.95	3.48	16.19
ν	6	8	9	10	11	12	13	14	15	19
σ_{fit}^2	20.760	0.000	0.013	0.071	0.073	0.077	0.109	0.139	0.232	0.852
ρ	0.363	1.000	0.997	0.981	0.979	0.975	0.962	0.947	0.904	0.384
$\rho = 0$	0.377	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.085
$\rho = 1$	0.145	—	0.909	0.765	0.737	0.705	0.624	0.549	0.398	0.009

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47), ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Da tabela anterior verificam-se algumas curiosidades:

1. o primeiro modelo (M14) que é o mais completo é aquele que apresenta pior desempenho quer na qualidade do ajuste (σ_{fit}^2) quer na descrição da resposta $y13$ (ρ) - a correlação múltipla encontrada é tendencialmente nula já que o valor de prova encontrado para que a correlação seja nula ($\alpha = 0.377$);
2. em termos de desempenho dos modelos (σ_{fit}^2 e ρ), (4.7), estes estimadores começam mal com M14, atingem um valor óptimo no caso do modelo exacto (M12) e depois vão piorando progressivamente até que terminam assumindo valores inadequados;
3. o parâmetro remanescente no modelo M01 é referente ao predictor x78 - é algo surpreendente esta conclusão já que a sua correlação com a resposta $y13$ é menos eficaz que a que apresenta com x84 [0.557] e x60 [0.526];

4. existe um modelo “exacto” (M12 com $p = 12$) para descrever a resposta y_{13} - este modelo contém as contribuições de x_{60} , x_{62} , x_{63} , x_{68} , x_{69} , x_{78} , x_{79} , x_{80} , x_{82} , x_{84} , x_{87} , contendo as contribuições x_{84} e x_{60} (correlações dadas como estatisticamente equivalentes à unidade) e a variável x_{58} (com correlação não nula), previamente encontradas na secção 4.2.1.1, na página 64.
5. existe um modelo parcimonioso (M08 com $p = 8$) para y_{13} que contém os predictores x_{58} , x_{60} , x_{68} , x_{69} , x_{78} , x_{80} , x_{82} , x_{84} ; nele estão preservadas as contribuições de x_{84} , x_{60} e x_{58} ; a correlação múltipla encontrada ($\rho = 0.975$) é excelente já que o valor de prova para $H_o : \rho = 1$ é de $\alpha = 0.705$ enquanto que para $\rho = 0$ é de $\alpha = 0.000$.

A figura (4.7) apresenta a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação (ρ) em função do número de parâmetros do modelo evidenciando o modelo parcimonioso (M08). Já a figura (4.8) apresenta a qualidade do ajuste conseguida com o modelo exacto (M12) e o modelo parcimonioso (M08).

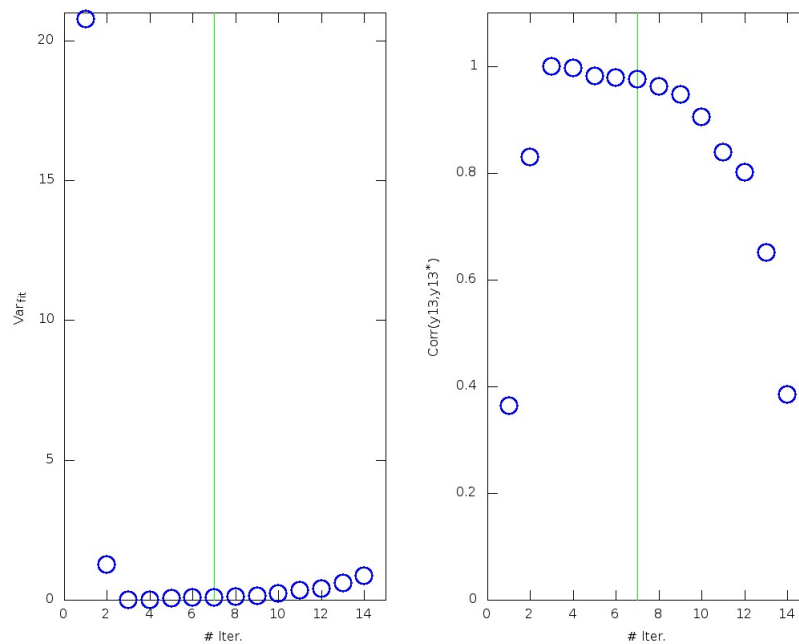


Figura 4.7: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa do melhor modelo. O traço vertical assinala o modelo parcimonioso.

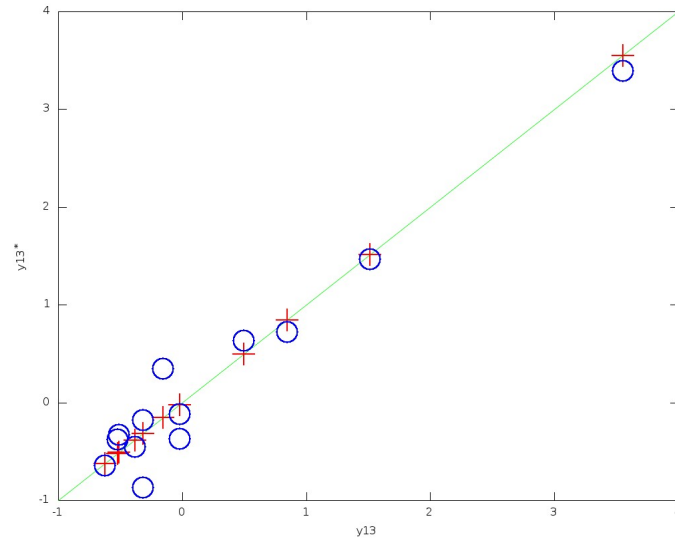


Figura 4.8: Capacidade de previsão dos modelos - representação dos valores previstos pelo modelo exacto (M12, cruces a vermelho) e parcimonioso (M08, círculos azuis) em função da resposta y_{13} . A linha oblíqua representa a bisectriz dos quadrantes ímpares.

As observações aqui encontradas podem ser explicadas se atendermos a que, a sobreparametrização do modelo confere-lhe um excesso de flexibilidade funcional que o torna instável quanto ao modo como justifica e prevê valores, conferindo-lhe um deficiente ajuste. Assim, o modelo com excesso de parâmetros (M14) apresenta um comportamento deficiente na forma como descreve os valores experimentais.

À medida que vão sendo removidos parâmetros do modelo ele vai adquirindo maior rigidez o que lhe confere estabilidade na modelação dos valores experimentais, melhorando o seu desempenho em termos de correlação (ρ) e de ajuste (σ_{jit}^2), ver figura (4.7). Nesse momento a correlação atinge um valor máximo enquanto que o estimador de erro do modelo passa pelo seu valor mínimo. Contudo não é forçoso que se obtenha um modelo “exacto” em que a correlação é unitária e o erro nulo².

Quando a remoção de parâmetros começa a ser excessiva, o modelo passa a ter dificuldade em acompanhar a variabilidade dos dados perdendo na forma como prevê os valores (diminui a correlação) e aumenta a dispersão em relação aos valores reais. Nesta trajectória descendente dos valores da correlação, ver figura (4.7), existe algures um compromisso de ajuste óptimo que é conseguido com um número restrito de parâmetros - essa situação corresponde à do modelo parcimonioso.

Na tabela (4.6) apresenta-se as estimativas paramétricas obtidas com o modelo parcimonioso (M08) segundo o ajuste polinomial explícito.

²Valores de desvio da ordem de 10^{-15} correspondem ao erro de truncatura numérica de valores reais em precisão dupla.

Tabela 4.6: Estimativas paramétricas obtidas com o modelo parcimonioso (M08).

Variável	β	$\sigma(\beta)$	%RSE	TV	α
x78	2.20	0.25	11.5	8.71	0.000
x84	1.51	0.18	12.2	8.21	0.000
x68	0.66	0.09	14.1	7.10	0.000
x69	0.60	0.11	18.0	5.56	0.000
x82	0.42	0.17	39.7	2.52	0.027
x60	-0.99	0.39	39.8	2.51	0.027
x58	-1.78	0.41	23.1	4.33	0.001
x80	-2.48	0.34	13.8	7.26	0.000

β - valor do parâmetro estimado, equação (2.49), pág 48; $\sigma(\beta)$ - erro padrão do parâmetro, equação (2.55), pág.49 ; %RSE - erro padrão relativo, equação (2.49), pág 48 ; TV e α - valor de teste e respectivo valor de prova referente ao teste de significado estatístico do parâmetro considerando a hipótese nula $H_0 : |\beta| = 0$.

Da tabela verifica-se que todos os parâmetros encontrados têm significado estatístico (assumindo $H_0 : |\beta| = 0$).

Na tabela (4.7) encontra-se uma compilação dos resultados obtidos por análise de correlação, modelação implícita (PLS) e modelação explícita (MOLS).

Tabela 4.7: Compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (w) e modelação explícita (β).

Variável	x55	x58	x59	x60	x62	x63	x68	x69	x78
ρ	0.47	-0.49	-0.49	0.53					
w	-0.26	0.28	0.28	-0.30	-0.42	0.21	0.54	-0.17	-0.22
β		-1.78	-1.78*	-0.99			0.66	0.60	2.20
Variável	x79	x80	x81	x82	x84	x86	x87	x101	
ρ	0.43	0.43	0.46		0.56			0.52	
w	-0.24	-0.24	-0.26	0.21	-0.31	0.23	0.23	-0.29	
β		-2.48		0.42	1.51			-0.99*	

ρ - estimativas obtidas por análise de correlação; w - contribuição evidenciada por PLS; β - efeito obtido por modelação explícita através do modelo parcimonioso (M08).

* As contribuições de x59 e x101 foram acrescidas a MOLS devido à sua forte interdependência com x58 e x60, respectivamente.

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso como sendo a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação, ver secção 4.2.1.1, na página 64, identificou 6 dos 10 predictores mais influentes embora em dois desses casos o tipo de contribuição seja antagónica - ver casos x60, x80 e x101. Já a modelação implícita (PLS) aponta para 17 contribuições relevantes das quais 12 sugeridas pelo primeiro factor latente ($FL1$). Curiosamente, todas as variáveis indicadas como relevantes pelo modelo parcimonioso estão presentes em $FL1$ indicando que a análise PLS é mais vantajosa para o rastreio de inter-relações entre predictores e respostas que a avaliação da correlação directa.

Deste estudo resulta que a resposta y13 depende de x58, x60, x68, x69, x78, x80, x82 e x84 - corresponde ao modelo definido como parcimonioso (M08) e em que todos os parâmetros possuem significado estatístico (com $\nu = 12$).

Devido à interdependência estreita entre x58-x59 e x60-x101, as variáveis de estudo (x59 e x101) serão agora incluídas nas conclusões.

Da tabela (4.7) verifica-se que as variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y13 são x78 [2.20], x84 [1.51], x68 [0.66], x69 [0.60] e x82 [0.42] - a dependência positiva indica que estas possuem uma influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio.

Em contrapartida, as variáveis x80 [-2.48], x58-x59* [-1.78] e x60-x101* [-0.99] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y13 indicam que o seu aumento têm como impacto a redução do seu valor.

4.2.2 Resposta y7

Neste caso em concreto a matriz dos predictores é também muito exígua - contém apenas 30 objectos ($N = 30$) representados sobre 7 dimensões ($M = 7$, variáveis Xe, x34, x37, x38, x39, Xf, x40) apresenta as dimensões $X_{(30 \times 7)}$.

4.2.2.1 Análise de correlação

Com $N = 30$, impondo o critério $\alpha = 0.03$, ver secção (3.3) na página 57, verificou-se quais os valores limites de correlação para $|\rho| = 0$ e $|\rho| = 1$, obteve-se os valores 0.397 e 0.687 indicando que valores absolutos de correlação inferiores a 0.397 são nulos enquanto que valores superiores a 0.687 serão considerados unitários.

Estudando a correlação interna ao sub-espaco predictor foi detectada uma elevada interdependência entre Xe-Xf [0.999].

Esta variável provavelmente terá de ser omitida no caso de ajuste polinomial explícito (MOLS, ver secção (2.4.5), na página 46).

Atendendo aos valores de aceitação e rejeição encontrados conclui-se que x34 [-0.567] e x37 [0.444] possuem apenas correlações não nulas.

Na figura (4.9) encontram-se representadas as dependências de y7 em função de x34 e x37.

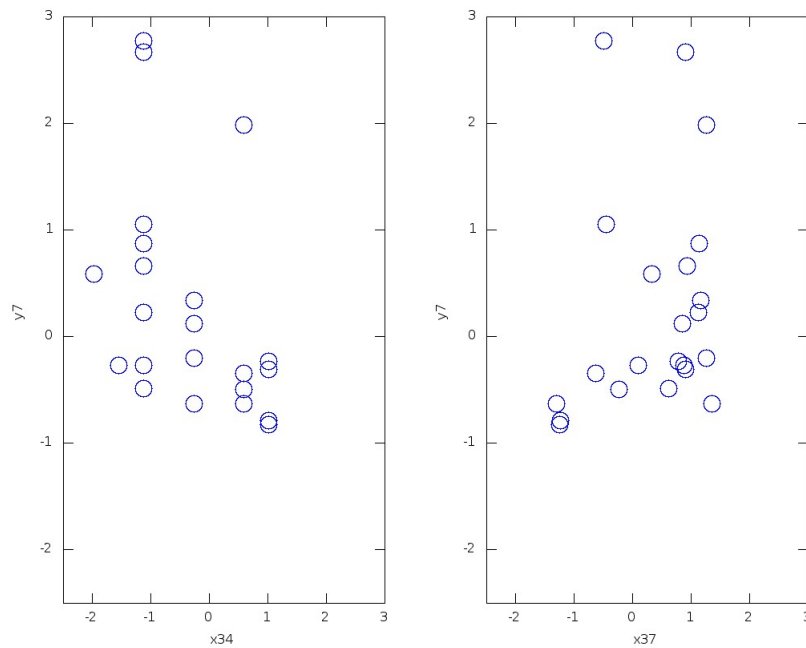


Figura 4.9: Representação das interdependências de $y7$ com as variáveis mais correlacionadas $x34$ [-0.567] e $x37$ [0.444].

Da figura anterior verifica-se que as fracas interdependências detectadas como não nulas ($x34$ e $x37$) apresentam uma variação mútua com $y7$ também difícil de assumir como relacionada.

4.2.2.2 Abordagem PLS

Como foi anteriormente mencionado, na secção (2.4.4), página 38, o algoritmo PLS procura descrever o sub-espço da resposta com base na informação contida no sub-espço dos predictores.

Na tabela (4.8) encontram-se as estimativas obtidas da análise PLS em relação à resposta $y7$.

Tabela 4.8: Resultados obtidos na abordagem da resposta y7 através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum(\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum(\%J_y)$	β_i	η_i
0	203.00	100.0	—	—	29.00	100.0	—	—	—	—
1	553.33	75.5	24.5	24.5	12.58	43.4	56.6	56.6	4.051	2.31
2	118.91	58.6	17.0	41.4	10.61	36.6	6.8	63.4	1.404	0.40
3	102.25	50.4	7.2	49.6	10.07	34.7	1.9	65.3	0.739	0.23
4	88.30	43.5	6.9	56.5	10.02	34.5	0.2	65.5	0.217	0.02
5	49.17	24.2	19.3	75.8	10.00	34.5	0.1	65.5	0.138	0.00
6	0.10	0.0	24.2	100.0	9.97	34.4	0.1	65.6	0.172	0.00

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos predictores e das respostas); $\%R_x$ e $\%R_y$ informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justifica a resposta em y ; $\sum(\%U_x)$ e $\sum(\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y/\%U_x$).

Da tabela anterior verifica-se que neste caso em concreto não se consegue obter uma boa descrição da resposta y7 - apenas se consegue descrever cerca de 65.6% da informação nela contida.

1. o primeiro factor latente ($FL1$) utiliza cerca de 24.5% da informação contida no sub-espaco dos predictores para descrever cerca de 56.6% da informação contida nesta resposta o que corresponde a uma eficiência de $\eta_1 = 2.31$, muito superior 1;
2. o segundo factor latente ($FL2$) requer 17.0% da informação contida nos predictores para descrever cerca de 6.8% da resposta possuindo uma baixa eficiência ($\eta_2 = 0.40$);
3. os restantes factores latentes apenas conseguem descrever cerca de 2.5% da informação residual da resposta ficando por descrever cerca de 34% desta.

As observações anteriores evidenciam que apenas o primeiro factor latente veicula informação pertinente, capaz de descrever alguma interdependência da resposta com o sub espaco dos predictores.

Na figura (4.10) encontram-se representados os scree plots referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

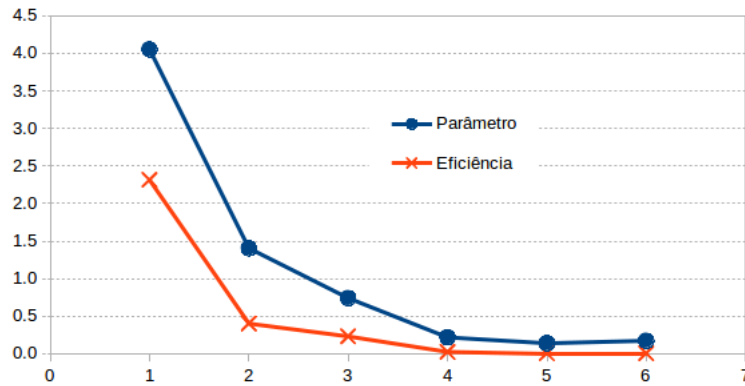


Figura 4.10: Scree plot dos parâmetros estimados pelo modelo e da eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Da figura anterior constata-se facilmente que as variáveis representadas (valores dos parâmetros do modelo e eficiência do factor latente) são concordantes e peremptórias em mostrar que apenas o primeiro factor latente, $FL1$, é relevante.

Na tabela(4.9) apenas se apresentam os pesos mais significativos sobre a resposta $y7$, atendendo à condição expressa pela equação (3.2) na página 58.

Tabela 4.9: Avaliação do impacto dos predictores mais relevantes sobre $y7$ através da análise das respectivas contribuições.

Predictor	FL1
x34	0.598
x37	-0.468
x40	0.391

FL1 - primeiro factor latente.

Atendendo aos valores da tabela (4.9) prevê-se a dependência de $y7$ com 3 predictores, designadamente: x34, x37 e x40.

Estas dependências serão avaliadas com base no ajuste explícito com polinómios de grau unitário.

4.2.2.3 Modelação explícita

A modelação explícita, realizada através da imposição directa de um modelo polinomial multivariado (M) de primeiro grau, (equação (3.3), na página 58), foi realizada através da abordagem matricial desenvolvida na secção 2.4.5, página 46.

Apesar da elevada interdependência observada anteriormente entre $Xe-Xf$ [0.999], atendendo que a dimensão do sub-espaco predictor é baixa ($M = 7$), numa primeira abordagem foram testadas todas as variáveis para servirem de predictores à modelação explícita não tendo ocorrido erros numéricos relacionados com o mau condicionamento de matrizes.

Segundo a estratégia da remoção sucessiva da variável cujo parâmetro possuía menor significado estatístico, foram realizadas 7 iterações - começando com um modelo completo ($p = 7$) convergiu-se para uma única variável ($p = 1$). Os resultados encontram-se sistematizados na tabela (4.10).

Tabela 4.10: Convergência iterativa do modelo polinomial obtido por MOLS; inicia-se o estudo com todas as variáveis (número de parâmetros $p = 7$) e vai-se rejeitando sucessivamente a variável que apresenta parâmetro com menor significado estatístico (que possui maior erro relativo). Os valores 1 e 0 associados às variáveis em estudo indicam a presença ou ausência do respectivo parâmetro no modelo polinomial.

Iteração	1	2	3	4	5	6	7
Xe	1	1	1	1	1	1	1
x34	1	1	1	1	1	0	0
x37	1	1	1	1	0	0	0
x38	1	0	0	0	0	0	0
x39	1	1	1	0	0	0	0
Xf	1	1	1	1	1	1	0
x40	1	1	0	0	0	0	0
Nm.	M07	M06	M05	M04	M03	M02	M01
p	7	6	5	4	3	2	1
SS	0.89	0.96	1.15	1.57	1.85	3.23	27.92
ν	23	24	25	26	27	28	29
σ_{fit}^2	0.039	0.040	0.046	0.060	0.068	0.115	0.963
ρ	0.984	0.983	0.980	0.973	0.968	0.943	0.193
$\rho = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.299
$\rho = 1$	0.675	0.657	0.620	0.553	0.511	0.371	0.000

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47); ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Neste caso o modelo parcimonioso encontrado apresenta $p = 5$ (M05) revelando a dependência de y_7 com Xe, x34, x37, x39 e Xf.

Da tabela anterior verificam-se também algumas curiosidades:

1. o primeiro modelo (M07) que é o mais completo ($p = 7$) é aquele que apresenta melhor desempenho quer na qualidade do ajuste (σ_{fit}^2) quer na descrição da resposta y_7 ($\rho = 0.984$) - a correlação múltipla atinge o valor de prova máximo (com $H_0 : |\rho| = 1$);
2. sendo o número de parâmetros relativamente baixo, o desempenho do modelo (σ_{fit}^2 e ρ) vai sofrendo deterioração à medida que se vai removendo parâmetros do modelo;
3. o parâmetro remanescente no modelo M01 é referente ao predictor Xe - é algo surpreendente esta conclusão já que a sua correlação com a resposta y_7 é menos eficaz que a que apresenta com x34 [-0.567] e x37 [0.444];
4. o modelo parcimonioso (M05, com $p = 5$) para y_7 que contém os preditores Xe, x34, x37, x39 e Xf ; nele estão preservadas as contribuições de

x34 e x37 identificados por análise directa de correlação;

5. a correlação múltipla encontrada com o modelo parcimonioso (M05, $\rho = 0.980$) é excelente já que o valor de prova para $\rho = 1$ é de $\alpha = 0.620$ enquanto que para $\rho = 0$ é de $\alpha = 0.000$, melhor que a que foi previamente detectada por análise directa de correlação.

A figura (4.11) apresenta a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação (ρ) em função do número de iteração do processo de pesquisa do modelo parcimonioso. Nela se encontra assinalado o modelo parcimonioso (M05) obtido na iteração $\# = 3$. Já a figura (4.12) apresenta a qualidade do ajuste conseguida com o modelo parcimonioso (M05).

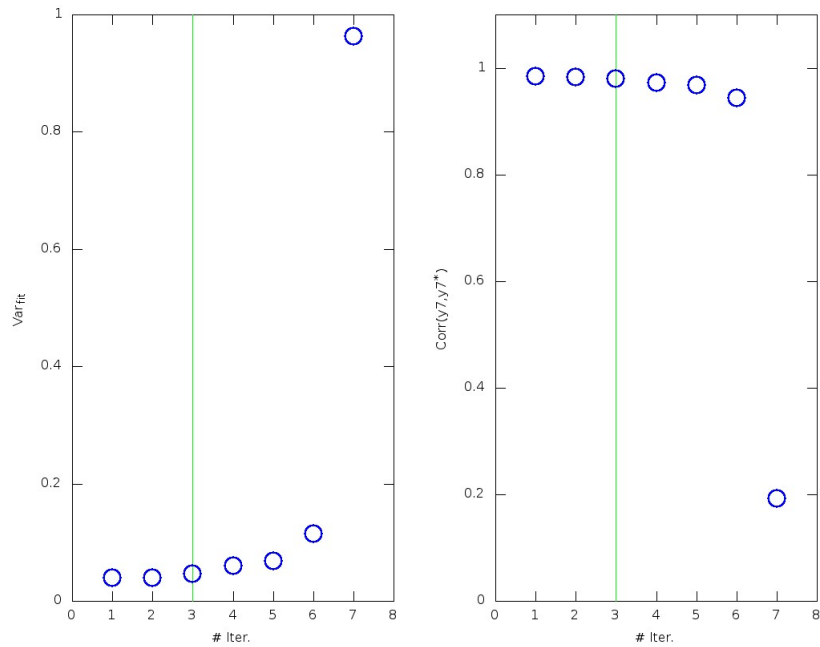


Figura 4.11: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa do melhor modelo. O traço vertical assinala o modelo parcimonioso.

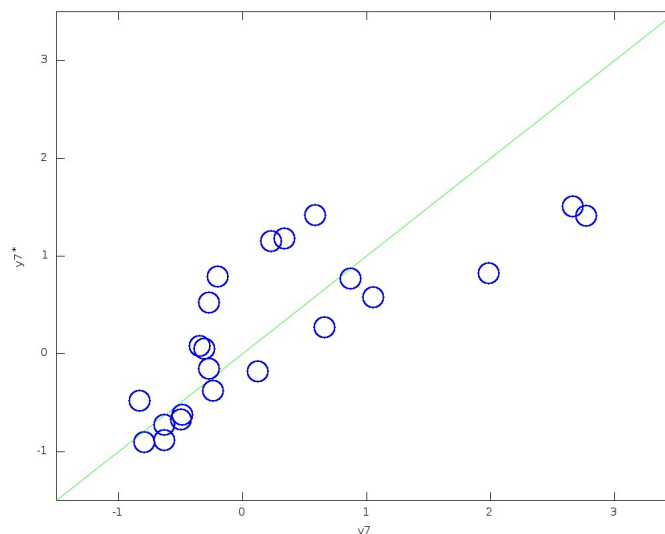


Figura 4.12: Capacidade de previsão do modelo - representação dos valores previstos pelo modelo parcimonioso (M05) em função da resposta y_7 ; a linha oblíqua representa a bissetriz dos quadrantes ímpares.

Da figura (4.11) verifica-se que neste caso em concreto em que a matriz de predictores possui uma dimensão baixa não estamos perante a situação de parametrização excessiva - não se verifica a instabilidade nas estimativas de desempenho dos modelos nem se atinge uma situação de modelação “exacta”. Assim, verifica-se que, à medida que vão sendo removidos parâmetros do modelo ele vai adquirindo maior rigidez passando a ter progressivamente maior dificuldade em acompanhar a variabilidade dos dados e, deste modo, perdendo na forma como prevê os valores (diminui a correlação) e aumentando a dispersão em relação aos valores reais.

Na figura (4.12) verifica-se uma boa relação entre os valores previstos pelo modelo parcimonioso (M05) e a resposta em causa (y_7).

Na tabela (4.11) encontram-se as estimativas dos parâmetros do modelo parcimonioso.

Tabela 4.11: Estimativas paramétricas obtidas com o modelo parcimonioso (M05).

Variável	β	$\sigma(\beta)$	%RSE	TV	α
Xe	23.4	1.2	5.3	18.98	0.000
Xf	-23.2	1.2	5.3	18.81	0.000
x34	-0.181	0.052	28.8	3.47	0.002
x37	0.166	0.051	30.7	3.26	0.003
x39	0.140	0.046	32.8	3.05	0.005

β - valor do parâmetro estimado, equação (2.49), pág 48; $\sigma(\beta)$ - erro padrão do parâmetro, equação (2.55), pág.49 ; %RSE - erro padrão relativo, equação (2.49), pág 48 ; TV e α - valor de teste e respectivo valor de prova referente ao teste de significado estatístico do parâmetro considerando a hipótese nula $H_o : |\beta| = 0$.

Da tabela verifica-se que todos os parâmetros encontrados têm significado estatístico ($\alpha < 0.03$ assumindo $H_0 : |\beta| = 0$).

Verifica-se ainda que as variáveis Xe e Xf possuem uma importância relativa muito elevada, comparativamente às restantes. Atendendo ainda à interdependência encontrada anteriormente entre Xe e Xf [0.999] pensou-se que estas variáveis estivessem a ser utilizadas para descrever algum desvio compensatório encontrado entre o modelo M05 e a resposta y7.

Para testar esta hipótese removeu-se uma delas (Xf) do conjunto inicial de predictores e reprocessou-se o método iterativo de pesquisa do modelo parcimonioso tendo culminado num modelo com $p = 3$ que usava exclusivamente Xe, x34 e x40 mas que apresentava nitidamente piores desempenhos ($\sigma_{fit}^2 = 0.401$, $\rho = 0.791$) quando comparados com os valores obtidos por M05, ver tabela (4.10).

Na tabela (4.12) encontra-se uma compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (PLS, representada por w) e modelação explícita (MOLS, representada por β).

Tabela 4.12: Compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (w) e modelação explícita (β).

Variável	Xe	Xf	x34	x37	x38	x39	x40
ρ			-0.57	0.44			
w			0.60	-0.47			0.39
β	23.4	-23.2	-0.18	0.17		0.14	

ρ - estimativas obtidas por análise de correlação; w - contribuição evidenciada por PLS; β - efeito obtido por modelação explícita através do modelo parcimonioso (M05).

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso representa a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação directa com y7 identificou 2 dos 5 predictores mais influentes revelando ainda o respectivo sentido de variação.

Já a modelação implícita (PLS) aponta para 3 contribuições relevantes, sugeridas pelo primeiro factor latente ($FL1$), das quais 2 correspondem a factores identificados pela modelação explícita (MOLS) embora com sinais opostos nas respectivas contribuições. Neste caso a deficiente modelação sugerida por PLS apresenta uma sugestão errada que não se revela em MOLS a não ser que se suprima um dos predictores mais relevantes (ex: Xf).

Olhando agora às estimativas obtidas com o modelo parcimonioso (M05) transtorna-nos um pouco o facto de este estar a dar uma contribuição relativa muito intensa às variáveis Xe e Xf e ainda o facto de estas possuírem sinais opostos, sugerindo que o método de mínimos quadrados estaria a tentar descrever a resposta com base num detalhe residual resultante da diferença Xe-Xf.

No sentido de averiguar melhor este fenómeno fez-se a representação de Xe, Xf, Xe-Xf e dos valores previstos pelo modelo ($y7^*$) em função da resposta y7, ver figura (4.13).

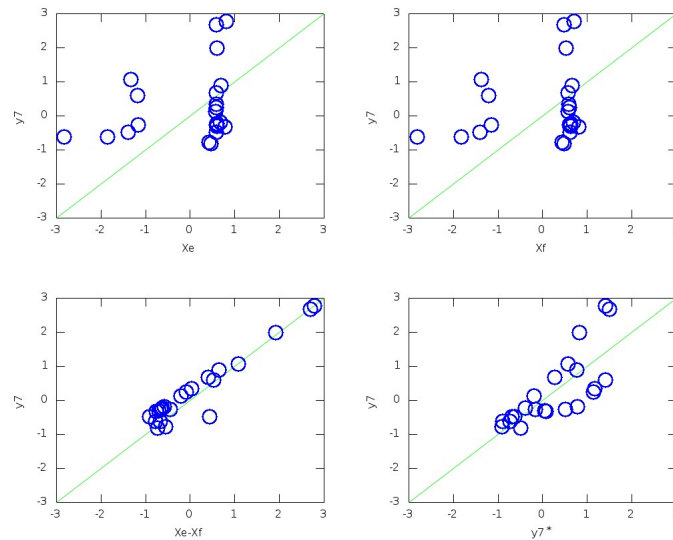


Figura 4.13: Verificação da contribuição das variáveis X_e , X_f e $X_e - X_f$ sobre a qualidade de ajuste em relação a y_7 por comparação com os valores previstos pelo modelo parcimonioso M05.

Os resultados exibidos pela figura anterior evidenciam a relevância do termo $X_e - X_f$ ($p = 2$) na explicação da resposta y_7 - esta contribuição é de facto tão relevante que o ajuste exclusivo de $X_e - X_f$ a y_7 parece ser melhor que o que é conseguido com o modelo parcimonioso (M05).

Voltando aos valores da tabela (4.10) e olhando para a figura (4.11), o modelo M02 ($p = 2$) corresponde ao que só contém informação relativa a X_e e X_f e este não é certamente o melhor modelo em termos de estimativa paramétrica (apresenta maior erro e menor correlação) e parcimonia, evidenciando que as restantes variáveis detectadas em M05 (x_{34} , x_{37} e x_{39}) são também importantes para descrever y_7 .

Deste estudo resulta que a resposta y_7 depende de x_e , x_f , x_{34} , x_{37} e x_{39} corresponde ao modelo definido como parcimonioso (M05) e em que todos os parâmetros possuem significado estatístico ($H_0 : |\beta_i| = 0$ com $\nu = 25$).

Da tabela (4.12) verifica-se que as variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y_7 são X_e [23.4], x_{37} [0.166] e x_{39} [0.140] - a dependência positiva indica que estas possuem uma influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio.

Em contrapartida, as variáveis X_f [-23.2] e x_{34} [-0.181] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y_7 indicam que o seu aumento têm como impacto a redução desse valor.

4.2.3 Resposta y1

Neste caso a matriz dos predictores é também muito exígua - contém apenas 26 objectos ($N = 26$) representados sobre 26 dimensões ($M = 26$, variáveis Xa, x1, x2, x3, x5, x8, x9, x10, x12, x14, x15, x16, x17, x18, x21, Xb, Xc, Xd, Xe, x34, x37, x39, Xf, x40) apresenta as dimensões $X_{(26 \times 26)}$.

4.2.3.1 Análise de correlação

Com $N = 26$, impondo o critério $\alpha = 0.03$, ver secção (3.3) na página 57, verificou-se quais os valores limites de correlação para $|\rho| = 0$ e $|\rho| = 1$, obteve-se os valores 0.427 e 0.635 indicando que valores absolutos de correlação inferiores a 0.427 são nulos enquanto que valores superiores a 0.635 considerados unitários.

Estudando a correlação interna ao sub-espaco predictor foram detectadas elevadas interdependências [0.999] entre Xa, Xb, Xc, Xd, Xe e Xf e também entre x15-x8 [1.000]. Tem-se de tomar atenção especial com estas variáveis no caso do ajuste polinomial explícito (MOLS) (ver secção (2.4.5), na página 46)..

Atendendo aos valores de aceitação e rejeição encontrados para a correlação conclui-se que x1 [-0.605], x2 [0.519], x5 [-0.514], x12 [0.563], x37 [0.620] e x39 [-0.450] possuem apenas correlações não nulas. Na figura (4.14) encontram-se representadas as dependências de encontradas com y1.

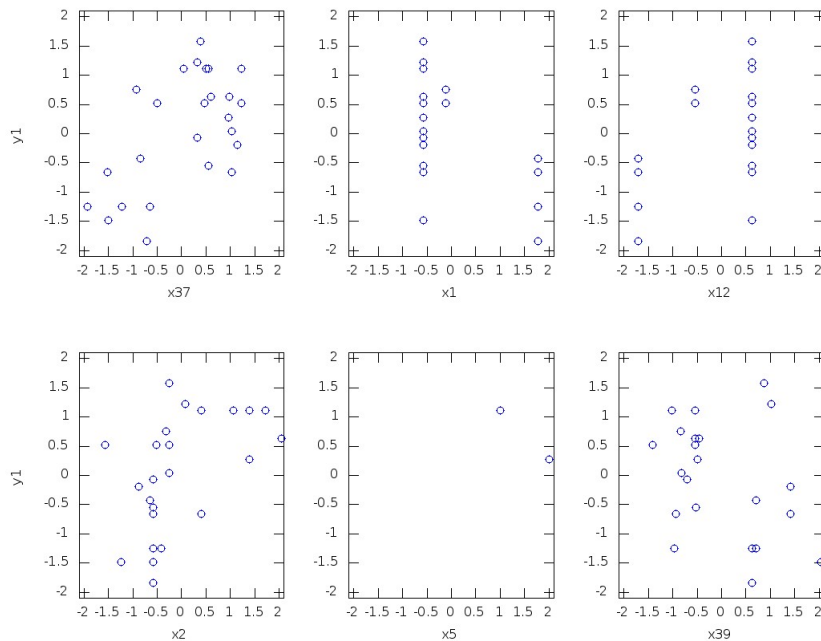


Figura 4.14: Representação das interdependências de y1 com x1 [-0.605], x2 [0.519], x5 [-0.514], x12 [0.563], x37 [0.620] e x39 [-0.450].

Da figura anterior verifica-se que as interdependências detectadas como não nulas apresentam uma variação mútua com y_1 pouco evidente em alguns casos.

4.2.3.2 Abordagem PLS

De modo similar fez-se o estudo da modelação implícita da resposta y_1 com o algoritmo PLS. O respectivos resultados encontram-se sistematizados na tabela (4.13).

Tabela 4.13: Resultados obtidos na abordagem da resposta y_1 através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum(\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum(\%J_y)$	β_i	η_i
0	650.00	100.0	—	—	25.00	100.0	—	—	—	—
1	479.84	73.8	26.2	26.2	11.24	45.0	55.0	55.0	3.709	2.10
2	311.13	47.9	26.0	52.1	5.95	23.8	21.2	76.2	2.301	0.82
3	253.17	38.9	8.9	61.1	3.75	15.0	8.8	85.0	1.481	0.98
4	209.88	32.3	6.7	67.7	2.92	11.7	3.3	88.3	0.910	0.50
5	168.78	26.0	6.6	74.0	2.50	10.0	1.7	90.0	0.654	0.27
6	146.22	22.5	3.5	15.5	0.00	8.0	2.0	92.0	0.706	0.57
7	120.16	18.5	4.0	81.5	1.75	7.0	1.0	93.0	0.495	0.24
8	94.40	14.5	4.0	85.5	1.63	6.5	0.5	93.5	0.345	0.12
9	72.29	11.1	3.4	88.9	1.51	6.0	0.5	94.0	0.349	0.14
10	47.58	7.3	3.8	92.7	1.42	5.7	0.4	94.3	0.297	0.09

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos predictores e das respostas); $\%R_x$ e $\%R_y$ informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justificar a resposta em y ; $\sum(\%U_x)$ e $\sum(\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y / \%U_x$).

Da tabela anterior verifica-se que neste caso consegue-se obter uma aceitável descrição da resposta y_1 com os dois primeiros factores latentes - estes descrevem cerca de 76.2% da informação contida em y_1 embora a eficiência de modelação, no segundo caso, seja inferior a 1.

1. o primeiro factor latente ($FL1$) utiliza cerca de 26.2% da informação contida no sub-espaco dos predictores para descrever cerca de 55.0% da informação contida nesta resposta o que corresponde a uma eficiência de $\eta_1 = 2.10$, superior a 1;
2. o segundo factor latente ($FL2$) utiliza cerca de 26.0% da informação contida nos predictores para descrever cerca de 21.2% da resposta possuindo uma baixa eficiência ($\eta_2 = 0.82$), embora próxima de 1;
3. após estes dois factores latentes removidos ficam ainda por descrever 23.8% da informação contida na resposta em causa.

Na figura (4.15) encontram-se representados os scree plots referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

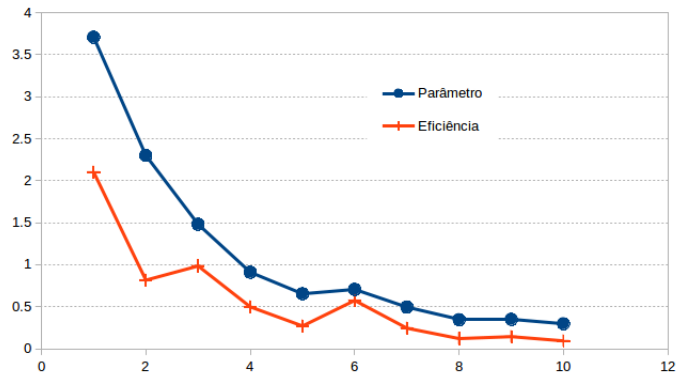


Figura 4.15: Scree plot dos parâmetros estimados pelo modelo e da eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Da figura anterior constata-se que há alguma dificuldade em estabelecer, com base no scree plot, a definição do número de factores latentes relevantes - o primeiro é sem dúvida o mais relevante mas não se percebe muito bem se basta considerar $f = 2$ ou $f = 3$ factores latentes para descrever esta resposta. Assim sendo e recorrendo à tabela (4.13), com $f = 2$ apenas se consegue descrever cerca de 76.2% da resposta enquanto que com $f = 3$ já se tem 85.0% dessa informação. Contudo, olhando ao acréscimo de informação justificada em y_1 , a informação contida em $FL2$ (21.2%) ultrapassa o dobro da contida em $FL3$ (8.8%). Vamos então assumir que $f = 2$.

Na tabela (4.14) apenas se apresentam os pesos mais significativos sobre a resposta y_1 , atendendo à condição expressa pela equação (3.2) na página 58.

Tabela 4.14: Avaliação do impacto dos predictores mais relevantes sobre y_1 através da análise das respectivas contribuições.

Predictor	FL1	FL2
Xa		0.302
x1	-0.384	
x2	0.329	
x3	0.260	
x5	-0.327	
x10		0.242
x12	0.357	
x16		0.222
x17	0.233	
x19		0.206
x21		0.268
Xb		0.305
Xc		0.297
Xd		0.305
Xe		0.300
x37	0.394	
x39	-0.286	0.245
Xf		0.303

FL1 - primeiro factor latente; FL2 - segundo factor latente.

Atendendo aos valores da tabela (4.14) prevê-se a dependência de y_1 com 18 predictores, designadamente: Xa, x1, x2, x3, x5, x10, x12, x16, x17, x19, x21, Xb, Xc, Xd, Xe, x37, x39 e Xf.

Estas dependências serão avaliadas com base no ajuste explícito com polinómios de grau unitário.

4.2.3.3 Modelação explícita

A modelação explícita, realizada através da imposição directa de um modelo polinomial multivariado (M) de primeiro grau, (equação (3.3), na página 58), foi realizada através da abordagem matricial desenvolvida na secção 2.4.5, página 46.

Atendendo à elevada interdependência observada anteriormente entre Xa, Xb, Xc, Xd, Xe e Xf e também entre x15-x8, não foi possível efectuar o ajuste explícito com todas estas 26 variáveis mas, após a remoção de Xb, Xc, Xd, Xe e Xf isso já foi possível.

Seguindo a estratégia da remoção sucessiva da variável cujo parâmetro possuía menor significado estatístico, foram realizadas 21 iterações - começando com um modelo completo ($p = 21$) convergiu-se para uma única variável ($p = 1$). Os resultados encontram-se sistematizados na tabela (4.15).

Tabela 4.15: Convergência iterativa do modelo polinomial obtido por MOLS; inicia-se o estudo com todas as variáveis (número de parâmetros $p = 21$) e vai-se rejeitando sucessivamente a variável que apresenta parâmetro com menor significado estatístico (que possui maior erro relativo). Os valores 1 e 0 associados às variáveis em estudo indicam a presença ou ausência do respectivo parâmetro no modelo polinomial.

Iteração	1	10	11	12	13	14	15	16	21
Xa	1	1	1	1	1	1	1	1	0
x1	1	1	1	0	0	0	0	0	0
x2	1	0	0	0	0	0	0	0	0
x3	1	0	0	0	0	0	0	0	0
x5	1	1	1	1	1	1	1	1	1
x8	1	1	1	1	1	1	1	1	0
x9	1	0	0	0	0	0	0	0	0
x10	1	1	1	1	1	1	1	1	0
x12	1	1	0	0	0	0	0	0	0
x14	1	0	0	0	0	0	0	0	0
x15	1	0	0	0	0	0	0	0	0
x16	1	1	1	1	1	1	1	1	0
x17	1	1	1	1	1	1	1	0	0
x18	1	0	0	0	0	0	0	0	0
x19	1	1	1	1	0	0	0	0	0
x20	1	1	1	1	1	0	0	0	0
x21	1	1	1	1	1	1	1	1	0
x34	1	0	0	0	0	0	0	0	0
x37	1	0	0	0	0	0	0	0	0
x39	1	0	0	0	0	0	0	0	0
x40	1	1	1	1	1	1	0	0	0
Nm.	M21	M12	M11	M10	M09	M08	M07	M06	M01
p	21	12	11	10	9	8	7	6	1
SS	0.91	1.31	1.91	1.97	2.55	3.53	4.24	5.93	18.38
ν	5	14	15	16	17	18	19	20	25
σ_{fit}^2	0.183	0.094	0.127	0.123	0.150	0.196	0.223	0.297	0.735
ρ	0.982	0.973	0.961	0.960	0.948	0.927	0.911	0.873	0.514
$\rho = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006
$\rho = 1$	0.838	0.670	0.593	0.575	0.508	0.419	0.359	0.258	0.009

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47), ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Neste caso o modelo parcimonioso encontrado apresenta $p = 9$ (M09) revelando a dependência de y_1 com Xa, x5, x8, x10, x16, x17, x20, x21 e x40.

Na figura (4.16) apresenta a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação (ρ) em função do número de parâmetros do modelo evidenciando o modelo parcimonioso (M09). Já a figura (4.17) apresenta a qualidade do ajuste conseguida com o modelo parcimonioso (M09).

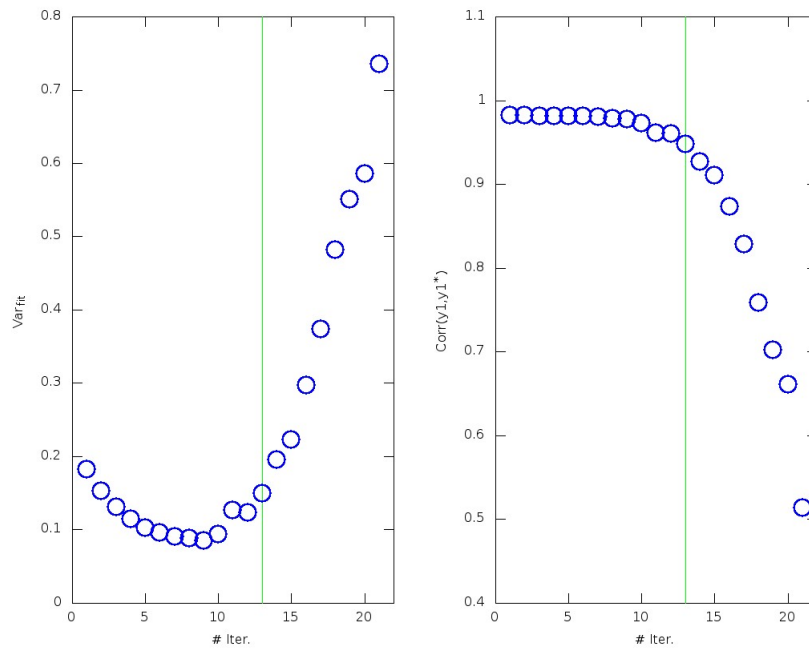


Figura 4.16: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa iterativa do melhor modelo. O traço vertical assinala a iteração em que se obteve o modelo parcimonioso (M09).

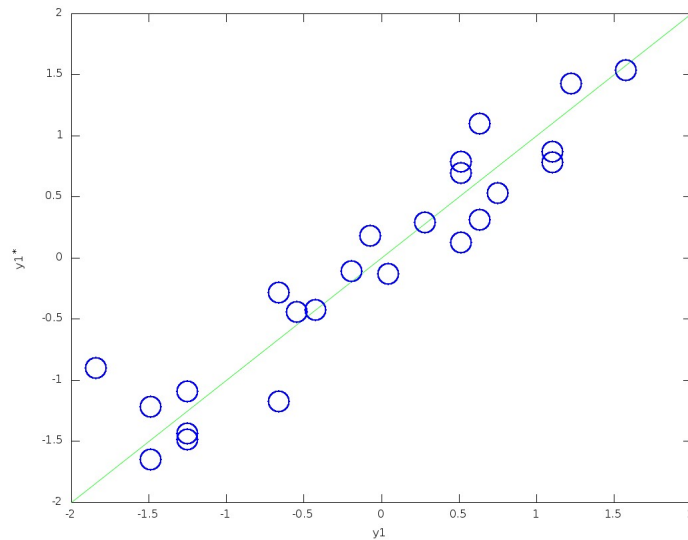


Figura 4.17: Capacidade de previsão do modelo - representação dos valores previstos ($y1^*$) pelo modelo parcimonioso (M09) em função da resposta $y1$; a linha oblíqua representa a bissetriz dos quadrantes ímpares.

Da figura (4.16) verifica-se que, neste caso concreto em que a matriz de preditores possui uma dimensão relativamente alta, se faz sentir uma nova situação de parametrização excessiva - a variância do ajuste decresce com a remoção de parâmetros atingindo um valor mínimo com $p = 13$; nesta situação a correlação ainda é superior ao valor 0.98. Continuando a remover parâmetros o erro do ajuste vai aumentando e a correlação assume uma queda progressiva. Com $p = 9$ atinge-se o melhor compromisso estatístico entre um bom desempenho do modelo e a maximização do número de graus de liberdade (atinge-se o modelo parcimonioso); daí em diante, a situação da qualidade do ajuste deteriora-se rapidamente, ver figura (4.16).

Na tabela (4.16) encontram-se as estimativas dos parâmetros do modelo parcimonioso.

Tabela 4.16: Estimativas paramétricas obtidas com o modelo parcimonioso (M09).

Variável	β	$\sigma(\beta)$	%RSE	TV	α
Xa	0.62	0.15	23.6	4.24	0.000
x5	-0.78	0.13	16.2	6.17	0.000
x8	-0.95	0.30	31.7	3.16	0.004
x10	-1.22	0.29	24.1	4.15	0.000
x16	0.63	0.20	31.1	3.22	0.004
x17	0.39	0.09	22.4	4.47	0.000
x20	0.22	0.09	39.3	2.55	0.017
x21	-0.59	0.14	23.8	4.20	0.000
x40	0.25	0.09	36.4	2.75	0.011

β - valor do parâmetro estimado, equação (2.49), pág 48; $\sigma(\beta)$ - erro padrão do parâmetro, equação (2.55), pág.49 ; %RSE - erro padrão relativo, equação (2.49), pág 48 ; TV e α - valor de teste e respectivo valor de prova referente ao teste de significado estatístico do parâmetro considerando a hipótese nula $H_o : |\beta| = 0$.

Da tabela (4.16) pode-se constatar que todos os parâmetros encontrados têm significado estatístico ($\alpha < 0.03$ assumindo $H_0 : |\beta| = 0$). Verifica-se também que as variáveis preservadas no modelo parcimonioso (M09) possuem contribuições relativamente equitativas - entre a maior contribuição (1.22) e a menor (0.22) a diferença é apenas da ordem de 5 vezes.

Na tabela (4.17) encontra-se uma compilação dos resultados obtidos por análise de correlação, modelação implícita (PLS) e modelação explícita (MOLS).

Tabela 4.17: Compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (w) e modelação explícita (β).

Variável	Xa	Xb	Xc	Xd	Xe	Xf	x1	x2	x3	x5	x8
ρ							-0.61	0.52		-0.51	
w	0.30	0.31	0.30	0.30	0.30	0.30	-0.38	0.33	0.26	-0.33	
β	-0.62	-0.62*	-0.62*	-0.62*	-0.62*	-0.62*				-0.78	-0.95
Variável	x10	x12	x16	x17	x19	x20	x21	x37	x39	x40	
ρ		0.56						0.11	-0.45		
w	0.24	0.36	0.22	0.23	0.21		0.27	0.39	-0.29		
β	-1.22		0.63	0.39		0.22	-0.59			0.25	

ρ - estimativas obtidas por análise de correlação; w - contribuição evidenciada por PLS; β - efeito obtido por modelação explícita através do modelo parcimonioso (M09).

* As contribuições de Xb, Xc, Xd, Xe e Xf foram acrescidas a MOLS devido à sua forte interdependência com Xa respectivamente.

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso representa a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação directa com y1 identificou apenas 1 dos 14 predictores mais influentes. Já a modelação implícita (PLS) aponta para 18 predictores dos quais 8 correspondem ao primeiro factor latente (FL1).

Deste estudo resulta que a resposta y1 depende de Xa, x5, x8, x10, x16, x17, x20, x21 e x40 - corresponde ao modelo definido como parcimonioso (M09) e em que todos os parâmetros possuem significado estatístico ($H_0 : |\beta_i| = 0$ com $\nu = 17$).

Devido à interdependência estreita entre X_a - X_b - X_c - X_d - X_e - X_f , as variáveis de estudo (X_b , X_c , X_d , X_e e X_f) serão agora incluídas nas conclusões.

Da tabela (4.17) verifica-se que as variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y_1 são x_{16} [0.63], x_{17} [0.39], x_{40} [0.25] e x_{20} [0.22] - a dependência positiva indica que estas possuem uma influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio.

Em contrapartida, as variáveis x_{21} [-0.59], $X_a/X_b/X_c/X_d/X_e/X_f$ [-0.62], x_5 [-0.78], x_8 [-0.95] e x_{10} [-1.22] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y_1 indicam que o seu aumento têm como impacto a redução desse valor.

4.2.4 Resposta y_{12}

Neste caso em concreto a matriz dos predictores é também muito exígua - contém apenas 20 objectos ($N = 20$) representados sobre 36 dimensões ($M = 36$, variáveis x_{55} , x_{56} , x_{57} , x_{58} , x_{59} , x_{60} , x_{61} , x_{62} , x_{63} , x_{64} , x_{65} , x_{66} , x_{67} , x_{68} , x_{69} , x_{70} , x_{75} , x_{76} , x_{77} , x_{78} , x_{79} , x_{80} , x_{81} , x_{82} , x_{84} , x_{85} , x_{86} , x_{87} , x_{91} , x_{92} , x_{93} , x_{101} , x_{102} , x_{103} , x_{107} , x_{108} variáveis) apresenta as dimensões $X_{(20 \times 36)}$.

4.2.4.1 Análise de correlação

Com $N = 20$, a correlação será considerada nula ($H_0 : |\rho| = 0$) se o módulo de correlação for inferior a 0.485 (indicando que não existe correlação significativa) enquanto que valores superiores a 0.530 indicam uma elevada correlação (valor considerado unitário). Já os valores intermédios indicam que apenas que estes não são nulos.

Estudando a correlação interna ao sub-espço predictor verificou-se a igualdade ($\rho = 1$) entre os predictores $x_{59} = x_{58}$, $x_{65} = x_{64}$ e $x_{103} = x_{102}$.

Impondo o valor limite de $\rho \geq 0.485$ no estudo da interdependência entre o sub-espço predictor e a resposta y_{12} , detectou-se apenas a forte interdependência com x_{84} [0.621].

Na figura (4.18) encontra-se representada a dependência de y_{12} com x_{84} .

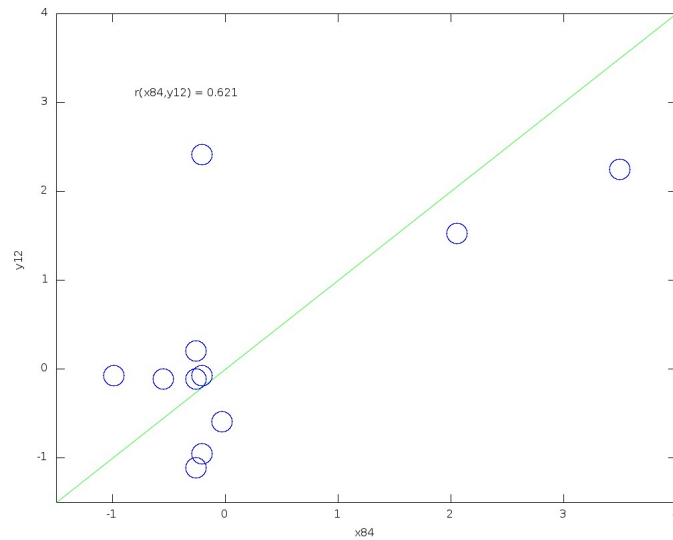


Figura 4.18: Representação das interdependências de y12 com x84 [0.621].

A análise directa da correlação evidenciou uma única correlação directa y12-x84 [0.621] como sendo muito relevante (similar à unidade). Da figura anterior verifica-se que, apesar da má consolidação da informação, parece de facto fazer sentido estatístico em haver alguma relação directa entre y12 e x84.

4.2.4.2 Abordagem PLS

Dado que a modelação implícita através do algoritmo PLS é suficientemente robusta para suportar a presença de variáveis redundantes, a matriz original de predictores ($X_{(20 \times 36)}$) foi fornecida como sub-espaco predictor para tentar descrever a resposta y12.

Na tabela 4.18 encontram-se as estimativas obtidas da análise PLS em relação à resposta y12.

Tabela 4.18: Resultados obtidos na abordagem da resposta y12 através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum (\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum (\%J_y)$	β_i	η_i
0	684.00	100.0	—	—	19.00	100.0	—	—	—	—
1	568.52	83.1	16.9	16.9	7.55	39.8	60.2	60.2	3.383	3.57
2	479.58	70.1	13.0	29.9	2.85	15.0	24.7	85.0	2.168	1.90
3	363.05	53.1	17.0	46.9	1.24	6.5	8.5	93.5	1.271	0.50
4	337.57	49.4	3.7	50.6	0.34	1.8	4.7	98.2	0.950	1.27
5	255.23	37.3	12.0	62.7	0.26	1.4	0.4	98.6	0.273	0.03
6	225.41	33.0	4.4	67.0	0.10	0.5	0.8	99.5	0.397	0.19
7	166.19	24.3	8.7	75.7	0.05	0.3	0.3	99.7	0.229	0.03
8	133.38	19.5	4.8	80.5	0.01	0.0	0.2	100.0	0.206	0.05
9	37.58	5.5	14.0	94.5	0.00	0.0	0.0	100.0	0.079	0.00
10	22.87	3.3	2.1	96.7	0.00	0.0	0.0	100.0	0.049	0.01

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos preditores e das respostas); $\%R_x$ e $\%R_y$ - informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justificar a resposta em y ; $\sum (\%U_x)$ e $\sum (\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y / \%U_x$).

Da tabela anterior verifica-se que com apenas dois factores latentes ($f = 2$) consegue-se descrever 85.0% da resposta em causa:

1. o primeiro factor latente ($FL1$) utiliza cerca de 16.9% da informação contida no sub-espaco dos preditores para descrever cerca de 60.2% da informação contida nesta resposta o que corresponde a uma eficiência de $\eta_1 = 3.57$, muito superior a 1;
2. O segundo factor latente ($FL2$) requer 13.0% da informação contida nos preditores para descrever cerca de 24.7% da resposta possuindo também uma eficiência bastante elevada ($\eta_2 = 1.90$);
3. após a remoção do efeito dos dois primeiros factores latentes a informação residual da resposta por modelar passa a ser de 15.0%.

As observações anteriores evidenciam que apenas os dois primeiros factores latentes veiculam informação pertinente, capaz de descrever a interdependência da resposta com o sub espaco dos preditores.

Na figura (4.19) encontram-se representados os scree plots referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

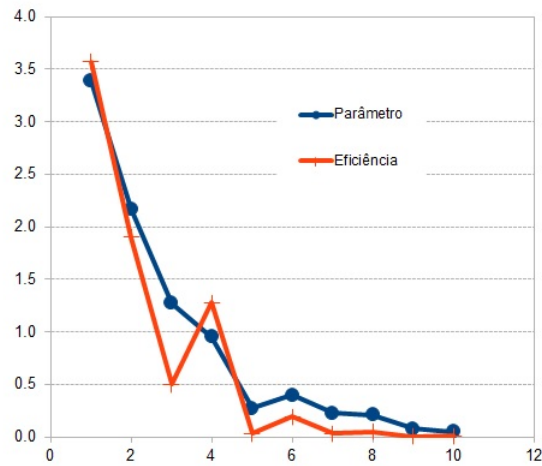


Figura 4.19: Scree plot dos parâmetros estimados pelo modelo e a eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Da figura anterior constata-se facilmente que as variáveis representadas (valores dos parâmetros do modelo e eficiência do factor latente) são concordantes e peremptórias em mostrar que apenas os dois primeiros factores latentes, $FL1$ e $FL2$, é que são relevantes.

Importa agora avaliar a informação contida em $FL1$ e $FL2$ em termos de representatividade dos predictores sobre a resposta.

Na tabela (4.19) apenas se apresentam os pesos mais significativos sobre a resposta y_{12} , atendendo à condição expressa pela equação (3.2) na página 58.

Tabela 4.19: Avaliação do impacto dos predictores mais relevantes sobre y_{12} através da análise das respectivas contribuições.

Predictor	FL1	FL2
x55	0.188	0.185
x56	-0.252	
x57		-0.188
x58	-0.189	0.307
x59	-0.189	0.307
x60	0.258	-0.262
x68	0.172	
x75	0.248	
x76	0.232	
x79		0.257
x80	0.214	
x82	-0.273	
x84	0.360	0.286
x86		0.254
x87		0.176
x93	-0.200	
x101	0.250	-0.263
x102		-0.232
x103		-0.232
x107	-0.255	
x108		-0.150

FL1 - primeiro factor latente; FL2 - segundo factor latente.

Atendendo a que ambos factores apresentam eficiências relevantes, da tabela (4.18) estamos a prever a dependência de y_{12} com 21 predictores, designadamente: x55, x56, x57, x58, x59, x60, x68, x75, x76, x79, x80, x82, x84, x86, x87, x93, x101, x102, x103, x107 e x108.

Estas dependências serão avaliadas com base no ajuste explícito com polinómios de grau unitário.

4.2.4.3 Modelação explícita

Dado que as variáveis x59, x65 e x103 são linearmente dependentes das x58, x64 e x102, estas foram inicialmente removidas do estudo de modelação explícita, tendo a matriz de predictores passado a $X_{(20 \times 33)}$. Esta simplificação ainda assim não foi suficiente já que a matriz de predictores possuía demasiadas variáveis ($M = 33$) relativamente ao número de pontos experimentais ($N = 20$) o que dificulta a estimativa paramétrica. A resolução MOLS deste sistema conduziu a valores imaginários.

Como aproximação inicial, foi-se removendo sucessivamente variáveis que apresentavam maiores correlações internas, em módulo, tendo sido removidas x70 ($\simeq x66$), x101 ($\simeq x60$), x107 ($\simeq x56$), x60 ($\simeq x58$), x79 ($\simeq x55$) sem se ter tido sucesso na estimativa MOLS em qualquer um dos casos.

Ao verificar que a única variável detectada como relevante através da análise de correlação directa (x84) estava representada no conjunto de predictores identificadas com o PLS ($M = 21$), e atendendo à experiência anterior em

que foi possível utilizar o sub-espço predictor na íntegra, suspeitou-se que o conjunto de predictores evidenciado por PLS seria suficiente para procurar descrever a resposta. Assim, usando o subconjunto de variáveis identificadas por PLS ($M = 21$) e removendo as 6 mais correlacionadas (x59, x60, x79, x101, x103 e x107) passou-se a um subconjunto mais restrito ($M = 15$, variáveis x55, x56, x57, x58, x68, x75, x76, x80, x82, x84, x86, x87, x93, x102 e x108) o qual já foi possível processar por ajuste explícito e utilizar a estratégia da remoção sucessiva da variável cujo parâmetro possuía menor significado estatístico com vista a identificar o modelo parcimonioso.

Os resultados encontram-se sistematizados na tabela (4.20).

Tabela 4.20: Convergência iterativa do modelo polinomial obtido por MOLS; inicia-se o estudo com todas as variáveis (número de parâmetros $p = 15$) e vai-se rejeitando sucessivamente a variável que apresenta parâmetro com menor significado estatístico (que possui maior erro relativo). Os valores 1 e 0 associados às variáveis em estudo indicam a presença ou ausência do respectivo parâmetro no modelo polinomial.

Iter	1	2	4	6	7	8	9	10	11	12	15
x55	1	1	1	1	1	1	1	1	1	1	0
x56	1	1	1	0	0	0	0	0	0	0	0
x57	1	1	0	0	0	0	0	0	0	0	0
x58	1	0	0	0	0	0	0	0	0	0	0
x68	1	1	1	1	1	1	1	1	0	0	0
x75	1	1	1	1	0	0	0	0	0	0	0
x76	1	1	1	1	1	1	1	1	1	1	0
x80	1	1	0	0	0	0	0	0	0	0	0
x82	1	1	1	1	1	0	0	0	0	0	0
x84	1	1	1	1	1	1	1	1	1	1	1
x86	1	1	1	1	1	1	1	1	1	1	0
x87	1	1	1	1	1	1	1	1	1	0	0
x93	1	1	1	1	1	1	1	0	0	0	0
x102	1	1	1	0	0	0	0	0	0	0	0
x108	1	1	1	1	1	1	0	0	0	0	0
Nm.	M15	M14	M12	M10	M09	M08	M07	M06	M05	M04	M01
p	15	14	12	10	9	8	7	6	5	4	1
SS	4.91	23.80	0.00	0.01	0.04	0.46	0.63	1.14	2.26	2.78	11.68
ν	5	6	8	10	11	12	13	14	15	16	19
σ_{fit}^2	0.982	3.967	0.000	0.001	0.004	0.039	0.048	0.081	0.151	0.174	0.615
ρ	0.905	0.820	1.000	1.000	0.999	0.988	0.983	0.970	0.939	0.924	0.621
$\rho = 0$	0.005	0.013	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
$\rho = 1$	0.638	0.470	—	0.969	0.941	0.790	0.746	0.649	0.501	0.438	0.049

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47), ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Da tabela anterior verificam-se algumas particularidades:

1. o segundo modelo (M14) que é o segundo mais completo e apresenta o pior desempenho quer na qualidade do ajuste (σ_{fit}^2) quer na descrição da

resposta y_{12} (ρ);

2. em termos de desempenho dos modelos (σ_{fit}^2 e ρ), estes estimadores começam mal, atingem um valor óptimo no caso do modelo exacto (M12) e depois vão piorando progressivamente até que terminam assumindo valores inadequados;
3. o parâmetro remanescente no modelo M01 é referente ao predictor x_{84} - coincide com o predictor mais correlacionado com y_{12} ;
4. o modelo parcimonioso (M07 com $p = 7$) para y_{12} contém os predictores x_{55} , x_{68} , x_{76} , x_{84} , x_{86} , x_{87} e x_{93} ; a correlação múltipla encontrada ($\rho = 0.983$) é excelente já que o valor de prova para $\rho = 1$ é de $\alpha = 0.746$ enquanto que para $\rho = 0$ é de $\alpha = 0.000$.

Na figura (4.20) apresenta-se a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação múltipla (ρ) em função do índice de iteração do processo de pesquisa do melhor modelo - a iteração 9 está assinalada com um traço vertical para evidenciar o desempenho do modelo parcimonioso (M07). Já a figura (4.21) apresenta a qualidade do ajuste conseguida com o modelo parcimonioso (M07).

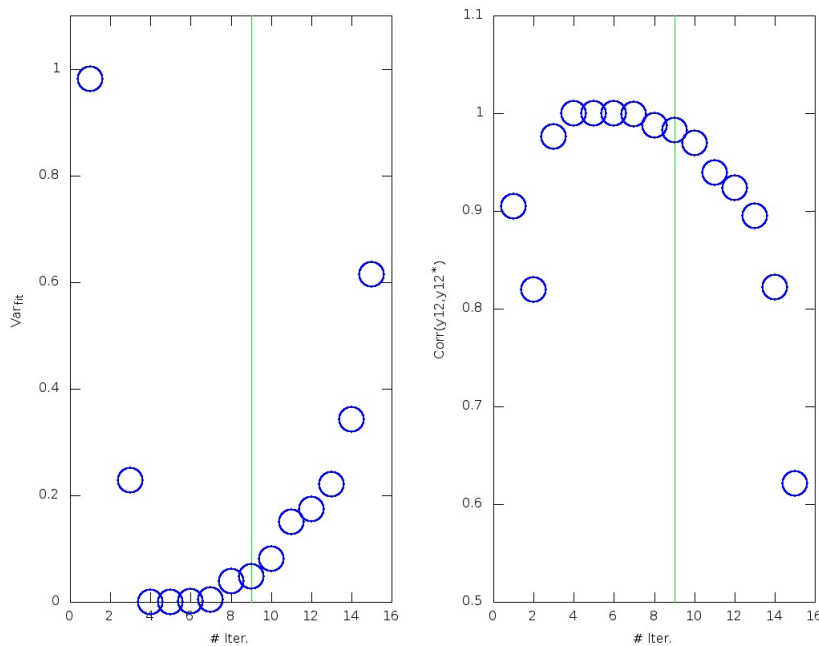


Figura 4.20: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa do melhor modelo. O traço vertical assinala o modelo parcimonioso.

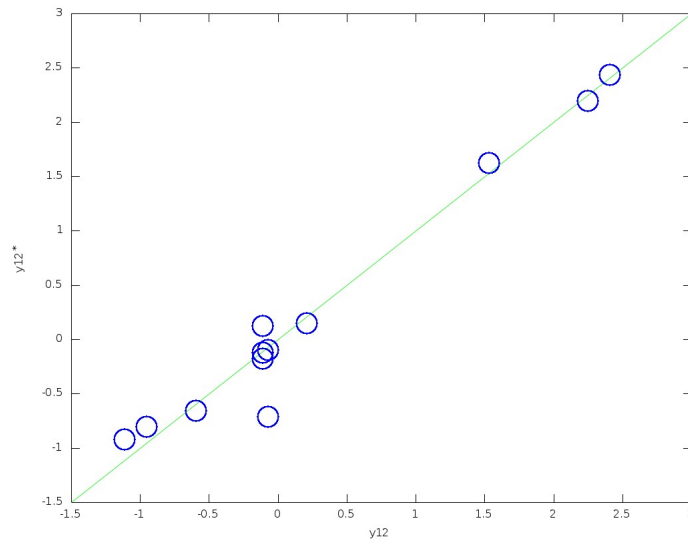


Figura 4.21: Capacidade de previsão dos modelos - representação dos valores previstos pelo modelo parcimonioso (M07, círculos azuis) em função da resposta y_{12} ; a linha oblíqua representa a bissetriz dos quadrantes ímpares.

Um vez mais as observações aqui encontradas podem ser explicadas em termos de sobreparametrização do modelo - o excesso inicial de parâmetros conduzem a más estimativas; à medida que se vai removendo parâmetros do modelo este começa a melhorar o seu desempenho (menores desvios e melhores correlações) até que se encontra um modelo exacto (apenas com erro de truncatura numérica) - contudo este modelo ainda contém um número excessivo de parâmetros e não deve ser por isso considerado. Continuando a retirar parâmetros atinge-se uma situação em que o modelo possui maior número de graus de liberdade e ainda assim excelente desempenho; prosseguindo a operação de simplificação o modelo entra em ruptura com os valores experimentais e começa de novo a dar desempenhos deficientes.

Na tabela (4.21) apresentam-se as estimativas paramétricas obtidas no caso do modelo parcimonioso (M07).

Tabela 4.21: Estimativas paramétricas obtidas com o modelo parcimonioso (M07).

Variável	β	$\sigma(\beta)$	%RSE	TV	α
x55	-0.95	0.16	16.4	6.10	0.000
x68	0.51	0.09	17.3	5.78	0.000
x76	2.21	0.20	9.3	10.80	0.000
x84	1.05	0.08	7.7	12.96	0.000
x86	2.42	0.28	11.7	8.58	0.000
x87	-1.11	0.17	15.8	6.34	0.000
x93	0.31	0.09	30.9	3.24	0.006

β - valor do parâmetro estimado, equação (2.49), pág 48; $\sigma(\beta)$ - erro padrão do parâmetro, equação (2.55), pág.49 ; %RSE - erro padrão relativo, equação (2.49), pág 48 ; TV e α - valor de teste e respectivo valor de prova referente ao teste de significado estatístico do parâmetro considerando a hipótese nula $H_o : |\beta| = 0$.

Da tabela verifica-se que todos os parâmetros encontrados têm significado estatístico (assumindo $H_o : |\beta| = 0$).

Na tabela (4.22) encontra-se uma compilação dos resultados obtidos por análise de correlação, modelação implícita (PLS) e modelação explícita (MOLS).

Tabela 4.22: Compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (w) e modelação explícita (β).

Variável	x55	x56	x57	x58	x59	x60	x68
ρ							
w	0.19	-0.25	-0.19	-0.19	-0.19	0.26	0.17
β	-0.95						0.51
Variável	x75	x76	x79	x80	x82	x84	x86
ρ						0.62	
w	0.25	0.23	0.26	0.21	-0.27	0.36	0.25
β		2.21				1.05	2.42
Variável	x87	x93	x101	x102	x103	x107	x108
ρ							
w	0.18	-0.20	0.25	-0.23	-0.23	-0.26	-0.15
β	-1.11	0.31					

ρ - estimativas obtidas por análise de correlação; w - contribuição evidenciada por PLS; β - efeito obtido por modelação explícita através do modelo parcimonioso (M07).

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso corresponde à situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação identificou apenas 1 dos 7 preditores mais influentes - x84 surge em ambos os casos com contribuição positiva. Já a modelação implícita (PLS) aponta para 21 contribuições relevantes das quais 14 sugeridas pelo primeiro factor latente ($FL1$) - sobre esta componente latente estão representados x55, x68, x76, x84 e x93 embora as contribuições x55 e x93 venham previstas em sentido oposto. Outras duas das variáveis indicadas como relevantes pelo modelo parcimonioso (x86 e x87) surgem como relevantes ao nível do segundo factor latente ($FL2$), vindo a x87 em sentido oposto - estas verificações são algo difíceis de entender mas, pelo menos o PLS tem a virtude de conseguir processar directamente sistemas mal condicionados e ainda assim evidenciar as variáveis mais relevantes.

Deste estudo resulta que a resposta y_{12} depende de x_{55} , x_{68} , x_{76} , x_{84} , x_{86} , x_{87} e x_{93} - corresponde ao modelo definido como parcimonioso (M07) e em que todos os parâmetros possuem significado estatístico ($H_0 : |\beta_i| = 0$ com $\nu = 13$).

Da tabela (4.22) verifica-se que as variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada por y_{12} são x_{68} [0.51], x_{76} [2.21], x_{84} [1.05], x_{86} [2.42] e x_{93} [0.31] - a sua influência crescente agrava a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio.

Em contrapartida, as variáveis x_{55} [-0.95] e x_{87} [-1.11] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada coma a resposta y_{12} indicam que o seu aumento têm como impacto a redução do seu valor.

4.2.5 Resposta y_9

No caso da resposta y_9 , matriz dos predictores contém apenas 77 objectos ($N = 77$) representados sobre 16 dimensões ($M = 16$, variáveis X_g , x_{41} , x_{42} , x_{43} , x_{44} , x_{45} , x_{46} , x_{47} , x_{48} , x_{49} , x_{50} , x_{51} , x_{52} , x_{53} , x_{54} , X_h) apresenta as dimensões $X_{(77 \times 16)}$.

4.2.5.1 Análise de correlação

Com $N = 77$, impondo o critério $\alpha = 0.03$, ver secção (3.3) na página 57, verificou-se quais os valores limites de correlação para $|\rho| = 0$ e $|\rho| = 1$, obteve-se os valores 0.247 e 0.877 indicando que valores absolutos de correlação inferiores a 0.247 são nulos enquanto que valores superiores a 0.877 serão considerados unitários.

Estudando a correlação interna ao sub-espço predictor verificou-se a igualdade ($\rho = 1$) entre os predictores $X_h = X_g$.

Impondo o valor limite de $\rho \geq 0.247$ no estudo da interdependência entre o sub-espço predictor e a resposta y_9 , detectaram-se fortes interdependências com quatro predictores: X_g [0.456], X_h [0.461], x_{43} [-0.383] e x_{47} [-0.263]. Na figura (4.22) encontra-se representada a dependência de y_9 com estes predictores identificados como relevantes.

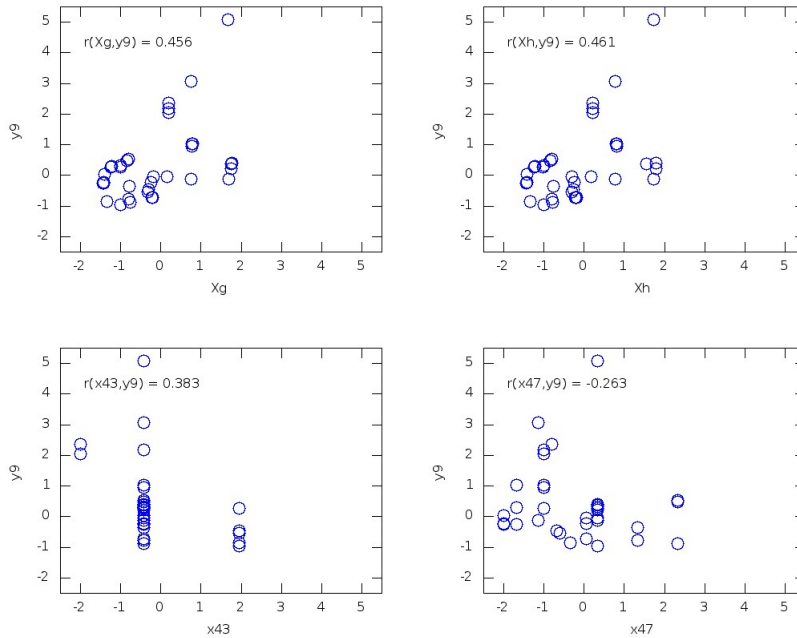


Figura 4.22: Representação das interdependências de $y9$ com Xg [0.456], Xh [0.461], $x43$ [-0.383] e $x47$ [-0.263].

Da figura anterior verifica-se que as interdependências detectadas apenas como sendo não nulas apresentam uma variação mútua com $y9$ de algum modo perceptível, em termos visuais.

4.2.5.2 Abordagem PLS

A matriz das variáveis independentes ($X_{(77 \times 16)}$) foi fornecida à modelação implícita com PLS como sub-espço predictor para descrever a resposta $y9$.

Na tabela (4.23) encontram-se as estimativas obtidas da análise PLS em relação à resposta $y9$.

Table 4.23: Resultados obtidos na abordagem da resposta y9 através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum(\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum(\%J_y)$	β_i	η_i
0	1216.00	100	—	—	76.00	100	—	—	—	—
1	1069.92	88.0	12.0	12.0	39.99	52.6	47.4	47.4	6.001	3.94
2	734.84	60.4	27.6	39.6	37.77	49.7	2.9	50.3	1.489	0.11
3	554.09	45.6	14.9	54.4	35.48	46.7	3.0	53.3	1.516	0.20
4	342.50	28.2	17.4	71.8	34.40	45.3	1.4	54.7	1.038	0.08
5	227.64	18.7	9.4	81.3	33.36	43.9	1.4	56.1	1.020	0.14
6	140.89	11.6	7.1	88.4	32.86	43.2	0.7	56.8	0.709	0.09
7	113.52	9.3	2.3	90.7	31.95	42.0	1.2	58.0	0.952	0.53
8	90.84	7.5	1.9	92.5	30.70	40.4	1.6	59.6	1.117	0.88
9	54.57	4.5	3.0	95.5	30.17	39.7	0.7	60.3	0.729	0.23
10	39.26	3.2	1.3	96.8	29.12	38.3	1.4	61.7	1.023	1.09

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos predictores e das respostas); $\%R_x$ e $\%R_y$ - informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justificar a resposta em y ; $\sum(\%U_x)$ e $\sum(\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y / \%U_x$).

Da tabela anterior verifica-se que não se consegue obter uma boa descrição de y9 - apenas se consegue descrever cerca de 61.7% da informação nela contida.

1. o primeiro factor latente ($FL1$) utiliza cerca de 12.0% da informação contida no sub-espaco dos predictores para descrever cerca de 47,4% da informação contida nesta resposta, o que corresponde a uma eficiência de $\eta_1 = 3.9$, muito superior a 1;
2. os restantes factores latentes têm uma eficiência baixa (inferior à unidade) e por isso não serão considerados para efeitos de descrição da resposta;
3. após a remoção do efeito do primeiro factor latente permanece cerca de 52.6% de informação residual da resposta por ser modelada evidenciando que, neste caso, a resposta não consegue ser bem descrita com base neste conjunto de predictores.

Na figura 4.23 encontram-se representados os scree plots referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

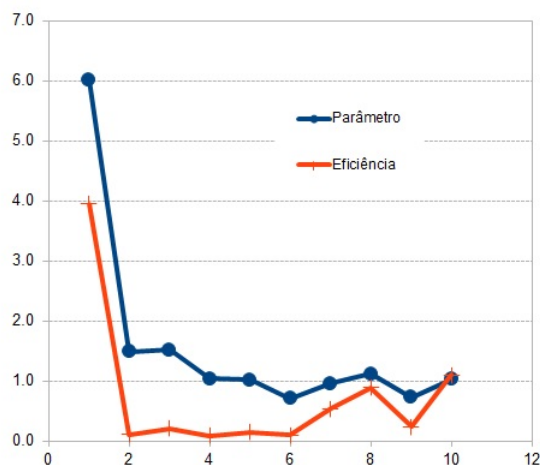


Figura 4.23: Scree plot dos parâmetros estimados pelo modelo e a eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Da figura anterior constata-se facilmente que as variáveis representadas (valores dos parâmetros do modelo e eficiência do factor latente) são concordantes e peremptórias em mostrar que apenas o primeiro factor latente, $FL1$, é que é relevante.

Importa agora avaliar a informação contida em $FL1$ em termos de representatividade dos predictores sobre a resposta.

Na tabela (4.24) apenas se apresentam os pesos mais significativos sobre a resposta $y9$, atendendo à condição expressa pela equação (3.2) na página 58.

Tabela 4.24: Avaliação do impacto dos predictores mais relevantes sobre $y9$ através da análise das respectivas contribuições.

Predictor	FL1
Xg	-0.501
x43	0.421
x47	0.289
x48	0.264
x52	-0.263
Xh	-0.506

FL1 - primeiro factor latente

Atendendo a que apenas existe neste caso um único factor latente relevante ($FL1$) e que ainda assim este é um mau descritor da resposta $y9$, da tabela (4.24) estamos a prever a dependência de $y9$ com 6 predictores, designadamente: Xg, x43, x47, x48, x52 e Xh.

Estas dependências serão avaliadas com base no ajuste explícito com polinómios de grau unitário (MOLS).

4.2.5.3 Modelação explícita

Devido à elevada interdependência entre $Xh = Xg$, a variável Xh foi removida deste estudo tendo a matriz de predictores passado a ser $X_{(77 \times 15)}$.

Os resultados encontram-se sistematizados na tabela (4.25).

Tabela 4.25: Convergência iterativa do modelo polinomial obtido por MOLS; inicia-se o estudo com todas as variáveis (número de parâmetros $p = 15$) e vai-se rejeitando sucessivamente a variável que apresenta parâmetro com menor significado estatístico (que possui maior erro relativo). Os valores 1 e 0 associados às variáveis em estudo indicam a presença ou ausência do respectivo parâmetro no modelo polinomial.

Iteração	1	8	9	10	11	12	13	14	15
Xg	1	1	1	1	1	1	1	1	1
x41	1	1	1	1	1	1	0	0	0
x42	1	0	0	0	0	0	0	0	0
x43	1	0	0	0	0	0	0	0	0
x44	1	0	0	0	0	0	0	0	0
x45	1	0	0	0	0	0	0	0	0
x46	1	0	0	0	0	0	0	0	0
x47	1	0	0	0	0	0	0	0	0
x48	1	1	1	1	1	1	1	1	0
x49	1	1	1	1	0	0	0	0	0
x50	1	1	0	0	0	0	0	0	0
x51	1	0	0	0	0	0	0	0	0
x52	1	1	1	0	0	0	0	0	0
x53	1	1	1	1	1	0	0	0	0
x54	1	1	1	1	1	1	1	0	0
Nm.	M15	M08	M07	M06	M05	M04	M03	M02	M01
p	15	8	7	6	5	4	3	2	1
SS	26.76	29.89	31.49	33.23	33.94	39.03	43.55	54.10	60.21
ν	62	69	70	71	72	73	74	75	76
σ_{fit}^2	0.432	0.433	0.450	0.468	0.471	0.535	0.588	0.721	0.792
ρ	0.805	0.779	0.765	0.750	0.744	0.697	0.653	0.537	0.456
$\rho = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\rho = 1$	0.012	0.005	0.003	0.002	0.002	0.001	0.000	0.000	0.000

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47), ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Da tabela anterior verifica-se:

- o primeiro modelo (M15) que é o mais completo ($p=15$), é aquele que apresenta melhor desempenho quer na qualidade do ajuste (σ_{fit}^2) quer na descrição da resposta (ρ) - a correlação múltipla atinge o valor de prova máximo (com $H_0 : |\rho| = 1$);
- em termos de desempenho dos modelos (σ_{fit}^2 e ρ) estes estimadores começam bem e depois vão piorando progressivamente até que terminam assumindo valores inadequados;

3. o modelo parcimonioso (M05 com $p = 5$) para y_9 contém os preditores $X_g, x_{41}, x_{48}, x_{53}$ e x_{54} ; a correlação múltipla encontrada ($\rho = 0.744$) não é infelizmente satisfatória já que o valor de prova para $\rho = 1$ é de apenas $\alpha = 0.002$.

Neste caso em concreto, o sub-espaço predictor utilizado para descrever a resposta y_9 é deficitário, razão pela qual o modelo mais completo não consegue fornecer um bom ajuste e a correlação deficiente que apenas é identificada, em termos estatísticos, com não nula. Assim, os fenómenos de sobreparametrização não são patentes e os parâmetros de desempenho do modelo vão-se deteriorando com a remoção sucessiva de parâmetros.

Na figura (4.24) a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação (ρ) em função do número da iteração do processo de pesquisa do melhor modelo - o traço vertical assinala a iteração em que foi atingido o modelo parcimonioso (M05). Já a figura (4.25) apresenta a qualidade do ajuste conseguida com o modelo parcimonioso (M05).

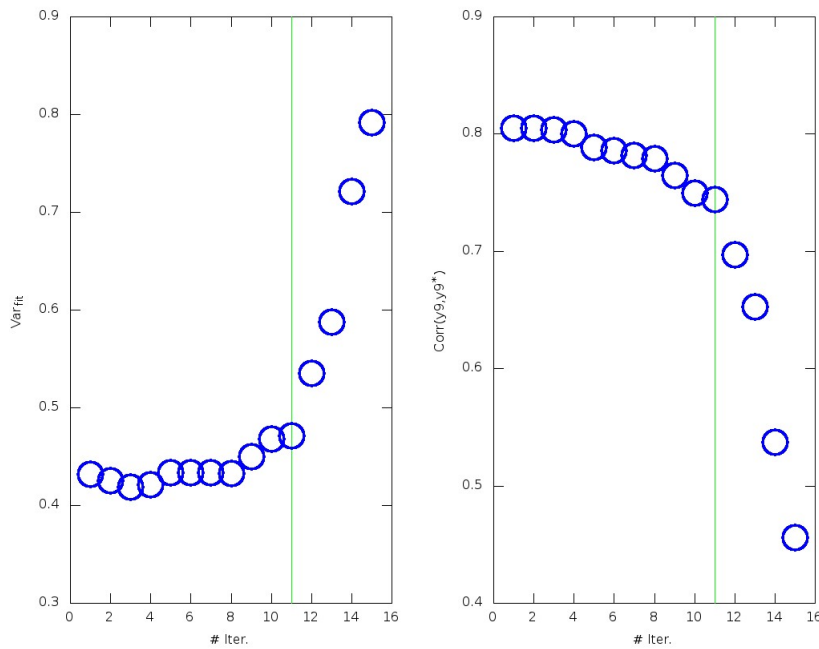


Figura 4.24: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa do melhor modelo. O traço vertical assinala o modelo parcimonioso.

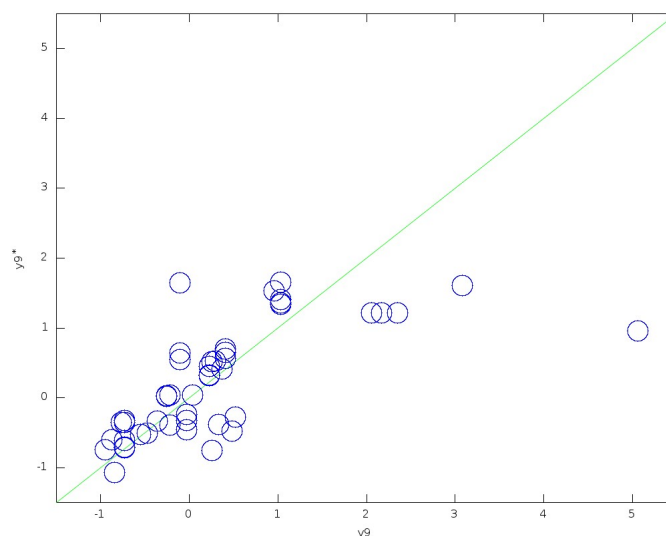


Figura 4.25: Capacidade de previsão do modelo - representação dos valores previstos (y_9^*) pelo modelo parcimonioso (M05) em função da resposta y_9 ; a linha oblíqua representa a bissetriz dos quadrantes ímpares.

Da figura 4.24 verifica-se a variância do ajuste decresce com a remoção de parâmetros. Continuando a remover parâmetros o erro do ajuste vai aumentando e a correlação assume uma queda progressiva. Com $p = 5$ atinge-se o melhor compromisso estatístico entre bom desempenho do modelo e maximização do número de graus de liberdade (modelo parcimonioso) e, daí para a frente, a situação da qualidade do ajuste deteriora-se rapidamente.

Na tabela (4.26) apresentam-se as estimativas paramétricas obtidas no caso do modelo parcimonioso (M05).

Tabela 4.26: Estimativas paramétricas obtidas com o modelo parcimonioso (M05).

Variável	β	$\sigma(\beta)$	%RSE	TV	α
Xg	0.82	0.10	12.0	8.31	0.000
x41	0.40	0.11	26.3	3.80	0.000
x48	-0.38	0.09	22.4	4.46	0.000
x53	-0.36	0.11	30.5	3.28	0.001
x54	0.48	0.11	22.3	4.48	0.000

β - valor do parâmetro estimado, equação (2.49), pág 48; $\sigma(\beta)$ - erro padrão do parâmetro, equação (2.55), pág.49 ; %RSE - erro padrão relativo, equação (2.49), pág 48 ; TV e α - valor de teste e respectivo valor de prova referente ao teste de significado estatístico do parâmetro considerando a hipótese nula $H_0 : |\beta| = 0$.

Da tabela verifica-se que todos os parâmetros encontrados têm significado estatístico (assumindo $H_0 : |\beta| = 0$).

Na tabela (4.27) encontra-se uma compilação dos resultados obtidos por análise de correlação, modelação implícita (PLS) e modelação explícita (MOLS).

Tabela 4.27: Compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (w) e modelação explícita (β).

Variável	Xg	x41	x43	x47	x48	x52	x53	x54	Xh
ρ	0.46		-0.38	-0.26					0.46
w	-0.50		0.42	0.29	0.26	-0.26			-0.51
β	0.82	0.40			-0.38		-0.36	0.48	0.82*

ρ - estimativas obtidas por análise de correlação; w - contribuição evidenciada por PLS; β - efeito obtido por modelação explícita através do modelo parcimonioso (M05).

* A contribuição de Xh foi acrescida a MOLS devido à sua forte interdependência com Xg.

Assumindo que a modelação explícita (MOLS), através do modelo parcimonioso, representa a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação directa com y9 identificou apenas 2 dos 6 predictores mais influentes - as variáveis Xg e Xh são ambas detectadas com o mesmo tipo de efeito positivo (crescente). Já a modelação implícita (PLS) aponta para 3 das 6 contribuições mais relevantes, sugeridas pelo primeiro factor latente (*FL1*) - continua a identificar Xg e Xh às quais adiciona a contribuição de x48, todas a serem contabilizadas em sentido contrário.

Estes resultados mostram ainda que, embora a descrição da resposta y9 seja deficiente, o algoritmo de modelação explícita conseguiu encontrar a melhor solução usando toda a informação compatível presente no subespaço predictor.

Deste estudo resulta que a resposta y9 depende de Xg=Xh, x41, x48, x53 e x54 - corresponde ao modelo definido como parcimonioso (M05) e em que todos os parâmetros possuem significado estatístico ($H_0 : |\beta_i| = 0$ com $\nu = 72$).

Devido à interdependência estreita entre Xg-Xh, a variável de estudo (Xh) será agora incluída nas conclusões.

Da tabela (4.27) verifica-se que as variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y9 são Xg e Xh* [0.82], x41 [0.40] e x54 [0.48] - influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio.

Em contrapartida, as variáveis x48 [-0.38] e x53 [-0.36] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y9 indicam que o seu aumento têm como impacto a redução do seu valor.

4.2.6 Resposta y4

No caso da resposta y4 a matriz dos predictores contém apenas 26 objectos ($N = 26$) representados sobre 26 dimensões ($M = 26$, variáveis Xa, x1, x2, x3, x5, x8, x9, x10, x12, x14, x15, x16, x17, x18, x19, x20, x21, Xb, Xc, Xd, Xe, x34, x37, x39, Xf e x40) apresentando-se sob a forma de $X_{(26 \times 26)}$.

4.2.6.1 Análise de correlação

Com $N = 26$, a correlação será considerada nula ($H_0 : |\rho| = 0$) se o seu módulo for inferior a 0.425 (indicando que não existe correlação significativa) enquanto que valores superiores a 0.635 indicam uma elevada correlação (valor considerado unitário).

Estudando a correlação interna ao sub-espço predictor verificou-se a igualdade ($\rho = 1$) entre os predictores $x_{15} = x_8$, $Xb = Xa$, $Xc = Xa$, $Xd = Xa$, $Xe = Xa$ e $Xf = Xa$.

Impondo o valor limite de $|\rho| \geq 0.425$ no estudo da interdependência entre o sub-espço predictor e a resposta y_4 , detectou-se apenas uma interdependência pouco significativa entre x_{39} - y_4 [0.524], representada na figura (4.26).

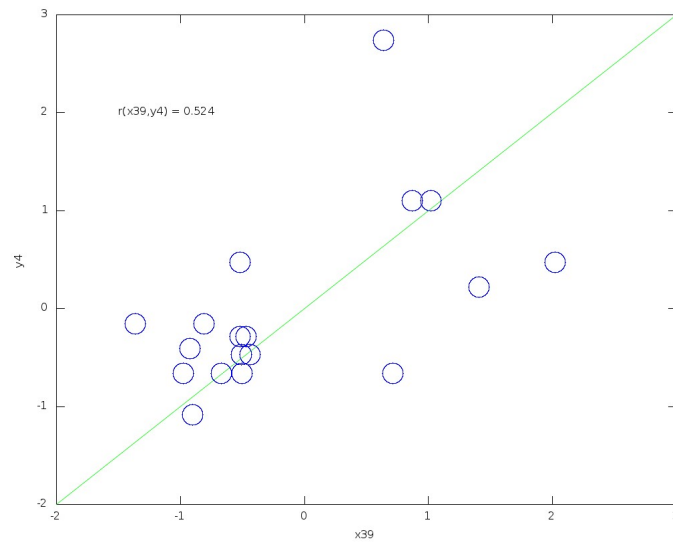


Figura 4.26: Representação das interdependências de y_4 com x_{39} [0.524].

Da figura anterior fica-se com a percepção visual de que a interação x_{39} - y_4 , identificada como não nula, parece fazer sentido especialmente se se prever que possam existir algumas situações anómalas (eventuais outliers).

4.2.6.2 Abordagem PLS

De modo similar ao que foi anteriormente apresentado, a matriz de variáveis independentes ($X_{(26 \times 26)}$) foi integralmente fornecida como sub-espço predictor para procurar descrever a resposta y_4 através da modelação implícita com PLS.

Na tabela (4.28) encontram-se as estimativas obtidas da análise PLS em relação à resposta y_4 ,

Tabela 4.28: Resultados obtidos na abordagem da resposta y4 através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum(\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum(\%J_y)$	β_i	η_i
0	650.00	100	—	—	25.00	100	—	—	—	—
1	483.90	74.4	25.6	25.6	17.09	68.4	31.6	31.6	2.812	1.24
2	323.68	49.8	24.6	50.2	11.47	45.9	22.5	54.1	2.371	0.91
3	275.15	42.3	7.5	57.7	6.12	24.5	21.4	75.5	2.312	2.86
4	221.82	34.1	8.2	65.9	4.26	17.0	7.5	83.0	1.365	0.91
5	164.18	25.3	8.9	74.7	3.66	14.6	2.4	85.4	0.777	0.27
6	136.28	21.0	4.3	79.0	2.56	10.2	4.4	89.8	1.047	1.02
7	114.31	17.6	3.4	82.4	1.58	6.3	3.9	93.7	0.990	1.16
8	88.76	13.7	3.9	86.3	1.31	5.2	1.1	94.8	0.524	0.28
9	64.01	9.8	3.8	90.2	1.15	4.6	0.6	95.4	0.392	0.16
10	45.07	6.9	2.9	93.1	0.92	3.7	0.9	96.3	0.483	0.32

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos predictores e das respostas); $\%R_x$ e $\%R_y$ informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justifica a resposta em y ; $\sum(\%U_x)$ e $\sum(\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y / \%U_x$).

Da tabela anterior verifica-se que com apenas dois factores latentes ($f = 2$) consegue-se descrever apenas 54.1% da resposta em causa:

1. o primeiro factor latente ($FL1$) utiliza cerca de 25.6% da informação contida no sub-espaco dos predictores para descrever cerca de 31.6% da informação contida nesta resposta, o que corresponde a uma eficiência de $\eta_1 = 1.24$, pouco superior a 1;
2. o segundo factor latente ($FL2$) requer 24.6% da informação contida nos predictores para descrever cerca de 22.5% da resposta apresentando uma baixa eficiência ($\eta_2 = 0.91$), embora próxima de 1;
3. após o efeito destes dois factores latentes terem sido removidos ficam ainda por descrever 45.9% da informação contida na resposta em causa.

Uma vez mais estamos perante uma situação de má descrição da resposta. Assim sendo, não havendo predictores relevantes para justificação da resposta, fica-se com a noção de que os modelos encontrados vão necessitar de suporte num grande número de variáveis para tentar modelar essa resposta. A tabela (4.28) prenúncio deste facto quando sugere a necessidade de recorrer a 4 factores latentes para conseguir descrever cerca de 80% da informação contida na resposta.

Na figura (4.27) encontram-se representados os scree plots referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

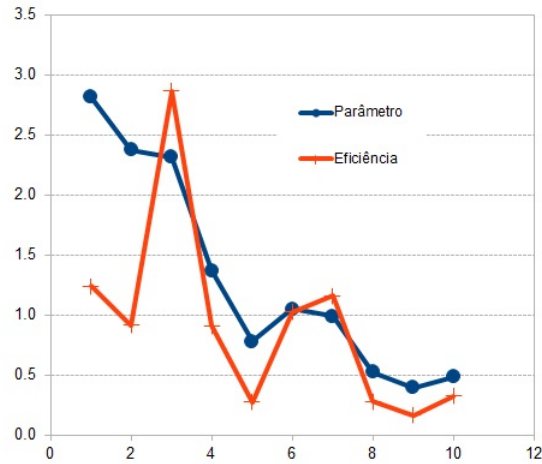


Figura 4.27: Scree plot dos parâmetros estimados pelo modelo e da eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Da figura anterior constata-se que há dificuldade em estabelecer o número de factores latentes necessários à modelação desta resposta.

Se nos cingirmos a apenas os dois primeiros factores latentes, $FL1$ e $FL2$ temos a limitação de apenas conter cerca de 50% da informação da resposta mas ainda assim teremos aqui representadas as contribuições mais relevantes sobre estes factores latentes.

Importa agora avaliar a informação contida em $FL1$ e $FL2$ em termos de representatividade dos predictores sobre a resposta.

Na tabela (4.29) apenas se apresentam os pesos mais significativos sobre a resposta y_4 .

Tabela 4.29: Avaliação do impacto dos preditores mais relevantes sobre y_4 através da análise das respectivas contribuições.

Predictor	FL1	FL2
Xa		-0.303
x1		0.283
x2	0.204	
x3	0.283	
x5	0.215	
x9		0.209
x12		-0.287
x14	-0.264	
x16		-0.208
x17	-0.273	
x19	0.197	-0.318
x20	0.203	
x21	-0.366	
Xb		-0.304
Xc		-0.304
Xd		-0.306
Xe		-0.306
x39	-0.538	
Xf		-0.302

FL1 - primeiro factor latente; FL2 - segundo factor latente.

Como tenha sido previsto anteriormente, a resposta y_4 também não consegue ser bem descrita através do conjunto de preditores utilizados. Não havendo a detecção de variáveis relevantes (factores evidentes), o modelo procura suportar-se num número elevado de variáveis para tentar descrever, ainda que deficientemente, a resposta - do conjunto original de 26 preditores a modelação implícita procura descrever a resposta y_4 com base em 19 desses preditores, designadamente com: Xa, x1, x2, x3, x5, x9, x12, x14, x16, x17, x19, x20, x21, Xb, Xc, Xd, Xe, x39, Xf.

4.2.6.3 Modelação explícita

Na matriz original das variáveis independentes ($X_{(26 \times 26)}$) foram identificadas elevadas interdependências entre $X_a=X_b=X_c=X_d=X_e=X_f$ e $x_8=x_{15}$ forçando a remover previamente as variáveis (Xb, Xc, Xd, Xe, Xf e x15) em detrimento de Xa e x8 que forma retidas. Assim sendo o sub espaço predictor ficou cingido às dimensões $X_{(26 \times 20)}$. Esta matriz mostrou ser invertível e, deste modo foi possível prosseguir com a abordagem de ajuste explícito multivariado.

Os resultados encontram-se sistematizados na tabela (4.30).

Tabela 4.30: Convergência iterativa do modelo polinomial obtido por MOLS; inicia-se o estudo com todas as variáveis (número de parâmetros $p = 20$) e vai-se rejeitando sucessivamente a variável que apresenta parâmetro com menor significado estatístico (que possui maior erro relativo). Os valores 1 e 0 associados às variáveis em estudo indicam a presença ou ausência do respectivo parâmetro no modelo polinomial.

Iteração	1	7	8	9	10	11	12	13	19	20
Xa	1	1	1	1	1	0	0	0	0	0
x1	1	1	1	1	1	1	1	1	1	0
x2	1	0	0	0	0	0	0	0	0	0
x3	1	1	1	1	1	1	1	0	0	0
x5	1	1	1	1	1	1	1	1	1	1
x8	1	1	1	1	1	1	1	1	0	0
x9	1	1	0	0	0	0	0	0	0	0
x10	1	1	1	1	1	1	1	1	0	0
x12	1	1	1	1	1	1	1	1	0	0
x14	1	1	1	1	0	0	0	0	0	0
x16	1	0	0	0	0	0	0	0	0	0
x17	1	0	0	0	0	0	0	0	0	0
x18	1	1	1	1	1	1	1	1	0	0
x19	1	1	1	1	1	1	1	1	0	0
x20	1	0	0	0	0	0	0	0	0	0
x21	1	1	1	0	0	0	0	0	0	0
x34	1	1	1	1	1	1	0	0	0	0
x37	1	0	0	0	0	0	0	0	0	0
x39	1	1	1	1	1	1	1	1	0	0
x40	1	0	0	0	0	0	0	0	0	0
Nm.	M20	M14	M13	M12	M11	M10	M09	M08	M02	M01
p	20	14	13	12	11	10	9	8	2	1
SS	0.08	0.24	0.31	0.32	0.38	0.57	0.70	1.07	16.66	23.90
ν	6	12	13	14	15	16	17	18	24	25
σ_{fit}^2	0.014	0.020	0.024	0.023	0.025	0.035	0.041	0.059	0.694	0.956
ρ	0.998	0.995	0.994	0.994	0.992	0.989	0.986	0.978	0.578	0.209
$\rho = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.294
$\rho = 1$	0.946	0.867	0.843	0.835	0.815	0.766	0.732	0.663	0.018	0.000

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47), ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Da tabela anterior verifica-se:

1. o primeiro modelo (M20) que é o mais completo ($p=20$), é aquele que apresenta melhor desempenho quer na qualidade do ajuste (σ_{fit}^2) quer na descrição da resposta (ρ) - a correlação múltipla atinge o valor de prova máximo (com $H_0 : |\rho| = 1$);;
2. em termos de desempenho dos modelos (σ_{fit}^2 e ρ), estes estimadores começam bem e depois vão piorando progressivamente até que terminam assumindo valores inadequados;

3. o modelo parcimonioso (M11 com $p = 11$) para y_4 contém os preditores X_a , x_1 , x_3 , x_5 , x_8 , x_{10} , x_{12} , x_{18} , x_{19} , x_{34} e x_{39} ; a correlação múltipla encontrada ($\rho = 0.992$) é excelente já que o valor de prova para $\rho = 1$ é de $\alpha = 0.815$, enquanto que para $\rho = 0$ é de $\alpha = 0.000$.

Na figura (4.28) encontra-se representada a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação (ρ) em função do número de iterações realizadas no processo de tentativa de identificação do melhor modelo - nas figuras encontra-se representada uma linha vertical que assinala a iteração em que foi obtido o modelo parcimonioso (M11). Já a figura (4.29) apresenta a qualidade do ajuste conseguida com o modelo parcimonioso (M11).

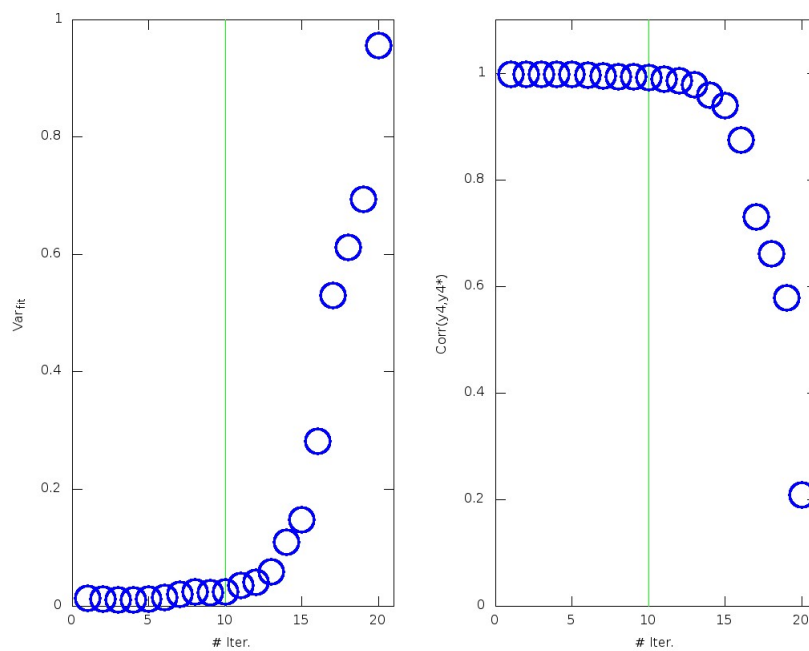


Figura 4.28: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa do melhor modelo. O traço vertical assinala o modelo parcimonioso.

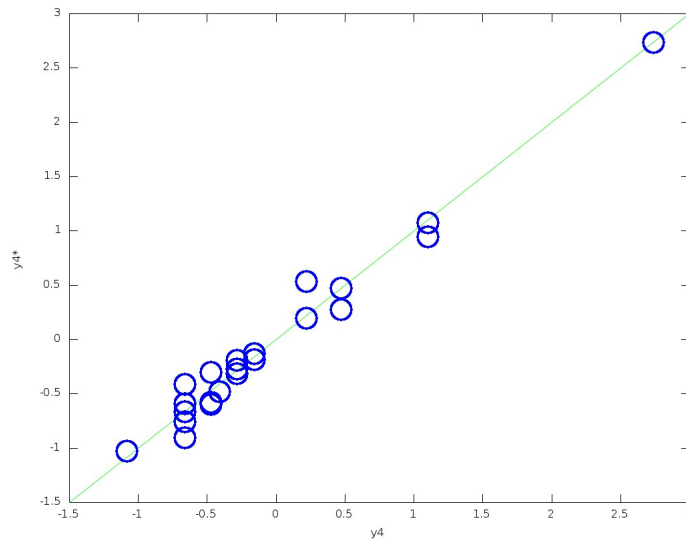


Figura 4.29: Capacidade de previsão do modelo - representação dos valores previstos (y_4^*) pelo modelo parcimonioso (M11) em função da resposta y_4 ; a linha oblíqua representa a bissetriz dos quadrantes ímpares.

Contrariamente ao que seria esperado, os resultados obtidos com MOLS revelam que com um número relativamente grande de parâmetros ($p = 11$) é possível descrever bastante bem a resposta y_4 - o coeficiente de correlação obtido apresenta um elevado valor de prova para ser considerado unitário. Este resultado é bastante explícito com a representação da figura (4.29).

Na tabela (4.31) apresentam-se as estimativas paramétricas obtidas no caso do modelo parcimonioso (M11).

Tabela 4.31: Estimativas paramétricas obtidas com o modelo parcimonioso (M11).

Variável	β	$\sigma(\beta)$	%RSE	TV	α
Xa	-0.21	0.08	36.4	2.74	0.015
x1	6.03	0.44	7.3	13.65	0.000
x3	-0.19	0.04	23.3	4.29	0.001
x5	-1.47	0.08	5.6	17.80	0.000
x8	-1.36	0.13	9.8	10.25	0.000
x10	-0.71	0.13	17.9	5.60	0.000
x12	4.87	0.45	9.2	10.92	0.000
x18	0.18	0.04	19.7	5.07	0.000
x19	-1.21	0.07	5.7	17.58	0.000
x34	-0.20	0.06	28.9	3.46	0.004
x39	0.34	0.05	16.0	6.26	0.000

β - valor do parâmetro estimado, equação (2.49), pág 48; $\sigma(\beta)$ - erro padrão do parâmetro, equação (2.55), pág.49 ; %RSE - erro padrão relativo, equação (2.49), pág 48 ; TV e α - valor de teste e respectivo valor de prova referente ao teste de significado estatístico do parâmetro considerando a hipótese nula $H_0 : |\beta| = 0$.

Da tabela verifica-se que todos os parâmetros encontrados têm significado estatístico (assumindo $H_0 : |\beta| = 0$).

Na tabela (4.32) encontra-se uma compilação dos resultados obtidos por análise de correlação, modelação implícita (PLS) e modelação explícita (MOLS).

Tabela 4.32: Compilação dos resultados obtidos por análise de correlação (ρ), modelação implícita (w) e modelação explícita (β).

Variável	Xa	x1	x2	x3	x5	x8	x9	x10
ρ								
w	-0.30	0.28	0.20	0.28	0.21		0.21	
β	-0.21	6.03		-0.19	-1.47	-1.36		-0.71
Variável	x12	x14	x15	x16	x17	x18	x19	x20
ρ								
w	-0.29	-0.26		-0.21	-0.27		0.20	0.20
β	4.87		-1.36*			0.18	-1.21	
Variável	x21	Xb	Xc	Xd	Xe	x34	x39	Xf
ρ							0.52	
w	-0.37	-0.30	-0.30	-0.31	-0.31		-0.54	-0.30
β		-0.21*	-0.21*	-0.21*	-0.21*	-0.20	0.34	-0.21*

ρ - estimativas obtidas por análise de correlação; w - contribuição evidenciada por PLS; β - efeito obtido por modelação explícita através do modelo parcimonioso (M11).

* As contribuições de Xb, Xc, Xd, Xe e Xf foram acrescida a MOLS devido à sua forte interdependência com Xa e x15 devido à forte interdependência com x8.

Deste estudo fica-se com a noção de que o processo iterativo de modelação explícita para pesquisa do modelo parcimonioso é muito poderoso - quando a correlação directa e a modelação implícita por PLS previam um resultado desastroso, o método de modelação explícita encontrou com êxito um modelo muito satisfatório para descrever a resposta y4.

Deste estudo resulta que a resposta y4 depende de Xa (Xb, Xc, Xd, Xe, Xf), x1, x3, x5, x8, x10, x12, x15, x18, x19, x34 e x39 - corresponde ao modelo definido como parcimonioso (M11) e em que todos os parâmetros possuem significado estatístico ($H_0 : |\beta_i| = 0$ com $\nu = 15$).

Da tabela (4.32) verifica-se que as variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada por y4 são x1 [6.03], x12 [4.87], x18 [0.18] e x39 [0.34] - influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio.

Em contrapartida, as variáveis Xa (Xb, Xc, Xd, Xe, Xf) [-0.21], x3 [-0.19], x5 [-1.47], x8 (x15) [-1.36], x19 [-1.21] e x34 [-0.20] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y9 indicam que o seu aumento têm como impacto a redução do seu valor.

4.2.7 Resposta y_5

No caso da resposta y_5 a matriz dos predictores contém apenas 107 objectos ($N = 107$) representados sobre 6 dimensões ($M = 6$, variáveis X_c , x_{29} , x_{30} , x_{31} , x_{32} e X_d) apresentando-se sob a forma de $X_{(107 \times 6)}$.

4.2.7.1 Análise de correlação

Com $N = 107$, a correlação será considerada nula ($H_0 : |\rho| = 0$) se o seu módulo for inferior a 0.210 (indicando que não existe correlação significativa) enquanto que valores superiores a 0.912 indicam uma elevada correlação (valor considerado unitário).

Estudando a correlação interna ao sub-espço predictor verificou-se a igualdade ($\rho = 1$) entre os predictores $X_d = X_c$.

Impondo o valor limite de $|\rho| \geq 0.210$ no estudo da interdependência entre o sub-espço predictor e a resposta y_5 não se detectou nenhuma interdependência não-nula. Neste estudo, maior valor de correlação directa conseguida entre estes predictores e a resposta y_5 foi obtida com x_{29} [0.121], e encontra-se representada na figura 4.30.

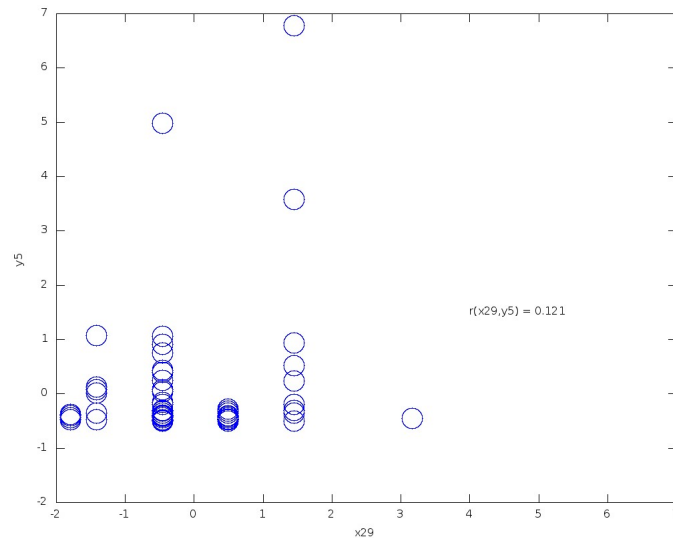


Figura 4.30: Representação das interdependências de y_4 com x_{29} [0.524].

Da figura anterior fica-se com a percepção visual de que há uma muito má interdependência entre y_5 e x_{29} , sendo esta quase que forçada por um conjunto muito restrito de valores.

4.2.7.2 Abordagem PLS

Um vez mais foi realizado o ajuste implícito, agora da resposta y_5 com o sub-espço predictor ($X_{(107 \times 6)}$).

Na tabela (4.33) encontram-se as estimativas obtidas da análise PLS em relação à resposta y_5 .

Tabela 4.33: Resultados obtidos na abordagem da resposta y_5 através da modelação implícita PLS.

#	SS_x	$\%R_x$	$\%U_x$	$\sum(\%U_x)$	SS_y	$\%R_y$	$\%J_y$	$\sum(\%J_y)$	β_i	η_i
0	636	100	—	—	106	100	—	—	—	—
1	429.75	67.6	32.4	32.4	103.76	97.89	2.11	2.11	1.496	0.065
2	312.89	49.2	18.4	50.8	101.98	96.21	1.68	3.79	1.335	0.092
3	276.18	43.4	5.8	56.6	99.48	93.85	2.36	6.15	1.581	0.409
4	192.79	30.3	13.1	69.7	99.30	93.68	0.16	6.32	0.418	0.013
5	0.17	0.0	30.3	100.0	99.23	93.61	0.07	6.39	0.267	0.002
6	0.00	0.0	0.0	100.0	29.40	27.74	65.88	72.26	8.356	2410.081

- iteração; SS_x e SS_y - somas de quadrados em x e em y (sub-espacos dos predictores e das respostas); $\%R_x$ e $\%R_y$ - informação residual relativa dos sub-espacos x e y (variabilidade não processada); $\%U_x$ e $\%J_y$ - quantidade relativa de informação utilizada em x para justificar a resposta em y ; $\sum(\%U_x)$ e $\sum(\%J_y)$ - cumulativos da informação utilizada em x e justificada em y ; β_i - parâmetro do modelo na iteração i ; η_i - eficiência relativa de cada factor latente ($\eta_i = \%J_y/\%U_x$).

Da tabela anterior verifica-se que a modelação implícita falha redondamente na descrição da resposta y_5 - o primeiro factor latente costuma ser aquele que melhor descreve a resposta conseguindo justificar pelo menos cerca de 30% dessa resposta. Neste caso em concreto verifica-se que apenas consegue descrever 2.1% da resposta em causa o que é um péssimo ajuste.

vejamos então a informação que nos dá esta tabela:

1. o primeiro factor latente ($FL1$) utiliza cerca de 32.4% da informação contida no sub-espaco dos predictores para descrever cerca de 2.11% da informação contida nesta resposta o que corresponde a uma eficiência de $\eta_1 = 0.065$, muito inferior a 1;
2. os restantes factores latentes, incluindo o primeiro factor latente, possuem uma eficiência muito inferior a 1 - neste caso a resposta não consegue ser bem descrita com base neste conjunto de predictores;
3. após o quarto factor latente o sub-espaco dos predictores ficou esvaziado ($\%R_x = 0.0$) e ainda só foi descrita cerca de 6.3% da resposta y_5 .

No nosso entender, estes resultados dificilmente poderiam ser piores, não se adivinhando um bom desfecho neste caso de descrição de factores.

Na figura (4.31) encontram-se representados os scree plots referentes às estimativas paramétricas e eficiências relativas de cada factor latente.

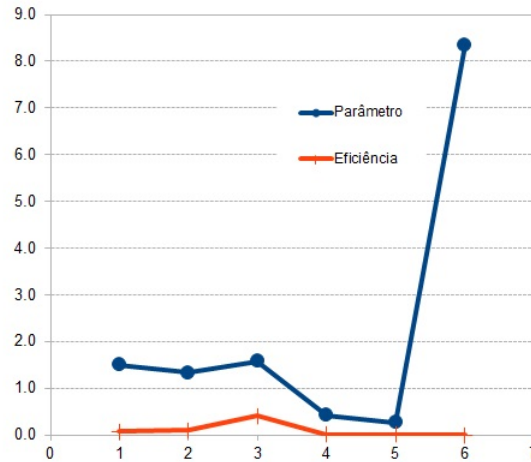


Figura 4.31: Scree plot dos parâmetros estimados pelo modelo e da eficiência dos factores latentes em função do número de iterações realizadas com o algoritmo PLS.

Se não se tiver em mente os resultados da tabela (4.33), a figura (4.31) pode induzir em erro - os três primeiros parâmetros superiores à unidade poderiam ser significado de algo relevante estar a ser evidenciado contudo, as eficiências nulas colmatam essa falha de interpretação - informam que são apenas resultado de uma anomalia numérica.

Assim sendo, neste caso em concreto, a modelação implícita através de PLS não permite avaliar a informação contida em y_5 ficando por isso esta abordagem terminada.

4.2.7.3 Modelação explícita

Apesar das evidências reveladas com a modelação implícita com PLS de que não é possível encontrar um modelo viável, com base nestes predictores, capaz de explicar, com significado estatístico, a resposta y_5 , foi removida temporariamente a variável X_d do sub-espço predictor para evitar singularidades numéricas e foi processado o método de pesquisa iterativa do melhor modelo por via explícita.

Os resultados encontram-se sistematizados na tabela (4.34).

Tabela 4.34: Convergência iterativa do modelo polinomial por MOLS, rejeitando sucessivamente a variável com menor significado estatístico.

Iteração	1	2	3	4	5
Xc	1	0	0	0	0
x29	1	1	1	0	0
x30	1	1	1	1	0
x31	1	1	1	1	1
x32	1	1	0	0	0
Nm.	M05	M04	M03	M02	M01
p	5	4	3	2	1
SS	99.60	99.71	100.03	102.82	105.27
ν	102	103	104	105	106
σ_{fit}^2	0.976	0.968	0.962	0.979	0.993
ρ	0.246	0.244	0.237	0.173	0.083
$\rho = 0$	0.012	0.012	0.014	0.075	0.392
$\rho = 1$	0.000	0.000	0.000	0.000	0.000

Nm. - Nome do modelo; p - número de parâmetros do modelo; SS - soma de quadrados residual do modelo (eq. 2.28, pág. 42); σ_{fit}^2 - variância do ajuste (eq.2.43, pág47), ρ - correlação múltipla do modelo (eq. 2.15, pág. 36); ν - graus de liberdade do ajuste ($N - p$); $\rho = 0$ - valor de prova para o teste de correlação nula assumindo $H_0 : |\rho| = 0$; $\rho = 1$ - valor de prova para o teste de correlação unitária assumindo $H_0 : |\rho| = 1$.

Da tabela anterior verifica-se que não foi possível encontrar um modelo parcimonioso - assim sendo o modelo mais completo teria de ser considerado como o melhor modelo (M05).

O melhor modelo não consegue sequer apresentar uma correlação significativa com os resultados da resposta - a correlação múltipla obtida ($\rho = 0.246$) apenas pode ser considerada como não-nula.

Na figura (4.32) representa-se a evolução da qualidade do ajuste (σ_{fit}^2) e da correlação (ρ) em função do número da iteração do processo de pesquisa do melhor modelo. Já a figura (4.33) apresenta a qualidade do ajuste conseguida à custa do modelo mais completo (M05).

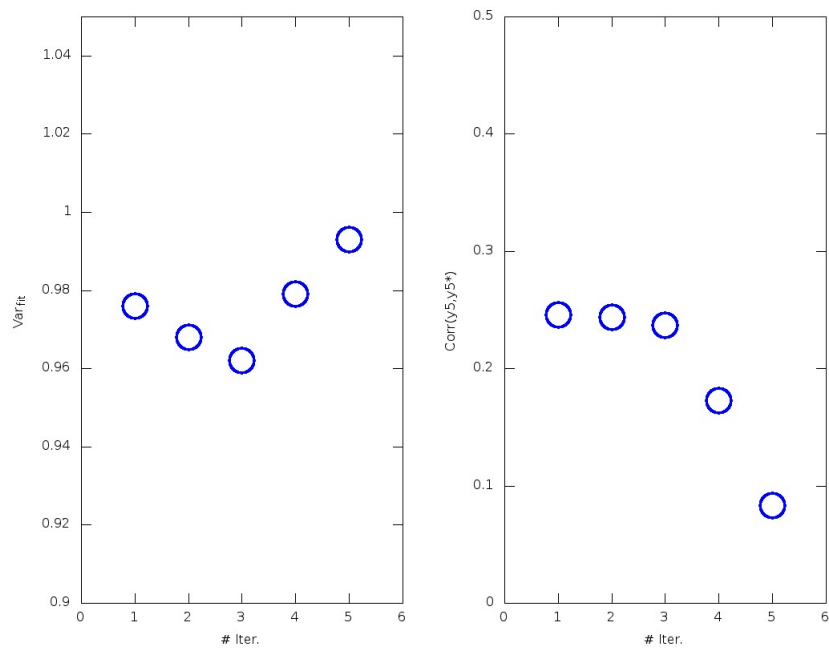


Figura 4.32: Evolução das estimativas referentes à qualidade do ajuste (σ_{fit}^2) e capacidade de descrever os valores experimentais (ρ) em função do número da iteração do processo de pesquisa do melhor modelo.

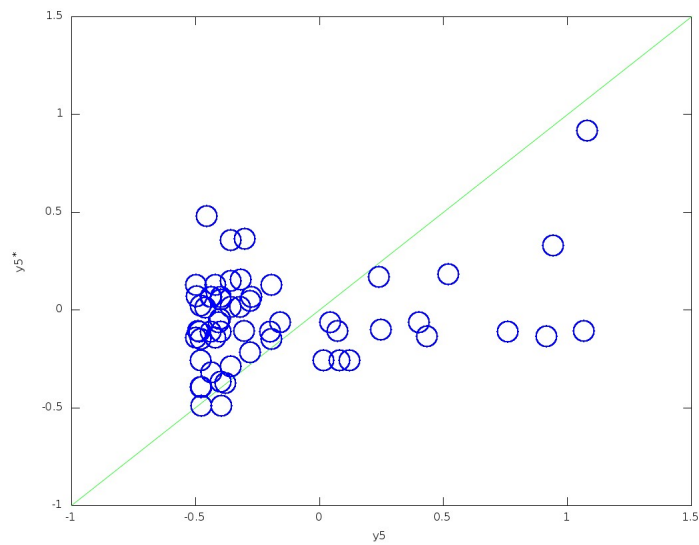


Figura 4.33: Capacidade de previsão dos modelos em função da resposta y_5 ; a linha oblíqua representa a bissetriz dos quadrantes ímpares.

De facto já se adivinhava este desfecho para a descrição da resposta y_5 - com este conjunto de descritores não faz sentido estar a tirar ilações sobre esta resposta razão pela qual este estudo se dá por terminado.

Capítulo 5

Conclusões

Dada a crescente importância da gestão de resíduos e atendendo à necessidade de melhoria contínua do sistema implementado na empresa, esta dissertação pretendeu auxiliar no desenvolvimento de uma gestão de resíduos interna mais eficiente.

Inicialmente começou por se definir a estratégia a adoptar e, para tal, foi essencial acompanhar os processos de fabrico e embalamento, assim como o circuito de cada produto, desde a pesagem das matérias-primas, passando pelas diferentes etapas do processo de fabrico até ao processo de embalamento.

Após analisar e perceber o modo de funcionamento do circuito de resíduos da empresa, direcionei a minha atenção para os detalhes dos respectivos produtos gerados. Deste modo, o meu trabalho incidiu na aplicação de ferramentas da qualidade e de métodos quimiométricos multivariados para o tratamento da informação relativa aos detalhes do processo de fabrico e embalamento de cada tipo de produto gerado na empresa. Esta metodologia foi também aplicada aos resíduos produzidos em cada etapa. No decorrer deste processo foi essencial uma análise crítica para detectar questões problemáticas, que no futuro poderão acarretar problemas associados à quantidade de resíduos produzidos e gastos acrescidos.

De um modo geral, as ferramentas da qualidade utilizadas constituíram uma forma eficaz de auxiliar o desenvolvimento da dissertação, uma vez que através destas foi possível concluir quais os resíduos com maior impacto, quer a nível de massa, quer a nível de custos.

Já a aplicação de métodos quimiométricos ao conjunto de dados a analisar, permitiu evidenciar qual ou quais os detalhes de fabrico mais relevantes para a empresa. Com esta avaliação foi possível identificar os parâmetros que geravam maior e menor quantidade de resíduos.

De seguida, apresenta-se uma descrição das várias conclusões retiradas de cada resposta de resíduos:

Através da análise de correlação, para o resíduo designado como y13, concluiu-se que x84 e x60 possuem uma correlação elevada com essa resposta enquanto que x101, x59 e x58 apresentam alguma interdependência, não nula. Já na análise de Mínimos Quadrados Parciais (PLS), verifica-se que com apenas dois factores latentes ($f=2$) consegue-se descrever 86.4% da resposta em causa. Neste método prevê-se a dependência de y13 com 17 predictores, designadamente: x55, x58, x59, x60, x62, x63, x68, x69, x78, x79, x80, x81, x82, x84, x86, x87 e x101.

Por fim, na modelação explícita, verifica-se que existe um modelo parcimonioso (M08) que contém os predictores x58, x60, x68, x69, x78, x80, x82, x84.

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso como sendo a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação identificou 6 dos 10 predictores mais influentes embora em dois desses casos o tipo de contribuição seja antagónica - ver casos x60, x80 e x101. Já a modelação implícita (PLS) aponta para 17 contribuições relevantes das quais 12 sugeridas pelo primeiro factor latente (*FL1*). As variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y13 são x78 [2.20], x84 [1.51], x68 [0.66], x69 [0.60] e x82 [0.42] - a dependência positiva indica que estas possuem uma influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio. Em contrapartida, as variáveis x80 [-2.48], x58-x59* [-1.78] e x60-x101* [-0.99] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y13 indica que o seu aumento tem como impacto a redução do seu valor.

Pela análise de correlação da resposta y7 conclui-se que x34 [-0.567] e x37 [0.444] possuem apenas correlações não nulas. Relativamente à análise de Mínimos Quadrados Parciais (PLS), apenas o primeiro factor latente veículo informação pertinente, capaz de descrever alguma interdependência da resposta com o sub-espaço dos predictores - prevê-se a dependência de y7 com 3 predictores, designadamente: x34, x37 e x40. No modelo de Mínimos Quadrados Multivariados (MOLS) verifica-se que existe um modelo parcimonioso (M05) que contém os predictores Xe, x34, x37, x39 e Xf.

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso representa a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação directa com y7 identificou 2 dos 5 predictores mais influentes revelando ainda o respectivo sentido de variação. Já a modelação implícita (PLS) aponta para 3 contribuições relevantes, sugeridas pelo primeiro factor latente (*FL1*), das quais 2 correspondem a factores identificados pela modelação explícita (MOLS) embora com sinais opostos nas respectivas contribuições. As variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y7 são Xe [23.4], x37 [0.166] e x39 [0.140] - a dependência positiva indica que estas possuem uma influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio. Em contrapartida, as variáveis Xf [-23.2] e x34 [-0.181] têm um efeito benéfico - à medida que estes parâmetros aumentam, a quantidade de resíduos diminui.

Da análise de correlação da resposta y1, conclui-se que x1 [-0.605], x2 [0.519], x5 [-0.514], x12 [0.563], x37 [0.620] e x39 [-0.450] possuem apenas correlações não nulas. Na análise de PLS verifica-se que os dois primeiros factores latentes ($f=2$) são os mais relevantes. Neste caso, prevê-se a dependência de y1 com 18 predictores, designadamente: Xa, x1, x2, x3, x5, x10, x12, x16, x17, x19, x21, Xb, Xc, Xd, Xe, x37, x39 e Xf. Nesta resposta existe um modelo parcimonioso (M09) que contém os predictores Xa, x5, x8, x10, x16, x17, x20, x21 e x40.

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso representa a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação directa com y1 identificou apenas 1 dos 14 predictores mais influentes. Já a modelação implícita (PLS) aponta para 18 predictores dos quais 8 correspondem ao primeiro factor latente

(FL1). As variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y_1 são x_{16} [0.63], x_{17} [0.39], x_{40} [0.25] e x_{20} [0.22] - a dependência positiva indica que estas possuem uma influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio. Em contrapartida, as variáveis x_{21} [-0.59], $X_a/X_b/X_c/X_d/X_e/X_f$ [-0.62], x_5 [-0.78], x_8 [-0.95] e x_{10} [-1.22] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y_1 indica que o seu aumento tem como impacto a redução desse valor.

Através da análise de correlação, verificou-se apenas a forte interdependência de y_{12} com x_{84} . Na análise de Mínimos Quadrados Multivariados (PLS) verifica-se que apenas os dois primeiros factores latentes, FL1 e FL2, é que são relevantes. Neste caso a prevê-se a dependência de y_{12} com 21 predictores, designadamente: x_{55} , x_{56} , x_{57} , x_{58} , x_{59} , x_{60} , x_{68} , x_{75} , x_{76} , x_{79} , x_{80} , x_{82} , x_{84} , x_{86} , x_{87} , x_{93} , x_{101} , x_{102} , x_{103} , x_{107} e x_{108} . Na análise de Mínimos Quadrados Multivariados (MOLS) verifica-se que existe um modelo parcimonioso em MO7 que contém os predictores x_{55} , x_{68} , x_{76} , x_{84} , x_{86} , x_{87} e x_{93} .

Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso como sendo a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação identificou apenas 1 dos 7 predictores mais influentes. Já a modelação implícita (PLS) aponta para 21 contribuições relevantes das quais 14 sugeridas pelo primeiro factor latente (FL1). As variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y_{12} são x_{68} [0.51], x_{76} [2.21], x_{84} [1.05], x_{86} [2.42] e x_{93} [0.31] - a sua influência crescente agrava a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio. Em contrapartida, as variáveis x_{55} [-0.95] e x_{87} [-1.11] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y_{12} indica que o seu aumento tem como impacto a redução do seu valor.

Na análise da correlação da resposta y_9 , detectaram-se fortes interdependências com quatro predictores: X_g [0.456], X_h [0.461], x_{43} [-0.383] e x_{47} [-0.263]. Relativamente aos Mínimos Quadrados Parciais, apenas o primeiro factor latente, FL1, é que é relevante, prevendo-se a dependência de y_9 com 6 predictores, designadamente: X_g , x_{43} , x_{47} , x_{48} , x_{52} e X_h .

Nesta resposta existe um modelo parcimonioso em MO5 que contém os predictores $X_g=X_h$, x_{41} , x_{48} , x_{53} e x_{54} . Assumindo que a modelação explícita (MOLS) através do modelo parcimonioso representa a situação mais correcta em termos de análise de efeitos de factores, verifica-se que a análise de correlação directa com y_1 identificou apenas 2 dos 6 predictores mais influentes. Já a modelação implícita (PLS) aponta para 3 das 6 contribuições mais relevantes, sugeridas pelo primeiro factores latentes (FL1). As variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y_9 são X_g e X_h^* [0.82], x_{41} [0.40] e x_{54} [0.48] - influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio. Em contrapartida, as variáveis x_{48} [-0.38] e x_{53} [-0.36] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y_9 indica que o seu aumento tem como impacto a redução do seu valor.

Já na análise da correlação de y_4 detectou-se uma interdependência pouco significativa com o predictor x_{39} . Relativamente aos Mínimos Quadrados Parciais, a resposta não consegue ser bem descrita, e portanto o modelo procura suportar-se num número elevado de variáveis para tentar descreve-la. Deste

modo, a modelação implícita procura descrever a resposta y_4 com base em 19 predictores, designadamente $X_a, x_1, x_2, x_3, x_5, x_9, x_{12}, x_{14}, x_{16}, x_{17}, x_{19}, x_{20}, x_{21}, X_b, X_c, X_d, X_e, x_{39}, X_f$. Na correlação explícita, encontrou-se um modelo parcimonioso em M_{11} que contém $X_a, x_1, x_3, x_5, x_8, x_{10}, x_{12}, x_{18}, x_{19}, x_{34}$ e x_{39} .

Deste estudo fica-se com a noção de que o processo iterativo de modelação explícita para a pesquisa do modelo parcimonioso é muito poderoso. Em contrapartida, a correlação directa e a modelação explícita de PLS previam um resultado desastroso. As variáveis que apresentam maior impacto positivo sobre a quantidade de resíduo avaliada em y_4 são x_1 [6.03], x_{12} [4.87], x_{18} [0.18] e x_{39} [0.34] - influência crescente para agravar a quantidade de resíduo gerado à medida que elas forem aumentadas no seu domínio. Em contrapartida, as variáveis X_a (X_b, X_c, X_d, X_e, X_f) [-0.21], x_3 [-0.19], x_5 [-1.47], x_8 (x_{15}) [-1.36], x_{19} [-1.21] e x_{34} [-0.20] têm um efeito benéfico - a sua influência negativa sobre a quantidade de resíduo avaliada em y_4 indica que o seu aumento tem como impacto a redução do seu valor.

Por fim, devido ao défice de detalhes presentes na resposta y_5 , concluiu-se que o maior valor de correlação directa foi conseguida com x_{29} . Relativamente à modelação implícita, esta resposta falha redondamente - a modelação implícita com PLS não permite avaliar a informação contida em y_5 .

Nesta resposta não foi possível encontrar um modelo parcimonioso. Considerou-se o modelo mais completo como sendo o melhor modelo, mas este não conseguiu representar uma correlação significativa com os resultados da resposta. Assim sendo, com este conjunto de predictores não faz sentido tirar ilações sobre esta resposta

Futuramente, pretende-se que a empresa venha a beneficiar deste trabalho exaustivo, no sentido de conseguir reduzir a produção de resíduos e tornar mais eficiente o seu Sistema de Gestão. Contudo, estamos cientes de que ainda existe muito trabalho a desenvolver para completar o estudo realizado.

Fica a esperança de se terem despertado consciências para a importância e a necessidade cada vez mais prementes de se desenvolverem esforços para minorar os problemas ambientais, numa óptica de gestão eficiente dos resíduos, preferentemente compatível com o desenvolvimento económico e financeiro da empresa.

Bibliografia

- [1] A. M. Craig and M. Malek. Market structure and conduct in the pharmaceutical industry. *Pharmacology & Therapeutics*, 66(2):301–337, 1995.
- [2] S. Ahmed and M. Hassan. Survey and case investigations on application of quality management tools and techniques in smis. *Internation Journal of Quality & Reliability Management*, 20(7):795–826, 2003.
- [3] H. Bunney and B. Dale. The implementation of quality management tools and techniques a study. *The TQM magazine*, 9(3):183–189, 1997.
- [4] J. A. Nunes and M. Matias. Controversia cientifica e conflitos ambientais em portugal: O caso da co-incineracao de residuos industriais perigosos. *Revista Critica de Ciencias Sociais*, 65:129–150, 2003.
- [5] L. B. Sohn. The stockholm declaration on the human environment. *The Tarvard International Law Journal*, 14(3):423–515, 1973.
- [6] F. I. Khan and S. A. Abbas. Major accidents in process industries and an analysis of causes and consequences. *Journal of Loss Prevention in the Process Industries*, 12(5):361–378, 1999.
- [7] K. Y. Foo. A vision on the role of environmental higher education contributing to the sustainable development in malaysia. *Journal of Cleaner Production*, 61(0):6–12, 2013.
- [8] O. J. Oliveira. Guidelines for the integration of certifiable management systems in industrial companies. *Journal of Cleaner Production*, 57(0):124–133, 2003.
- [9] P. E. D. Love and Z. Iran. A project management quality cost information system for the construction industry. *Information & Management*, 40(7):649–661, 2003.
- [10] DI n. 178/2006 de 5 de setembro. Technical report, Assembleia Nacional da Republica Portuguesa, 2006.
- [11] C. Vieira, B. Machado, N. Ferraz, J. Monteiro, and S. Roque. *Manual de Gestao de Residuos Industriais*. Associacao Empresarial de Portugal, AEP, 2011.
- [12] Plano nacional de residuos industriais -2011-2020. Technical report, Agencia Portuguesa do Ambiente, APA, 2011.

- [13] Portaria n.172/2009 de 17 de fevereiro. Technical report, Assembleia Nacional da Republica Portuguesa, 2009.
- [14] Dl n.73/2011 de 17 de junho. Technical report, Assembleia Nacional da Republica Portuguesa, 2011.
- [15] Portaria n.209/2004 de 3 de março. Technical report, Assembleia Nacional da Republica Portuguesa, 2003.
- [16] S. J. Formosinho, C. A. Pio, J. H. Barros, and J. R. Cavalheiro. *Parecer relativo ao tratamento de residuos industriais perigosos*, volume 1. Principia -Publicacoes Universitarias e Cientificas, 2000.
- [17] A. Valente and V. M. M. Lobo. *Corrosao: fundamentos, prevencao e efluentes*. ECEMEI, 2000.
- [18] Dl n.183/2009 de 10 de agosto. Technical report, Assembleia Nacional da Republica Portuguesa, 2011.
- [19] N. P. Cheremisinoff. *Handbook of solid waste management and waste minimization technologies*. Elsevier Science, 2003.
- [20] G. Tchobanoglous, H. Theisen, and S. Vigil. *Integrated solid waste management: engineering principles and management issues*. McGraw-Hill, 1th edition, 1993.
- [21] S. J. Formosinho, C. A. Pio, J. H. Barros, and J. R. Cavalheiro. *Parecer relativo ao tratamento de residuos industriais perigosos*, volume 2. Principia -Publicacoes Universitarias e Cientificas, 2000.
- [22] L. N. Prista, A. C. Alves, R. Morgado, and J. S. Lobo. *Tecnologia Farmaceutica*, volume 1. Fundacao Caloust Gulbenkian, 8th edition, 2011.
- [23] *Guides to pollution prevention -the pharmaceutical industry*. Tecnology transfer, 1991.
- [24] Q. Zhu, J. Cordeiro, and J. Sarkis. Institutional pressures, dynamic capabilities and environmental management systems: Investigating the iso 9000 -environmental management system implementation linkage. *Journal of Environmental Management*, 114(0):232–242, 2013.
- [25] *Quality assurance of pharmaceuticals: a compendium of guidelines and related materials. Good manufacturing practices and inspection*, volume 2. World Health Organization, 2th edition, 2007.
- [26] ISO 14001:2004 -environmental management system. Technical report, International Organization for Standardization ISO, 2004.
- [27] M. Ribeiro. Norma interna de gestao de residuos. Technical report, Blupharma -Industria Farmaceutica S.A, 2014.
- [28] U. S. Government and M. E. Phillips. *Guidance for industry: quality systems approach to pharmaceutical CGMP regulations*. General books LLC, 2011.

- [29] ISO 9000:2005 -quality management systems. Technical report, International Organization for Standardization, ISO, 2005.
- [30] ISO 9001:2008 -quality management systems. Technical report, International Organization for Standardization, ISO, 2008.
- [31] Regulamento EMAS -sistema comunitario de ecogestao e auditorias. Technical report, Sistema comunitario de ecogestao e auditorias, EMAS, 2009.
- [32] I. Petrosillo, A. Marco, S. Botta, and C. Comoglio. EMAS in local authorities: Suitable indicators in adopting environmental management systems. *Ecological Indicators*, 13(1):263–274, 2012.
- [33] OHSAS 18001:2007 - safety and occupation health management systems. Technical report, occupation health and safety management systems , OHSAS, 2007.
- [34] R. Cirillo. *The Economics of Vilfredo Pareto*. Routledge, 1978.
- [35] V. Tarascio. *Pareto's Methodological Approach to Economics: A Study in the History of Some Scientific Aspects of Economic Thought*. University of North Carolina Press, 1th edition, 1968.
- [36] J. M. Andrade and M. P. Gomez-Carracedo. Notes on the use of mandel's test to check for nonlinearity in laboratory calibrations. *Anal. Methods*, 5:1145–1149, 2013.
- [37] F. E. Grubbs. Sample criteria for testing outlying observations. *Ann. Math. Stat.*, 21:27, 1950.
- [38] F.E. Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21, 1969.
- [39] R. G. Brereton. *Chemometrics: data analysis for the laboratory and chemical plant*. Wiley, 1th edition, 2003.
- [40] C. Vervaet and J. Remon. Continuous granulation in the pharmaceutical industry. *Chemical Engineering Science*, 60(14):3949–3957, 2005.
- [41] L. Lachman, H. Lieberman, and J. Kanig. *Teoria e pratica na industria farmaceutica*, volume 2. Fundacao calouste Gulbenkian Lisboa, 2th edition, 2001.
- [42] A. Mendez, G. Carli, and C. Garcia. Evaluation of powder mixing operation during batch production: Application to operational qualification procedure in the pharmaceutical industry. *Powder Technology*, 198(2):310–313, 2010.
- [43] L. Lachman, H. Lieberman, and J. Kanig. *Teoria e pratica na industria farmaceutica*, volume 1. Fundacao calouste Gulbenkian Lisboa, 2th edition, 2001.
- [44] D. Kremer. A numerical investigation of air flow during tablet compression. *Chemical Engineering Science*, 61(24):7963–7978, 2006.

Apêndice A

Tabelas

Tabela A.1: Valores críticos para teste de Grubbs ($\alpha = 0.05$) [ISO 17025, ASTM E-178].

$$G = \frac{|x? - \bar{x}|}{s_x}$$

n	3	4	5	6	7	8	9	10	11
G	1.15	1.48	1.71	1.89	2.02	2.13	2.21	2.29	2.34
n	12	13	14	15	16	17	18	19	20
G	2.41	2.46	2.51	2.55	2.54	2.62	2.65	2.68	2.71
n	21	22	23	24	25	30	35	40	50
G	2.73	2.76	2.78	2.80	2.82	2.91	2.98	3.04	3.13
n	60	70	80	90	100	110	120	130	140
G	3.20	3.26	3.31	3.35	3.38	3.42	3.44	3.47	3.49

Tabela A.2: Valores críticos para o teste de Cochran referente à homogeneidade da variância ($\alpha = 0.05$).

$$C = \frac{s_{max}^2}{\sum_{i=1}^n s_i^2}$$

$m = \text{número de réplicas da estimativa de cada variância}$								
n	2	3	4	5	6	7	10	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8010	0.5000
3	0.9969	0.8709	0.7977	0.7457	0.7071	0.6771	0.6167	0.3333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5017	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4214	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3682	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3259	0.1429
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2439	0.1000
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1357	0.0500
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.0958	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0745	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0520	0.0167

Tabela A.3: Valores normalizados para o estudo da resposta y7.

xe	x34	x37	x38	x39	xf	x40	y7
0.588	-1.125	0.900	-0.724	-0.152	0.489	-1.123	2.666
0.598	-1.125	1.119	-0.489	-0.097	0.601	-0.301	0.231
0.692	-1.125	1.132	-0.489	-0.143	0.669	0.520	0.876
0.597	-1.125	0.926	-0.489	-0.060	0.582	1.342	0.661
0.595	-1.125	0.617	-0.489	-0.152	0.628	2.985	-0.485
0.584	-1.125	0.874	-0.394	-0.143	0.607	2.163	-0.270
-1.322	-1.125	-0.451	0.077	-0.642	-1.362	-1.123	1.055
-1.383	0.584	-0.233	-0.017	-0.614	-1.399	-0.301	-0.495
-1.383	0.584	-0.233	-0.017	-0.614	-1.399	-0.301	-0.495
-1.383	0.584	-0.233	-0.017	-0.614	-1.399	-0.301	-0.495
-1.851	0.584	-1.301	1.679	4.837	-1.822	-1.123	-0.628
0.592	-0.271	1.158	-0.017	-0.476	0.591	-1.123	0.339
0.674	-0.271	1.261	-0.300	-0.152	0.696	-0.301	-0.198
0.575	-0.271	0.849	-0.630	-0.152	0.582	1.342	0.124
-2.827	-0.271	1.351	-0.300	-0.152	-2.802	0.520	-0.628
0.466	1.011	-1.250	-0.630	0.337	0.492	-0.301	-0.825
0.466	1.011	-1.250	-0.630	0.337	0.492	-0.301	-0.825
0.466	1.011	-1.250	-0.630	0.337	0.492	-0.301	-0.825
0.466	1.011	-1.250	-0.630	0.337	0.492	-0.301	-0.825
0.426	1.011	-1.237	-0.441	0.402	0.446	0.520	-0.789
0.426	1.011	-1.237	-0.441	0.402	0.446	0.520	-0.789
0.426	1.011	-1.237	-0.441	0.402	0.446	0.520	-0.789
0.426	1.011	-1.237	-0.441	0.402	0.446	0.520	-0.789
0.609	1.011	0.772	-0.112	-0.337	0.631	0.520	-0.234
0.609	1.011	0.900	-0.347	-0.134	0.634	-0.301	-0.306
0.603	0.584	1.261	0.878	-0.707	0.533	-1.123	1.985
-1.161	-1.552	0.089	-0.347	-0.504	-1.145	-0.301	-0.270
-1.180	-1.979	0.321	4.083	-1.169	-1.200	-1.123	0.589
0.814	-1.125	-0.503	1.632	-0.448	0.712	-1.123	2.773
0.792	0.584	-0.632	1.114	-0.328	0.820	-0.301	-0.341