



Márcia Patrícia Ferreira Determinação do número de octano por
Gonçalves cromatografia gasosa



Márcia Patrícia Ferreira Gonçalves **Determinação do número de octano por cromatografia gasosa**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Métodos Instrumentais e Controlo de Qualidade Analítica, realizada sob a orientação científica do Doutor Armando Silvestre e do Doutor João Coutinho, Professores Associados do Departamento de Química da Universidade de Aveiro.

o júri

presidente:

Doutor **Armando da Costa Duarte**, Professor Associado da Universidade de Aveiro.

Doutor **Joaquim Carlos Gomes Esteves da Silva**, Professor Associado com Agregação da Faculdade de Ciências da Universidade do Porto.

Doutor **João Manuel da Costa e Araújo Pereira Coutinho**, Professor Associado com Agregação da Universidade de Aveiro. (**Co-Orientador**)

Doutor **Armando Jorge Domingues Silvestre**, Professor Associado com Agregação da Universidade de Aveiro. (**Orientador**)

Doutor **João António Baptista Pereira de Oliveira**, Professor Associado da Universidade de Aveiro.

agradecimentos

Este trabalho só foi possível de concretizar com o apoio de várias pessoas. A todas elas deixo aqui o meu sincero agradecimento, especialmente:

Aos meus orientadores, Doutor Armando Silvestre e Doutor João Coutinho, por toda a colaboração no desenvolvimento deste trabalho.

Ao Doutor Armando Silvestre, em especial, pela forma paciente e compreensiva com que lidou com os meus atrasos e adiamentos e por todo o apoio que me deu.

À Galp Energia – Laboratório da Refinaria do Porto, na pessoa da Dr^a Carla Santos, pela disponibilidade que me deu para poder frequentar esta pós-graduação e por permitir o uso de dados da empresa para o desenvolvimento deste trabalho.

À Dr^a Carla Santos, pelo incentivo e apoio que sempre me deu ao longo destes dois últimos anos, pela bibliografia dispensada e por todo o tempo que amavelmente dedicou à análise dos resultados.

Ao Doutor João Oliveira, pela disponibilidade demonstrada para a colaboração neste trabalho e por todo o tempo que amavelmente dedicou à análise estatística dos resultados.

palavras-chave

Número de Octano, Reformado, Cromatografia Gasosa Multidimensional, Regressão Linear Múltipla.

resumo

A gasolina é uma mistura complexa de hidrocarbonetos com ponto de ebulição <200°C, em que os hidrocarbonetos que constituem esta gama de ebulição têm 4 a 12 átomos de carbono. Esta pode ser produzida por vários processos, e devido às suas diferentes composições é necessário elaborar misturas, dos chamados Componentes de Gasolina, processo normalmente designado por *Blending*. O Reformado é uma das correntes de refinaria usada como Componente de Gasolina.

No sentido de caracterizar o comportamento da gasolina e seus componentes, na resistência à detonação, foi introduzido o conceito de *Número de Octano*. A determinação do número de octano é feita pelos métodos ASTM D 2699 e ASTM D 2700, respectivamente designados por RON (Research Octane Number) e MON (Motor Octane Number), através de um motor de referência, designado por CFR (Cooperative Fuel Research).

O que se propõe neste trabalho é o desenvolvimento de um método alternativo (modelo matemático por regressão linear múltipla), mais económico e mais rápido para a determinação do Número de Octano, no Reformado a partir da sua composição. A composição é determinada por cromatografia gasosa multidimensional que separa a amostra por grupos de compostos (*i*-parafinas, *n*-parafinas, *i*-olefinas, *n*-olefinas, *c*-olefinas, Nafténicos e Aromáticos) e dentro de cada grupo por números de carbono.

Para tal foram consideradas as amostras de Reformado no período de 2003 a 2006 e desenvolvidos vários modelos pelo método de selecção de regressores *Stepwise*, recorrendo ao SPSS. O melhor modelo obtido apresenta um R^2_{ajustado} de 0,80, e de acordo com o teste *t*, os resultados obtidos pelo motor CFR e pelo modelo não são estatisticamente diferentes. A diferença entre o RON determinado pelo motor CFR e pelo modelo é apenas em 12% dos casos superior à reprodutibilidade de método ASTM D 2699 (0,7).

keywords

Octane Number, Reformate, Multidimensional Gas Chromatography, Multiple Linear Regression.

abstract

Gasoline is a complex mixture of hydrocarbons with a boiling point $<200^{\circ}\text{C}$, the hydrocarbons that constitute this range of boiling points can have 4 to 12 carbon atoms. This can be produced from various production processes and due to the different composition it is necessary blend this different fuel streams, referred to as Gasoline Components, this process is known as Blending. Reformate is one of the refinery streams used as Gasoline Component. To characterize the anti-knocking behaviour of gasoline and its components it was introduced the concept of Octane Number. The Octane Number is measured in a standardized CFR (Cooperative Research Fuel) engine following the standard methods ASTM D 2699 and ASTM D 2700, respectively known as RON (Research Octane Number) and MON (Motor Octane Number). The purpose of this work is the development of an alternative method (Multiple Linear Regression Model) more economic and faster for the determination of the Octane Number, in Reformate from its composition. The composition is determined by multidimensional gas chromatography that separates the sample in groups of components (*i*-paraffins, *n*-paraffins, *i*-olefins, *n*-olefins, *c*-olefins, naphthenes and aromatics) and within each group in carbon numbers. For this purpose it was considered the Reformate samples between 2003 and 2006 and several models were developed based on the *Stepwise* method, recurring to SPSS. The best model showed an $R^2_a=0,80$ and according with the *t test* the results of the CFR engine and model were not statistically different. The difference between the RON that was determined by the CFR engine and the model is only superior in 12% of the cases of the reproducibility method ASTM D 2699 (0,7).

ÍNDICE

1. INTRODUÇÃO	1
2. CARACTERIZAÇÃO DE PETRÓLEOS BRUTOS E PRODUTOS DE REFINAÇÃO	5
2.1 COMPOSIÇÃO DOS PETRÓLEOS BRUTOS	5
2.2 HIDROCARBONETOS PRESENTES NOS PETRÓLEOS BRUTOS	6
2.3 CLASSIFICAÇÃO DOS DIFERENTES PETRÓLEOS BRUTOS	8
2.4 PRODUTOS DA REFINAÇÃO DO PETRÓLEO BRUTO	8
3. CARACTERIZAÇÃO DA GASOLINA	10
3.1 PROCESSO DE PRODUÇÃO DA GASOLINA	10
3.2 ESPECIFICAÇÃO COMERCIAL DA GASOLINA	12
3.3 CARACTERÍSTICAS FUNDAMENTAIS DA GASOLINA	12
4. APLICAÇÃO DA GASOLINA NO MOTOR	16
4.1 FUNCIONAMENTO DO MOTOR A GASOLINA	16
4.2 FENÓMENO DE DETONAÇÃO	17
4.3 DEFINIÇÃO DO NÚMERO DE OCTANO	18
4.4 NÚMERO DE OCTANO vs DESEMPENHO DO MOTOR	19
4.5 NÚMERO DE OCTANO DAS DIFERENTES FAMÍLIAS DE HIDROCARBONETOS	20
5. MÉTODOS PARA A DETERMINAÇÃO DO NÚMERO DE OCTANO	22
5.1 MÉTODOS DE REFERÊNCIA – ASTM D 2699 E ASTM D 2700	22
5.2 MÉTODOS ALTERNATIVOS	25
5.3 APLICAÇÃO DOS MÉTODOS ALTERNATIVOS AO REFORMADO	34

6. METODOLOGIA PROPOSTA NESTE TRABALHO PARA A DETERMINAÇÃO DO RON	37
6.1 AMOSTRAS	37
6.2 DADOS DO MOTOR CFR	38
6.3 DADOS CROMATOGRÁFICOS	38
6.4 ANÁLISE ESTATÍSTICA DOS DADOS	42
6.4.1 Programa Informático para Estatística- SPSS	42
6.4.2 O modelo de regressão linear	43
6.4.3 Estimação dos coeficientes de regressão – métodos dos mínimos quadrados	44
6.4.4 Inferências sobre o modelo de regressão linear	45
6.4.4.1 A análise de variância do modelo de regressão linear	45
6.4.4.2 Testes aos coeficientes do método de regressão	46
6.4.4.3 O coeficiente de determinação	47
6.4.5 Validação dos pressupostos do modelo de regressão linear	48
6.4.5.1 A análise de resíduos	48
6.4.5.2 Multicolinearidade	49
6.4.6 Seleccção de regressores	49
6.4.6.1 Método Exaustivo	50
6.4.6.2 Método Progressivo (ou <i>Forward</i>)	50
6.4.6.3 Método Regressivo (ou <i>Backward</i>)	51
6.4.6.4 Método de Regressão Passo a Passo (ou <i>Stepwise</i>)	51
6.5 ABORDAGEM AOS DADOS	51
7. RESULTADOS OBTIDOS COM A METODOLOGIA PROPOSTA	53
7.1 MODELO COM 4 VARIÁVEIS INDEPENDENTES	53
7.1.1 Análise das Variáveis	53
7.1.2 Estimação do modelo	55
7.1.3 Validação dos pressupostos do modelo estimado	58

7.1.3.1 Homocedasticidade	58
7.1.3.2 Covariância Nula	59
7.1.3.3 Normalidade dos Resíduos	59
7.1.3.4 Multicolinearidade	60
7.1.4 Análise dos Valores Atípicos	60
7.2 MODELO COM 4 VARIÁVEIS INDEPENDENTES EXCLUÍDO OS VALORES ATÍPICOS	62
7.2.1 Análise das Variáveis	62
7.2.2 Estimação do Modelo	62
7.2.3 Validação dos pressupostos do modelo estimado	65
7.2.3.1 Homocedasticidade	65
7.2.3.2 Covariância Nula	65
7.2.3.3 Normalidade dos Resíduos	65
7.2.3.4 Multicolinearidade	66
7.2.4 Previsão do RON usando o modelo	66
7.3 MODELO COM 7 VARIÁVEIS INDEPENDENTES	69
7.3.1 Análise das Variáveis	69
7.3.2 Estimação do Modelo	71
7.3.3 Validação dos pressupostos do modelo estimado	74
7.3.3.1 Homocedasticidade	74
7.3.3.2 Covariância Nula	74
7.3.3.3 Normalidade dos Resíduos	74
7.3.3.4 Multicolinearidade	75
7.3.4 Previsão do RON usando o modelo	76
7.4 MODELO COM 47 VARIÁVEIS INDEPENDENTES	77
7.4.1 Análise das Variáveis	77
7.4.2 Estimação do Modelo	78
7.4.3 Validação dos pressupostos do modelo estimado	81
7.4.3.1 Homocedasticidade	81
7.4.3.2 Covariância Nula	81

7.4.3.3 Normalidade dos Resíduos	82
7.4.3.4 Multicolinearidade	82
7.4.4 Previsão do RON usando o modelo	83
7.5 MODELO COM 47 VARIÁVEIS NORMALIZADAS	84
7.6 MODELO COM 47 VARIÁVEIS INDEPENDENTES DADOS DE 2003	85
7.6.1 Análise das Variáveis	85
7.6.2 Estimação do Modelo	86
7.6.3 Validação dos pressupostos do modelo	88
7.6.3.1 Homocedasticidade	88
7.6.3.2 Covariância Nula	88
7.6.3.3 Normalidade dos Resíduos	89
7.6.3.4 Multicolinearidade	90
7.6.4 Previsão do RON usando o modelo	90
7.7 MODELO COM 47 VARIÁVEIS COM OS DADOS DE 2003 A 2004	92
7.7.1 Análise das Variáveis	92
7.7.2 Estimação do Modelo	92
7.7.3 Diagnóstico do Modelo Estimado	94
7.7.3.1 Homocedasticidade	94
7.7.3.2 Covariância Nula	95
7.7.3.3 Normalidade dos Resíduos	95
7.7.3.4 Multicolinearidade	95
7.7.4 Previsão do RON usando o modelo	96
7.8 MODELO COM 47 VARIÁVEIS COM OS DADOS DE 2003 A 2005	97
7.8.1 Análise das Variáveis Independentes	97
7.8.2 Estimação do Modelo	97
7.8.3 Diagnóstico do Modelo Estimado	100
7.8.3.1 Homocedasticidade	100
7.8.3.2 Covariância Nula	100
7.8.3.3 Normalidade dos Resíduos	100

7.8.3.4 Multicolinearidade	101
7.8.4 Previsão do RON usando o modelo	101
8. DISCUSSÃO E CONCLUSÃO	104
9. REFERÊNCIAS BIBLIOGRÁFICAS	108

1. INTRODUÇÃO

O petróleo é uma mistura complexa de hidrocarbonetos constituídos essencialmente por átomos de carbono e hidrogénio. A sua formação teve origem há centenas de milhões de anos, a partir de restos vegetais e animais e encontra-se depositado no solo a várias profundidades. A sua composição exacta varia consoante o local de extracção [1].

Apesar de existirem referências históricas sobre a sua existência cerca de 3000 anos antes de Cristo, o petróleo foi olhado ao longo da maior parte da história da humanidade como uma mera curiosidade. A sua emergência como recurso estratégico de fundamental importância económica é uma realidade com pouco mais de um século, e associada ao desenvolvimento tecnológico que marcou os processos de extracção e refinação, assim como ao alargamento da sua utilização aos mais diversos sectores da actividade humana [1].

Nos dias de hoje, o petróleo assume uma importância ímpar no âmbito dos recursos estratégicos fundamentais. Os choques petrolíferos de 1973 e 1979 colocaram a descoberto a verdadeira dependência das economias mais avançadas do mundo em relação a este recurso, bem como reforçaram o sentimento de poder dos países detentores das maiores reservas [1].

A indústria petrolífera está claramente dividida em duas grandes áreas: uma que compreende a pesquisa, exploração e produção e outra que engloba o transporte, refinação e logística. Na primeira importa avaliar parâmetros tais como o domínio mineiro sobre a zona de interesse, o risco político e económico subjacente e a logística, antes de iniciar esta etapa. Na segunda, e uma vez que o petróleo bruto contém centenas de diferentes tipos de hidrocarbonetos, é necessário separá-los e transformá-los a fim de se obterem os produtos pretendidos. A diversidade desses produtos, assim como as suas características implicam um conjunto complexo de operações de refinação, as quais se podem dividir em 3 grupos fundamentais [2]:

- *Operações físicas de separação* - incluem o fraccionamento por destilação e a extracção por solvente. São operações que permitem separar vários produtos utilizando as diferenças de pontos de ebulição, a densidade ou a solubilidade em solventes selectivos [2].

- *Operações químicas de conversão molecular* - reúnem todas as operações que se destinam a alterar a estrutura molecular dos componentes do petróleo de forma a obter produtos com determinadas características. Estas operações efectuam-se em condições variáveis de pressão e temperatura, podendo utilizar-se ou não um catalisador e decorrerem em atmosfera de hidrogénio [2].
- *Operações físicas ou químicas de tratamento* - destinadas a melhorar a qualidade dos vários derivados do petróleo por remoção de certas impurezas [2].

Das áreas da indústria petrolífera, em Portugal apenas existe a área da refinação, uma vez que a pesquisa de reservas de hidrocarbonetos ainda não detectou petróleo em quantidade que permita a sua extracção no nosso país, em condições económicas favoráveis [1].

No âmbito da refinação, Portugal possui duas refinarias, uma situada no Porto e outra em Sines. Ambas são propriedade da Galp Energia, detentora do monopólio da refinação em Portugal. A Refinaria de Sines possui uma capacidade de destilação instalada de 10,4 milhões de toneladas/ano e está orientada para maximizar a produção de gasolinas. A Refinaria do Porto tem uma capacidade de destilação instalada de 4,8 milhões de toneladas/ano e está orientada para a produção de combustíveis, lubrificantes, ceras de petróleo, produtos aromáticos de base e solventes industriais. As duas refinarias dispõem de instalações tecnologicamente actualizadas sendo que as suas valências não se sobrepõem, mas complementam-se relativamente às necessidades do país [2].

A Refinaria do Porto iniciou a sua laboração em 1969 e conjuga as Fábricas de Combustíveis, Óleos Base, Lubrificantes e Aromáticos [2].

No caso da Fábrica de Combustíveis, é constituída por 14 Unidades Processuais, Figura 1, que utilizando o petróleo bruto como matéria-prima obtêm toda a gama de produtos combustíveis incluindo grande parte da matéria-prima que alimenta a Fábrica de Aromáticos. Dada a flexibilidade e interligação das suas instalações, trata uma grande variedade de petróleos brutos, escolhidos caso a caso, de acordo com as suas características e as necessidades do mercado [2].

Assim, o petróleo bruto alimenta a Unidade de Destilação Atmosférica da Fábrica de Combustíveis (Unidade 3000), que o separa em cinco grandes fracções mais simples que serão tratadas em unidades processuais a jusante, passando a constituir componentes de produtos comerciais. Estas cinco fracções são designadas por: *Produtos de Topo* (leves), *Gasolina Pesada*, *Petróleo*, *Gasóleo* e *Resíduo Atmosférico* [2].

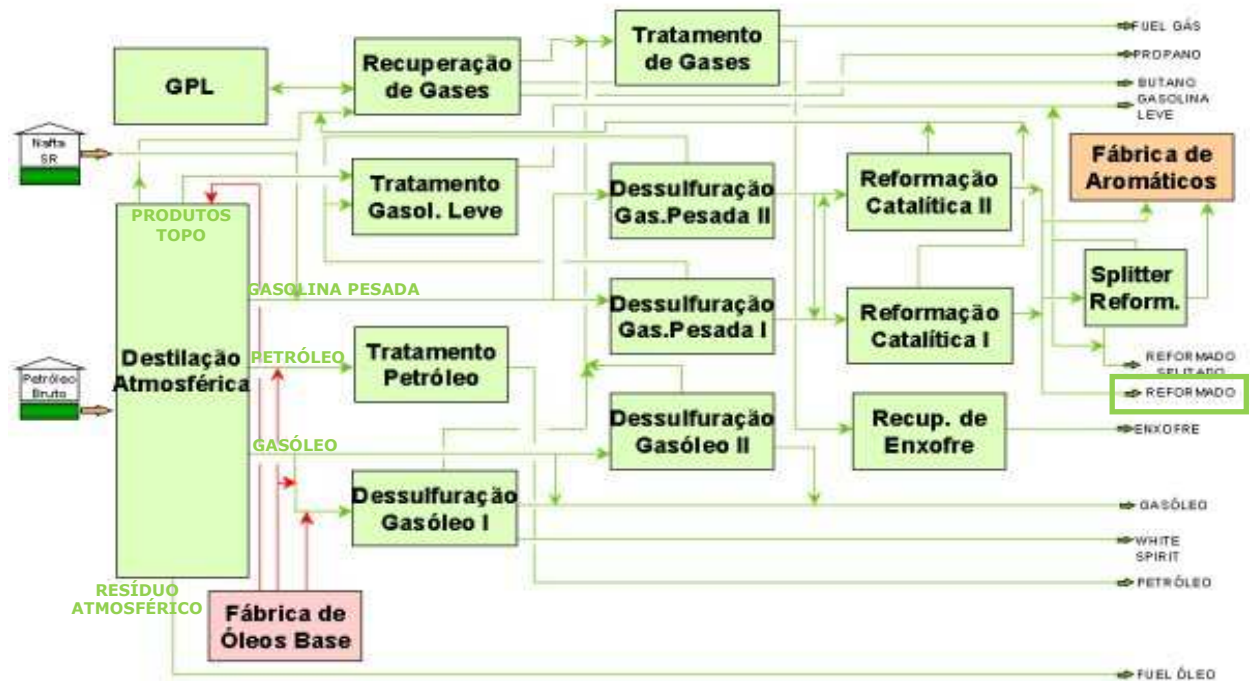


Figura 1- Esquema das várias unidades e correntes processuais que constituem a Fábrica de Combustíveis da Refinaria do Porto da Galp Energia [2].

Ao petróleo separado na destilação à pressão atmosférica são removidos alguns contaminantes e é usado como combustível para aviação designado por *Jet-Fuel*. O gasóleo proveniente da destilação à pressão atmosférica é dessulfurado, reduzindo drasticamente o seu teor em enxofre. O produto de topo é redestilado, dando origem a gases incondensáveis (*Fuel Gás*), gases de petróleo liquefeitos (*GPL*), naftas leve e pesada aos quais são posteriormente removidos os compostos de enxofre. A gasolina pesada é enviada para a Unidade de *Platforming semi-regenerativo e de regeneração contínua* a qual converte os hidrocarbonetos nafténicos e parafínicos em aromáticos (*reformação catalítica*). O produto obtido é designado por *Reformado*,

apresenta um elevado teor de hidrocarbonetos aromáticos e um número de octano também bastante elevado. Este produto é usado como componente de gasolinas ou como matéria-prima para a Fábrica de Aromáticos [2].

As correntes processuais das várias unidades produtivas de cada uma das fábricas acima referidas, são analisadas de acordo com o plano de inspecção e ensaios (PIE) de cada fábrica. Este plano foi elaborado por cada uma das fábricas em conjunto com o laboratório e o seu objectivo é definir quais as características analisadas em cada uma das correntes processuais e qual a periodicidade dessas análises. O cumprimento deste plano permite um controlo diário do processo de produção, a partir dos resultados obtidos para cada uma das características analisadas nas várias unidades. Nas fases de movimentação de produtos e expedição, todos os produtos finais são também analisados a fim de verificar a conformidade com as especificações em vigor antes da sua expedição. Neste sentido o Laboratório da Refinaria do Porto está equipado com todos os equipamentos necessários à realização dos ensaios químicos, físico-químicos e mecânicos que garantam o controlo processual e a comprovação da qualidade dos produtos [2].

A presente dissertação procura contribuir para a melhoria contínua do desempenho do Laboratório. Assim o que se propõe neste trabalho é o desenvolvimento de um método alternativo, mais rápido e económico para a determinação do *Número de Octano*, no *Reformado*, a partir da sua composição cromatográfica. Este método tem por base um modelo de Regressão Linear Múltipla que permite a partir de um único ensaio determinar duas características, reduzindo assim o tempo e os recursos humanos necessários e melhorando com isso a capacidade de resposta do Laboratório.

Face ao objectivo do trabalho, e uma vez que mais à frente será usada a composição do *Reformado* para o desenvolvimento do modelo, no ponto seguinte será feita uma caracterização da composição do petróleo bruto uma vez que está na origem de todo o processo de refinação.

De seguida será feita também uma caracterização da gasolina, uma vez que o *Reformado*, corrente processual em estudo neste trabalho, é um dos componentes usado na produção de gasolina. Assim será apresentada a sua forma de produção, composição, características de interesse, e a importância de algumas delas no funcionamento dos motores automóveis.

2. CARACTERIZAÇÃO DE PETRÓLEOS BRUTOS E PRODUTOS DE REFINAÇÃO

2.1 COMPOSIÇÃO DOS PETRÓLEOS BRUTOS

Os petróleos brutos são misturas naturais de hidrocarbonetos e de derivados orgânicos de enxofre, azoto e oxigénio, existentes no subsolo, no estado líquido. A estas misturas juntam-se normalmente outras substâncias, designadamente água, matéria inorgânica e gás sulfídrico, cuja posterior separação não altera a composição orgânica da mistura original [3].

Do ponto de vista físico, o petróleo bruto varia desde líquidos relativamente pouco viscosos, de cor palha escura, até fluídos muito viscosos, semi-sólidos. A sua massa volúmica oscila entre 790 e 1010 kg/m³ e a viscosidade varia muito desde 0,7 até 42.000 cP. Os petróleos brutos são substâncias inflamáveis em condições ambientais e exalam cheiros que variam dum quase fragância até um odor repulsivo (característico dos compostos orgânicos de enxofre) [3].

Apesar da grande diversidade observada nas características físicas, tal não se verifica no que diz respeito à composição elementar, que varia entre limites relativamente estreitos, Tabela 1. A explicação para esta particularidade é o facto dos diferentes petróleos brutos serem constituídos por hidrocarbonetos pertencentes a reduzidas séries ou famílias, Tabela 2 [3].

Tabela 1- Composição elementar média dos petróleos brutos [3].

Elemento	%m/m
Carbono	83,9-86,8
Hidrogénio	11,0-14,0
Enxofre	0,04-6,0
Azoto	0,10-1,50
Oxigénio	0,10-0,50
Metais	0,005-0,015

Tabela 2-Composição típica de alguns petróleos brutos de diferentes origens [4].

Fonte de Petróleo Bruto	Parafinas (%vol)	Aromáticos (%vol)	Naftênicos (%m/m)	Enxofre (%m/m)	Densidade API	Número de Octano
Nigéria (Leve)	37	9	54	0,2	36	60
A.Saudita (Leve)	63	19	18	2	34	40
A.Saudita (Pesado)	60	15	25	2,1	28	35
Venezuela (Pesado)	35	12	53	2,3	30	60
Venezuela (Leve)	52	14	34	1,5	24	50
USA- Midcont.	-	-	-	0,4	40	-
USA- Texas	46	22	32	1,9	32	55
Brent- Mar do Norte	50	16	34	0,4	37	50

2.2 HIDROCARBONETOS PRESENTES NOS PETRÓLEOS BRUTOS

Os petróleos brutos são constituídos por moléculas de hidrocarbonetos que podem ter desde 1 a 60 átomos de carbono. As propriedades desses hidrocarbonetos dependem do número e dos diferentes arranjos dos átomos de carbono e hidrogênio nas moléculas [3].

Os hidrocarbonetos que contêm até 4 átomos de carbono estão normalmente no estado gasoso, os que têm entre 5 e 19 átomos de carbono estão no estado líquido e os que têm 20 átomos de carbono, ou mais, encontram-se no estado sólido. Estes hidrocarbonetos agrupam-se da seguinte forma [4]:

- **Parafinas**, também designadas por alcanos, são hidrocarbonetos lineares, as *n*-parafinas, ou ramificados, as *i*-parafinas, de fórmula C_nH_{2n+2} normalmente com 1 a 20 átomos de carbono. Exemplos de parafinas: metano, etano, propano, *n*-butano, isobutano, pentano, *n*-hexano, *n*-heptano, *n*-octano, isoctano, etc, Figura 2 [4].

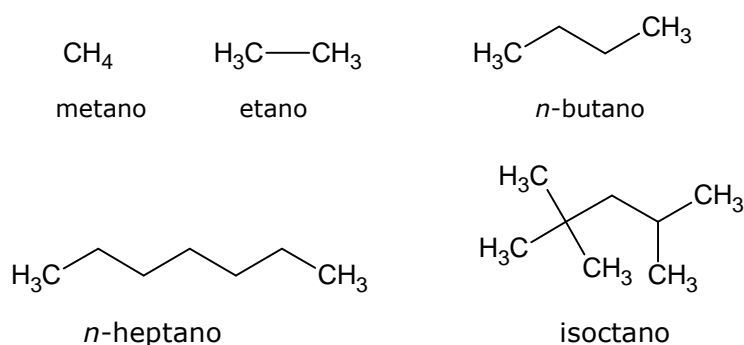


Figura 2- Estrutura química de algumas parafinas.

- **Nafténicos**, também designados por cicloparafinas ou cicloalcanos, são hidrocarbonetos com estrutura em anel mas sem ligações duplas, de fórmula C_nH_{2n} e com um ou mais anéis por molécula. Exemplos de hidrocarbonetos nafténicos: ciclopentano, ciclohexano, etc, Figura 3 [4].

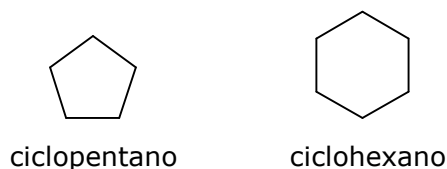


Figura 3- Estrutura química de alguns compostos nafténicos.

- **Olefinas**, também designadas por alcenos, são hidrocarbonetos lineares, ramificados ou cíclicos, com ligações duplas de fórmula C_nH_{2n} . Exemplos de olefinas: etileno, buteno, isobuteno, etc, Figura 4. Estes compostos raramente estão presentes no petróleo bruto, e surgem normalmente durante o processo de refinação [4].

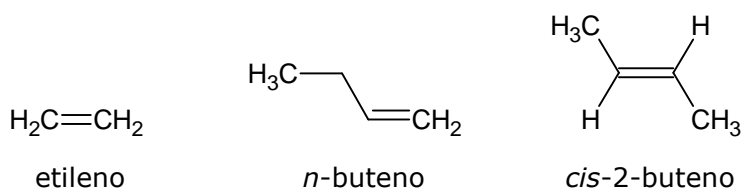


Figura 4- Estrutura química de algumas olefinas.

- **Aromáticos**, são hidrocarbonetos com estruturas em anel (anel de benzeno com 6 átomos de carbono e ligações duplas alternadas), com um ou mais anéis por molécula. Exemplos de hidrocarbonetos aromáticos: benzeno, tolueno, *m*-xileno, naftaleno, etc, Figura 5 [4].

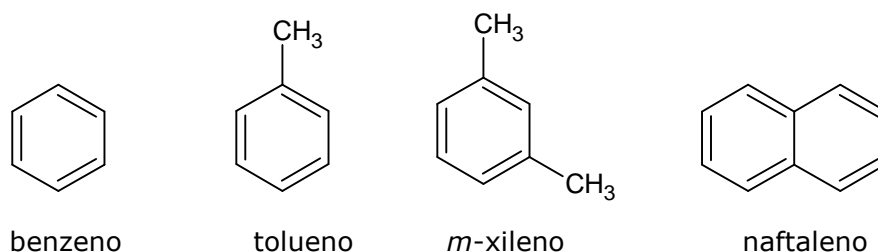


Figura 5- Estrutura química de alguns compostos aromáticos.

2.3 CLASSIFICAÇÃO DOS DIFERENTES PETRÓLEOS BRUTOS

Dada a diversidade de características dos petróleos brutos é frequente agrupá-los segundo uma propriedade física ou química comum [3].

A densidade API (American Petroleum Institute) é uma característica física usada para agrupar os petróleos brutos e classifica-os em quatro grandes categorias: Leves, Médios, Pesados e Muito Pesados. Quanto maior for a densidade API mais leve é o petróleo e mais baixa é a massa volúmica [3].

Os petróleos brutos também são agrupados por famílias, de acordo com as famílias predominantes na sua composição química. Designam-se por: *Petróleos Brutos Parafínicos*, *Nafténico-Parafínicos*, *Nafténicos*, *Aromáticos*, *Aromático-Asfálticos* e *Aromático-Nafténicos* [3].

Uma outra forma de agrupar os petróleos brutos considera como característica distintiva o teor em enxofre total. De acordo com este critério os petróleos brutos dividem-se em 'sweet' e 'sour' consoante o teor em enxofre é menor que 1% m/m ou maior e igual a 1% m/m respectivamente [3].

2.4 PRODUTOS DA REFINAÇÃO DO PETRÓLEO BRUTO

Os diferentes componentes do petróleo bruto têm pesos moleculares diferentes e moléculas com diferentes configurações e, como tal, diferentes pontos de ebulição. Assim é possível separá-los por destilação em que no topo da coluna se recolhem os hidrocarbonetos mais leves (essencialmente GPL). Na base da coluna, no resíduo, recolhem-se os hidrocarbonetos mais pesados e normalmente recolhem-se ao longo da coluna de destilação produtos intermédios tais como nafta, gasolina e gasóleo, por ordem crescente de pontos de ebulição [5]. Tal como esquematizado na [Figura 6](#) e já mencionado no ponto 1 quando foi apresentada a Unidade 3000 da Fábrica de Combustíveis da Refinaria do Porto da Galp Energia.

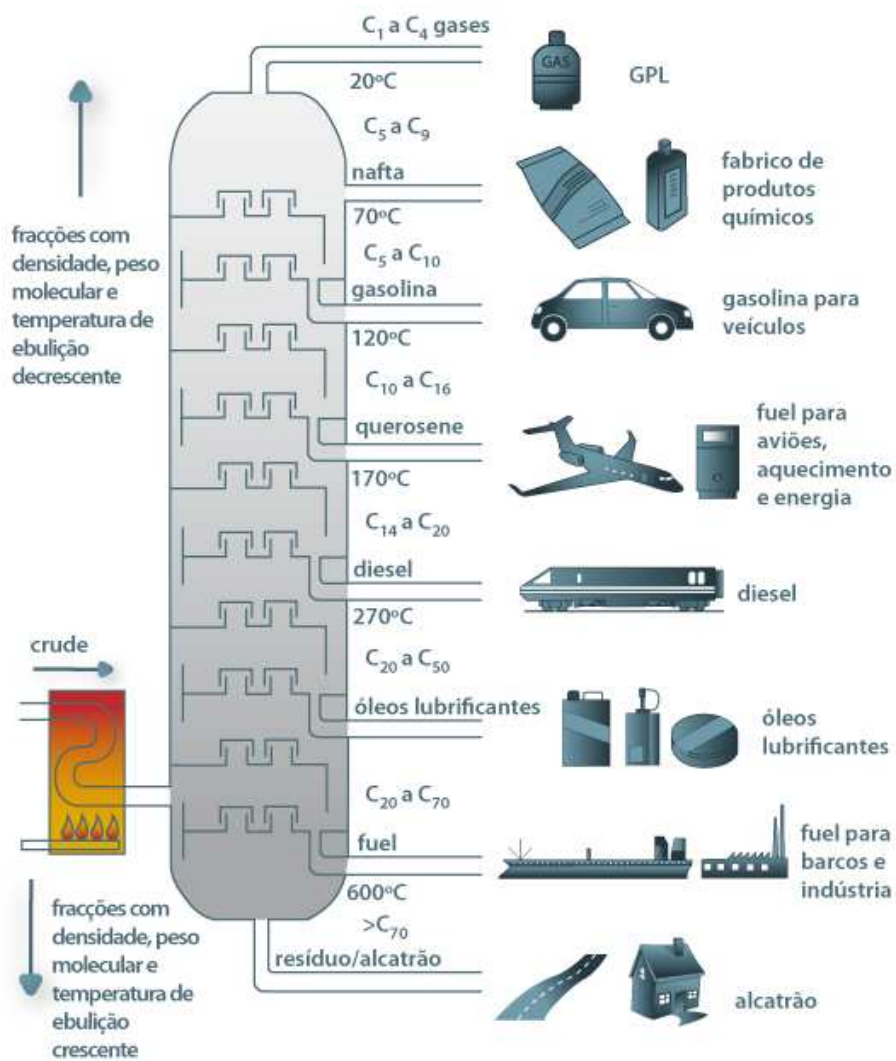


Figura 6- Esquema da destilação fraccionada do petróleo bruto [5].

3. CARACTERIZAÇÃO DA GASOLINA

3.1 PROCESSO DE PRODUÇÃO DA GASOLINA

A gasolina é uma combinação complexa de hidrocarbonetos, constituídos principalmente por compostos parafínicos, nafténicos, aromáticos e olefínicos, com número de carbonos predominantemente maiores que C₃ (de C₄ a C₁₂) e uma gama de temperaturas de ebulição de 30°C a 220°C [6].

Esta combinação é produzida a partir da mistura de várias correntes de refinaria de modo a obter um produto final com as características desejadas. Muitas gasolinas na Europa contêm também componentes de origem não petrolífera especialmente oxigenados (maioritariamente éteres e álcoois), Figura 7, e aditivos que podem ser usados para melhorar o desempenho do combustível [6].

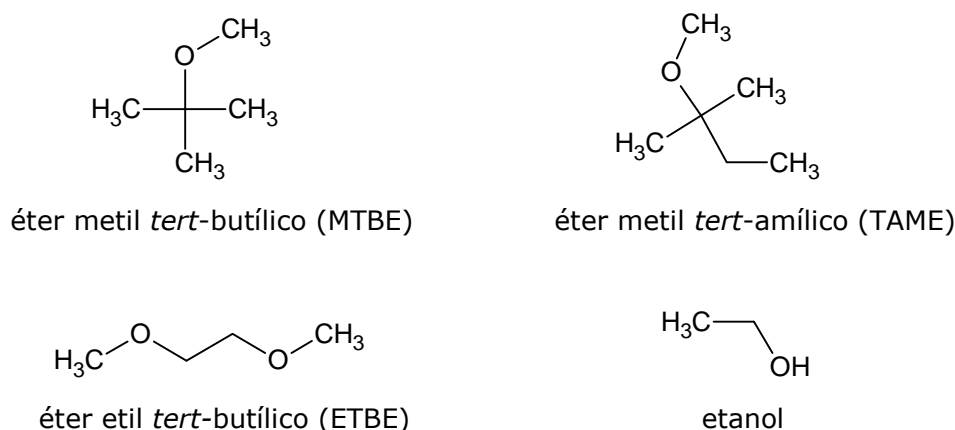


Figura 7- Estrutura química de alguns dos compostos oxigenados adicionados às gasolinas.

A composição dos hidrocarbonetos presentes na gasolina varia muito, dependendo do tipo e da natureza de cada petróleo bruto processado, dos processos de refinação disponíveis, das condições do processo e do balanço global exigido entre a gasolina e os outros produtos de refinação [6].

As correntes de refinaria mais usadas para a produção de gasolinas, são designadas por *Componentes de Gasolina* e derivam de vários processos produtivos [7]:

- *Cracking Catalítico* – consiste na quebra dos hidrocarbonetos maiores e com maior ponto de ebulição, em hidrocarbonetos na gama dos

existentes na gasolina e contendo cerca de 30% de compostos aromáticos e 20-30% de olefinas [7];

- *Hidrocracking* – quebra de moléculas e hidrogenação, produzindo fracções de gasolina mais saturadas e estáveis [7];
- *Isomerização* – conversão de hidrocarbonetos de cadeia linear em hidrocarbonetos de cadeia ramificada que como consequência produz um aumento do número de octano da gasolina [7];
- *Reforming* – converte os hidrocarbonetos saturados, com baixo número de octano, em hidrocarbonetos com elevado número de octano contendo cerca de 60% de compostos aromáticos [7];
- *Alquilação* – reacção de uma corrente gasosa de olefinas com isobutano de modo a produzir *i*-parafinas com elevado número de octano [7].

Devido às diferentes composições dos *Componentes de Gasolina*, para se obter uma gasolina final, é necessário proceder à sua mistura - *Blending*. O processo físico de *Blending* de componentes é simples mas a escolha dos componentes a usar, assim como a percentagem de cada um na mistura, é mais complexo [8]. A operação é feita por bombagem simultânea de todos os componentes em que cada propriedade, de cada componente, é uma variável e o seu efeito na mistura final é considerável. Assim é fundamental conhecer as características de cada componente antes de proceder ao seu *Blending*.

Nos últimos anos surgiu uma preocupação crescente dos governos em criar regulamentos para proteger o meio ambiente. Muitos desses regulamentos foram feitos com o objectivo de minimizar as emissões dos automóveis, nomeadamente em CO (monóxido de carbono), CO₂ (dióxido de carbono), NO_x (compostos de azoto), SO_x (compostos de enxofre), VOCs (compostos orgânicos voláteis) e partículas, para a atmosfera [7].

Inicialmente a maioria dos regulamentos foram dirigidos para os automóveis e tiveram resultados nas tecnologias, aumentando a eficiência dos motores, o que reduziu significativamente as emissões [7]. Posteriormente as regulamentações foram dirigidas para a mudança da composição dos combustíveis (fim dos aditivos à base de compostos de chumbo, menor volatilidade, menor teor em olefinas e aromáticos) com o objectivo de reduzir ainda mais as emissões dos veículos e torná-las menos tóxicas [9].

Estas novas considerações ambientais alteraram as especificações comerciais das gasolinas e introduziram uma série de dificuldades na sua produção [9]. As refinarias tiveram que alterar os seus processos de fabrico de modo a produzir gasolinas menos voláteis, com menor teor em olefinas e em compostos aromáticos, nomeadamente benzeno. Assim passaram a incorporar componentes de origem não petrolífera, compostos oxigenados e outros aditivos, com o objectivo de compensar as perdas no desempenho do combustível provocadas por estas variações na composição [7].

3.2 ESPECIFICAÇÃO COMERCIAL DA GASOLINA

Actualmente a composição da gasolina varia de refinaria para refinaria e não é apenas um único critério que caracteriza a sua qualidade [9]. A gasolina, para além de satisfazer as exigências dos motores automóveis, está adicionalmente exposta a uma variedade de condicionantes mecânicas, físicas e químicas. Assim as suas propriedades têm que ser um compromisso entre o bom desempenho do motor, perante condições operatórias adversas, e o cumprimento dos parâmetros ambientais [10].

Em 2004, a Directiva Europeia 2003/17/CE que regulamenta as especificações de Gasolinas e Gasóleos foi transposta para a lei Portuguesa pelo Decreto de Lei 235/2004. Esta especificação tem como objectivo assegurar que toda a gasolina comercializada apresenta um padrão de qualidade mínimo para as características fundamentais. A Tabela 3 [11] apresenta os valores das especificações para essas características nas Gasolinas Euro Super e Super Plus.

3.3 CARACTERÍSTICAS FUNDAMENTAIS DA GASOLINA

Todas as características presentes na especificação comercial das gasolinas são importantes e têm que ser cumpridas, umas por motivos ambientais, outras para permitirem o bom funcionamento dos motores automóveis. No caso das características associadas ao motor há quatro que têm predominante importância: a volatilidade, a estabilidade, a corrosão, e as características de combustão [7].

Tabela 3 – Especificações para as principais características das Gasolinas Euro Super e Super Plus [11].

Característica	Unidades	Euro Super		Super Plus	
		Mínimo	Máximo	Mínimo	Máximo
Aspecto		Claro e límpido		Claro e límpido	
Cor		Violeta		Violeta	
Massa Volúmica a 15°C	Kg/m ³	720	775	720	775
Número de Octano - RON		95	-	98	-
Número de Octano - MON		85	-	87	-
Pressão de Vapor:					
De 1 de Maio a 30 de Setembro	kPa	45,0	60,0	45,0	60,0
Meses de Outubro a Abril	kPa	60,0	90,0	60,0	90,0
De 1 de Novembro a 31 de Março	kPa	60,0	90,0	60,0	90,0
Destilação					
Evaporado a 70°C					
De 1 de Maio a 30 de Setembro	%v/v	20,0	48,0	20,0	48,0
De 1 de Outubro a 30 de Abril	%v/v	22,0	50,0	22,0	50,0
Evaporado a 100°C	%v/v	46,0	71,0	46,0	71,0
Evaporado a 150°C	%v/v	75,0	-	75,0	-
Ponto Final	°C	-	210	-	210
Resíduo	%v/v		2	-	2
Análise de hidrocarbonetos					
Olefinas	%v/v	-	18,0	-	18,0
Aromáticos	%v/v	-	42,0	-	42,0
Benzeno	%v/v	-	1,00	-	1,00
Teor em oxigénio	%m/m	-	2,7	-	2,7
Teor em enxofre	mg/kg	-	150	-	150
Teor de chumbo	g/l	-	0,005	-	0,005
Estabilidade à Oxidação	min	360	-	360	-
Gomas existentes (lavadas com solvente)	mg/100 ml	-	5	-	5
Corrosão à lâmina de Cobre (3h a 50°C)			1		1
Compostos Oxigenados					
Metanol	%v/v	-	3,0	-	3,0
Etanol	%v/v	-	5,0	-	5,0
Álcool Isopropílico	%v/v	-	10,0	-	10,0
Álcool Terbutílico	%v/v	-	7,0	-	7,0
Álcool Isobutílico	%v/v	-	10,0	-	10,0
Éteres com 5 ou mais átomos de carbono por molécula	%v/v	-	15,0	-	15,0
Outros Compostos Oxigenados	%v/v	-	10,0	-	10,0

A volatilidade afecta de várias formas o comportamento do motor. Se for muito baixa, o motor pode ter dificuldade em arrancar a frio e leva muito tempo a aquecer. Se for muito elevada, pode provocar problemas de falhas do motor ou funcionamento irregular em tempo quente. A volatilidade também tem efeito na economia e energia, podendo influenciar ainda a formação de depósitos no motor. A relação entre a volatilidade e o comportamento do motor está associada a grupos de hidrocarbonetos que se encontram em zonas distintas do intervalo de destilação da gasolina. Para a refinação isto representa uma limitação e portanto um cuidadoso equilíbrio dentro do intervalo de destilação, de modo que a gasolina permaneça entre os limites de vaporização especificados [12]. A volatilidade é controlada pela *pressão de vapor* e pelo *intervalo de destilação* [7].

A estabilidade da gasolina é conseguida com a incorporação de aditivos que a tornam mais estável, promovem a limpeza e evitam a formação de depósitos nos motores. Caso contrário a gasolina poderia originar sujidade nos motores (formação de gomas) provocada pela sua deterioração por oxidação [12]. A estabilidade da gasolina é controlada por duas das características mencionadas na especificação: a *estabilidade à oxidação* e as *gomas existentes* [7].

O enxofre presente na gasolina pode provocar corrosão, e depois da combustão pode formar gases corrosivos que atacam o motor, o escape e o meio ambiente. A *corrosão à lâmina de cobre* avalia a actividade do enxofre e o *teor em enxofre* reporta o enxofre total presente na gasolina. Estas duas características são usadas para avaliar a qualidade do combustível relativamente à corrosão [7].

Os combustíveis queimam-se nos motores dos automóveis para deslocar os êmbolos, tal como descrito à frente no ponto 4.1. Uma medida com interesse é comparar o volume do cilindro após o êmbolo ter sido totalmente deslocado pelos gases em expansão, com o volume do cilindro quando o êmbolo está no seu ponto superior. Isto representa a taxa de compressão de cada motor. Um automóvel com taxa de compressão de 7,5 necessita, como modelo típico, de uma gasolina com número de octano de 88 para evitar o fenómeno de detonação, definido à frente no ponto 4.2. Aumentando a taxa de compressão para 9 passa a necessitar de um número de octano de cerca de 98. Este

aumento de número de octano de 10 tem a acompanhá-lo um aumento de cerca de 10% de rendimento, mas esta proporção de aumento de 1% de rendimento para cada 1% de aumento do número de octano não se mantém indefinidamente. Taxas de compressão cada vez maiores traduzem-se em diferenças de rendimento cada vez menores [12]. Assim quanto maior for a relação de compressão de um motor, maior será a economia de combustível e a potência desenvolvida. Esta relação, ou taxa de compressão, é contudo limitada pelo aparecimento da detonação. O poder anti-detonante de um combustível é expresso em termos do *número de octano*. Esta qualidade é, como a volatilidade, uma propriedade fundamental das gasolinas. Quanto maior for o número de octano de uma gasolina, melhor poderá resistir ao aparecimento da detonação e melhor será o rendimento do motor [12].

4. APLICAÇÃO DA GASOLINA NO MOTOR

4.1 FUNCIONAMENTO DO MOTOR A GASOLINA

Os motores dos veículos, cujo combustível usado é a gasolina, são motores de explosão ou combustão interna, e seguem um ciclo denominado por *Ciclo a Quatro Tempos* ou *Ciclo de Otto*. Os quatro tempos do ciclo designam-se por: *Admissão*, *Compressão*, *Explosão* e *Escape*, Figura 8 [13].

Para ter início o funcionamento do motor, a válvula de admissão abre-se enquanto o pistão se move para baixo. A gasolina é bombeada a partir do depósito e misturada com o ar no cilindro, esta fase é designada por *Admissão*. O pistão sobe e comprime a mistura ar-gasolina, na fase de *Compressão*, o que torna a explosão mais potente. Quando o pistão atinge o topo do seu curso, a vela de ignição produz uma faísca que inflama a gasolina. A gasolina no cilindro entra em combustão e expande-se empurrando o êmbolo ao longo do cilindro e gerando assim a potência necessária para criar o movimento. Logo, a primeira exigência de funcionamento para uma gasolina é que a sua combustão seja rápida e completa, no interior das câmaras de combustão, de modo a libertar o máximo de energia – *Explosão* [13].

De seguida o pistão atinge o topo do seu curso, a válvula de escape abre-se e os gases queimados saem do cilindro pelo tubo existente para esse fim – *Escape* [13].

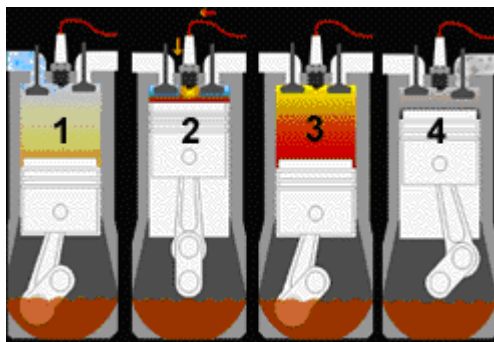


Figura 8 – Quatro tempos de um motor a gasolina: 1-Admissão, 2-Compressão, 3-Explosão e 4-Escape [14].

4.2 FENÓMENO DE DETONAÇÃO

Nos motores de combustão interna, são possíveis vários tipos de combustão. O processo normal é uma combustão, rápida e suave, da mistura combustível com o ar no motor, devido à propagação da chama proveniente da faísca entre os eléctrodos da vela, de acordo com o descrito no ponto anterior [13].

No entanto, quando bolsas com uma mistura de ar e combustível explodem espontaneamente na câmara antes da chama, ocorre um fenómeno indesejável chamado *detonação*. Como consequência há um aumento de pressão local seguido de vibrações, as quais são atenuadas apenas quando a pressão é igual em todos os pontos da câmara de combustão. Este fenómeno é acompanhado por um ruído característico e similar a uma série de estalares metálicos [15]. A Figura 9 mostra a câmara de combustão em condições normais e em condições de detonação, assim como os respectivos gráficos de pressão ao longo de um ciclo de funcionamento do motor.

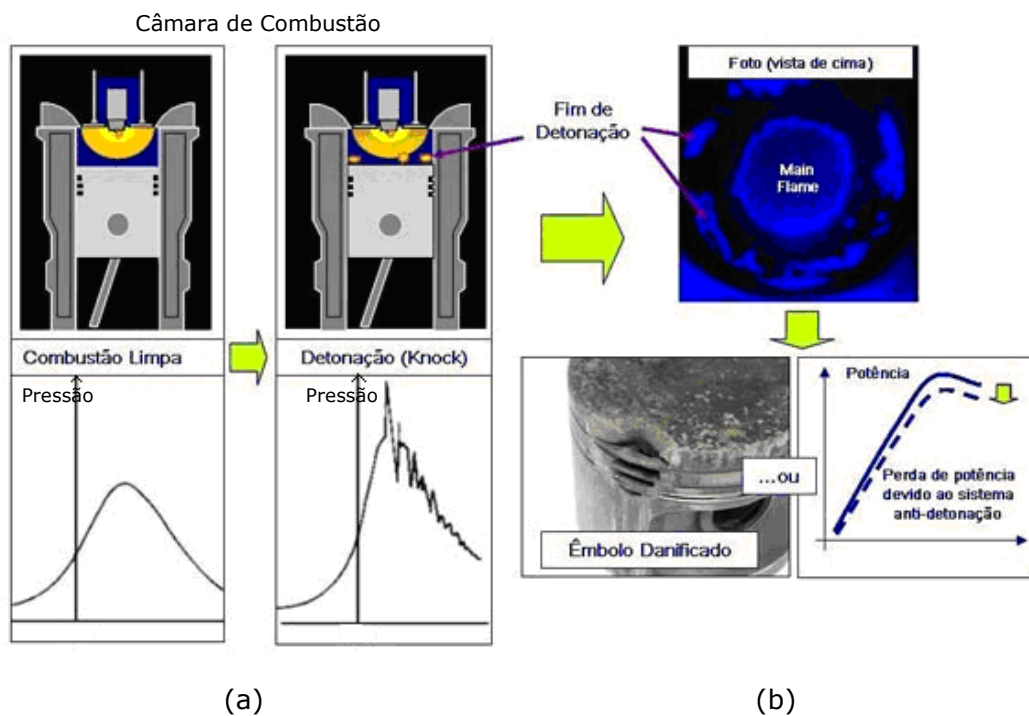


Figura 9 – (a) Diagrama de combustão em condições normais e de detonação.

(b) Consequências da detonação [15].

As detonações no motor levam a uma redução do desempenho, mas também podem danificar gravemente o motor. Podem causar sérios problemas

mecânicos, térmicos e incidentes destrutivos como a ruptura da cabeça do cilindro da junta, falha ou parcial fusão dos pistões, deterioração da cabeça dos cilindros e válvulas [13]. A [Figura 9](#) mostra um êmbolo danificado como consequência do fenômeno de detonação.

No caso dos motores mais recentes, o sistema de gestão do motor encarrega-se das detonações, dessincronizando o motor para evitar que isso ocorra. Obviamente isto significa que não ocorre detonação, no entanto o desempenho dos veículos reduz-se significativamente já que o motor está a proteger-se a si mesmo dos efeitos potencialmente nocivos da detonação, [Figura 9](#) [15].

4.3 DEFINIÇÃO DO NÚMERO DE OCTANO

Os pontos anteriores mostram como é importante que os combustíveis tenham grande resistência à auto-detonação. Qualitativamente são conhecidos os hidrocarbonetos com estruturas preferenciais nesta matéria: parafinas muito ramificadas, olefinas e aromáticos (benzeno, tolueno, xilenos) são muito resistentes à auto-detonação; inversamente as cadeias longas e sem ramificações de parafinas e olefinas, com mais do que quatro átomos de carbono, são mais susceptíveis aos fenômenos de auto-detonação [13].

De modo a caracterizar o comportamento dos combustíveis ou dos seus componentes relativamente à resistência à detonação durante a compressão e antes da ignição, sem envolver a composição química, foi introduzido e usado universalmente há mais de 50 anos o conceito de *Número de Octano* [13].

De acordo com este conceito, o combustível em teste é comparado com dois hidrocarbonetos puros escolhidos como combustíveis de referência primários. O primeiro é o 2,2,4-trimetilpentano ou isoctano, [Figura 2](#), o qual é muito resistente à auto-ignição e ao qual foi dado arbitrariamente o Número de Octano 100. O outro componente é o *n*-heptano, [Figura 2](#), o qual tem pouca resistência à detonação e ao qual foi atribuído o Número de Octano zero. Assim, um combustível com Número de Octano X , é uma mistura com X % v/v de isoctano e $(100-X)$ %v/v de *n*-heptano. As misturas do binário isoctano/*n*-heptano são consideradas combustíveis de referência [13].

A determinação do Número de Octano é feita recorrendo ao uso de um motor de referência designado por CFR (Cooperative Fuel Research) e existem dois processos padrão para a sua determinação, designados por RON (Research Octane Number) e MON (Motor Octane Number). A diferença entre os dois consiste essencialmente na velocidade de rotação, temperatura de admissão e avanço da faísca. O RON é mais usado do que o MON, e o valor obtido pelo MON é normalmente 10 a 12 pontos inferior ao RON. Esta diferença é designada por *sensibilidade* e é uma indicação da sensibilidade do combustível à modificação das condições experimentais, mais extremas (em particular ao aumento da temperatura) em que se processa o MON [13]. Convencionalmente as gasolinas têm um RON entre 90 e 100 e um MON entre 80 e 90 [13].

4.4 NÚMERO DE OCTANO vs DESEMPENHO DO MOTOR

Uma vez conhecido o Número de Octano do combustível, os motores devem ser construídos de modo a prevenir a detonação em quaisquer que sejam as condições de operação. As variáveis em causa são a razão de compressão e o avanço da ignição as quais determinam o desempenho do motor (eficiência térmica e potência). A potência pode sempre ser mantida por avanços tecnológicos como deslocamento de cilindros e razão de transmissão mas a eficiência térmica está sempre condicionada pelo Número de Octano. Assim quanto maior o Número de Octano, melhor será o seu desempenho e menor o consumo de combustível [13].

Verifica-se que a detonação depende de fenómenos físicos e químicos complexos e muito relacionados com o *design* do motor e as suas condições de operação. Assim não é possível caracterizar completamente o desempenho anti-detonante de uma gasolina com uma única medida, pois este está intimamente relacionado com o motor no qual é usada assim como com as condições operacionais do mesmo [10].

4.5 NÚMERO DE OCTANO DAS DIFERENTES FAMÍLIAS DE HIDROCARBONETOS

O RON dos hidrocarbonetos depende da sua estrutura química e varia com o ponto de ebulição de cada família, Figura 10. Para as *n*-parafinas, o RON é muito elevado para os componentes mais leves, e decresce com o aumento da cadeia alcançando o zero para o *n*-heptano. O RON, assim como o MON, aumenta sempre com o número e complexidade das ramificações laterais. O MON para parafinas é geralmente 2 a 3 pontos inferiores ao RON [13].

O RON para as olefinas, assim como para as parafinas, depende do comprimento e das ramificações da cadeia. O RON das olefinas é geralmente mais elevado do que o das parafinas para o mesmo esqueleto carbonado. A localização das ligações duplas no 'centro' da molécula tende a aumentar o RON, pelo menos para os primeiros membros da série. O MON das olefinas é sempre inferior ao correspondente RON numa média de 10 a 15 pontos, o que é explicado pela elevada energia de activação no processo cinético de auto-ignição. Na prática, o MON dos combustíveis comercializados correlaciona-se sempre fortemente com o conteúdo em olefinas [13].

Os compostos nafténicos têm sempre valores de RON e MON, superiores aos dos seus homólogos não-ciclícos. Por exemplo o RON do *n*-hexano é 24,8 enquanto o do ciclo-hexano é 83 [13].

O RON dos compostos Aromáticos é sempre superior a 100 (acima de 115-120). Estes valores estão numa ordem de magnitude em que há pouca precisão na determinação do RON. O MON dos Aromáticos também é superior a 100 mas sempre cerca de 10 pontos inferiores ao respectivo RON [13].

Para além dos hidrocarbonetos alguns compostos orgânicos oxigenados como álcoois e éteres são usados na formulação de gasolinas. São eles o metanol, etanol, propanol e butanol, assim como metil e etil éteres obtidos a partir de olefinas de C₄ e C₅: éter metil *terc*-butílico (MTBE), éter etil *terc*-butílico (ETBE) e o éter metil *terc*-amílico (TAME). Todos estes componentes são caracterizados por um elevado RON. O MON também é superior a 100 mas a diferença RON-MON é elevada, na ordem de 15 a 20 pontos. Os seus valores são geralmente estimados, uma vez que o procedimento do motor CFR não está adaptado a estas características. Estas substâncias têm um comportamento bastante diferente dos hidrocarbonetos derivados do petróleo.

Normalmente é mais importante conhecer o comportamento destes componentes na mistura do que no seu estado puro e o seu incremento nas gasolinas resulta sempre num ganho substancial no número de octano [13].

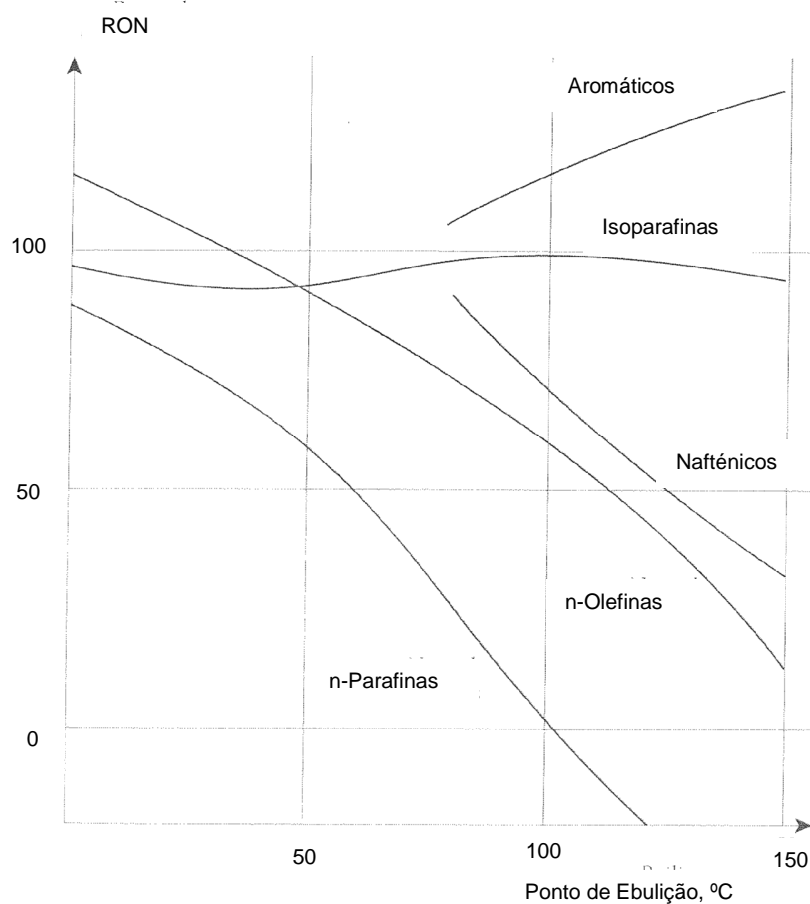


Figura 10 – Relação entre o RON e o ponto de ebulição das diferentes famílias de hidrocarbonetos [13].

5. MÉTODOS PARA A DETERMINAÇÃO DO NÚMERO DE OCTANO

5.1 MÉTODOS DE REFERÊNCIA- *ASTM D 2699* e *ASTM D 2700*

A razão anti-detonante da gasolina é medida num motor laboratorial de um único cilindro, de acordo com os métodos desenvolvidos para o efeito: *ASTM D 2699*- Caracterização do combustível pelo método *Research* (RON) e *ASTM D 2700*- Caracterização do combustível pelo método *Motor* (MON) [10].

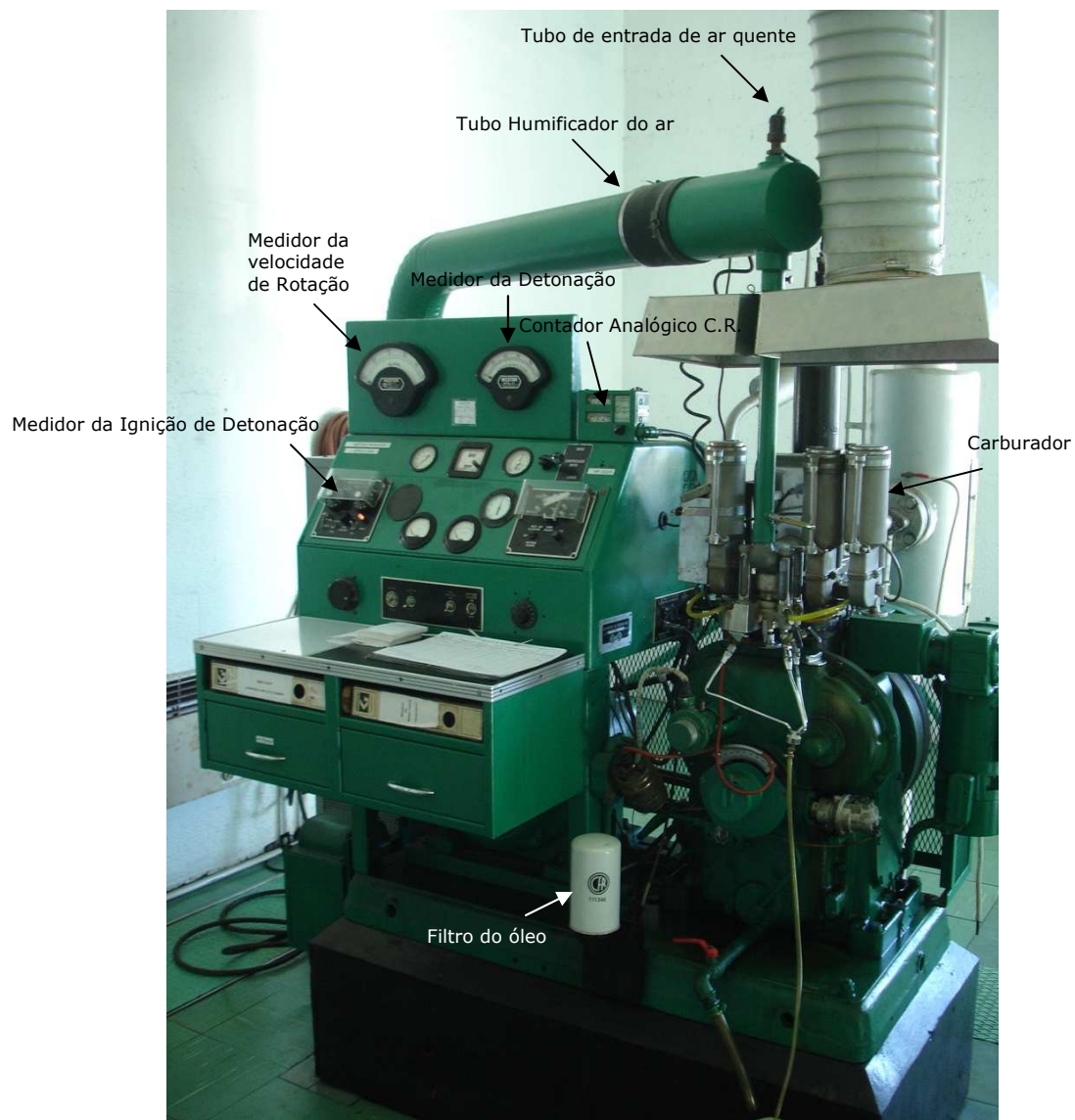


Figura 11 – Motor CFR do Laboratório da Refinaria do Porto.

O motor CFR (Cooperative Research Fuel), Figuras 11 e 12, funciona de acordo com as condições de operação definidas para cada método, usa um

único cilindro, um ciclo a quatro tempos e razão de compressão variável. A escala do número de octano é definida pela composição volumétrica das misturas Combustível de Referência Primárias (PRF). Estas misturas são preparadas a partir de isoctano e *n*-heptano, em proporções volumétricas de modo a obter misturas com números de octano de 0 a 100. A intensidade da detonação (KI) da amostra combustível é comparada com a de uma ou mais misturas PRF [16, 17].



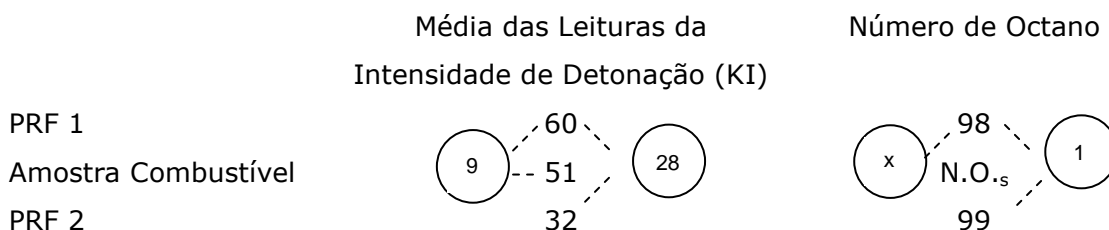
Figura 12 – Pormenores do Motor CFR: 1- Medidor da Velocidade de Rotação, 2- Medidor da Detonação, 3- Medidor da Ignição da Detonação, 4-Contador Analógico.

A escala do Número de Octano cobre a gama de 0 a 120 mas estes métodos têm uma gama de trabalho apenas de 40 a 120 [16,17].

A diferença entre os dois métodos consiste essencialmente na velocidade de rotação, temperatura de admissão do ar e da mistura. Durante a determinação do RON, o motor CFR opera a 600 ± 6 rpm, com temperatura de admissão do ar de $52 \pm 1^\circ\text{C}$ e sem pré-aquecimento do combustível. Em contraste o MON opera a 900 ± 9 rpm, com temperatura de admissão do ar de $38 \pm 2,8^\circ\text{C}$ e temperatura da mistura combustível de $149 \pm 1^\circ\text{C}$ [16,17].

O procedimento usado para a determinação do RON e MON é o da *Razão de Compressão*. Neste procedimento os valores de KI são estabelecidos a partir de PRF analisados com Números de Octano próximos da amostra a analisar e os vários parâmetros do motor são ajustados de acordo com as indicações do método em uso. A amostra a analisar é introduzida no carburador, o motor é colocado em funcionamento e o valor obtido no contador analógico é convertido em Número de Octano usando as '*Tabelas Guias de Intensidade de Detonação*' que constam nos métodos, para esse efeito [16,17].

Devem ser feitas duas determinações que não podem variar entre si mais do que 0,2, para ambos os métodos, e o Número de Octano é obtido por interpolação dessa média com as misturas padrão, como mostra o Esquema 1 [16,17].



$$N.O._s = N.O._{PRF1} + \left(\frac{K.I._{PRF1} - K.I._s}{K.I._{PRF1} - K.I._{PRF2}} \right) \cdot (N.O._{PRF2} - N.O._{PRF1}) = 98 + \left(\frac{60 - 51}{60 - 32} \right) \cdot (99 - 98) = 98.32$$

Esquema 1- Exemplo de cálculo da determinação do Número de Octano.

Em que:

Os valores dos círculos representam diferenças.

$N.O._s$ = Número de Octano da amostra combustível;

$N.O._{PRF1}$ = Número de Octano do PRF inferior;

$N.O._{PRF2}$ = Número de Octano do PRF superior;

$K.I._s$ = Intensidade de detonação da amostra combustível;

$K.I._{PRF1}$ = Intensidade de detonação do PRF inferior;

$K.I._{PRF2}$ = Intensidade de detonação do PRF superior.

A determinação do RON e MON usando o motor CFR é complexa e relativamente dispendiosa, uma vez que o motor requer manutenção constante e calibrações frequentes [16,17]. Nas gasolinas finais o número de

octano tem que ser determinado por estes métodos para cumprir as especificações e o produto poder ser colocado no mercado. No entanto para as correntes processuais e componentes de gasolina podem ser utilizados métodos alternativos mais rápidos e menos dispendiosos.

5.2 MÉTODOS ALTERNATIVOS

A complexidade da determinação do RON e MON pelo motor CFR, suscitou desde sempre interesse por parte de inúmeros investigadores, para o desenvolvimento de métodos alternativos à sua execução.

Uma vez que muitas das propriedades físicas e químicas das gasolinas estão directamente relacionadas com a sua composição, uma alternativa para a determinação do RON é o uso da composição da gasolina.

Existem várias técnicas analíticas para a determinação da composição das gasolinas, a partir da qual se pode calcular o RON. A desvantagem da maioria delas, dada a precisão necessária à obtenção do RON, é o facto da composição obtida ser pouco detalhada. Quanto mais detalhada for a composição obtida para a gasolina maior precisão é conseguida na previsão do RON [9].

Técnicas analíticas como a Adsorção com Indicador de Fluorescência, Infra-Vermelho e Ressonância Magnética Nuclear podem apenas fornecer informações limitadas sobre a composição correspondente a grupos estruturais como compostos aromáticos, olefinas e compostos saturados. Estas limitações analíticas na análise detalhada da composição fazem com que estes métodos não sejam tão precisos na previsão do número de octano, com a excepção dos casos em que a gasolina é constituída predominantemente por um único grupo estrutural [9].

Pelo contrário, os métodos que usam Cromatografia Gasosa (GC) com modernas colunas capilares, promovem uma excelente separação da maioria dos hidrocarbonetos constituintes da gasolina, e proporcionam a obtenção de dados de composição detalhados muito para além dos grupos estruturais, podendo por isso ser usados para a previsão mais precisa do número de octano [9].

Em 1955, *Schoen e Mrstik* [18] desenvolveram uma correlação gráfica para prever o número de octano para misturas binárias. Esta correlação tinha por

base o número de octano e a concentração em olefinas de cada um dos dois componentes que se pretendiam misturar. Este método foi redefinido, em 1959 por *Stewart* [19], e aplicado a misturas multi-componentes, com resultados mais consistentes.

Healy, et al. [20], em 1959, correlacionaram misturas de componentes de gasolina com diferentes níveis do número de octano e diferentes tipos de hidrocarbonetos em cada um dos componentes de gasolina. No entanto se este modelo for usado para prever o comportamento da mistura de novos componentes, podem ser obtidos valores pouco razoáveis, especialmente se o tipo de hidrocarbonetos ou o número de octano dos novos componentes estiver fora da gama dos componentes previamente testados.

Auckland e Charnock [21] desenvolveram, em 1969, um índice para o RON em misturas lineares obtido por combinação linear das propriedades molares parciais dos componentes que as constituem. Este índice de mistura é um valor hipotético obtido por extrapolação e a obtenção deste valor é análoga à obtenção da propriedade molar de uma solução. Desde que as propriedades molares parciais dependam da composição, é possível determinar o índice de mistura. No entanto esta aproximação é limitada, este método só pode ser usado para encontrar valores de componentes de mistura com uma composição particular e não pode ser usado para prever os seus valores de mistura em outras misturas [21].

Três anos mais tarde, em 1972, *Anderson et al.* [22], agruparam os compostos obtidos por cromatografia, na caracterização de gasolinas e componentes de gasolina, em 31 grupos, Tabela 4. Alguns desses grupos, consistiam apenas num componente enquanto outros eram misturas de componentes, agrupados pelo facto de terem estruturas químicas idênticas e por pertencerem a uma estreita gama de pontos de ebulição. Definiram um modelo para o cálculo do RON a partir da definição do RON numa mistura de *n*-heptano com o isoctano, eq.1 [22].

$$\text{RON} = a_1x_1 + a_2x_2 \quad (\text{eq.1})$$

Em que x_1 e x_2 são as fracções em volume do *n*-heptano e do isoctano na mistura e a_1 e a_2 são os valores do RON do *n*-heptano (RON=0) e do isoctano

(RON=100). E generalizaram esta equação para os restantes componentes da mistura, eq.2 [22].

$$RON = \sum_{g=1}^{g=31} b_g \cdot W_g \quad (\text{eq.2})$$

Em que:

W_g é a fracção em massa do grupo g na mistura.

b_g é o número de octano do grupo g na mistura.

No entanto, em misturas com outros componentes para além do n -heptano e do isoctano o RON efectivo dos compostos é função da sua concentração e da composição da mistura. O que significa que para se obter um RON preciso da mistura deve-se ter em conta os termos de interacção e os termos não-lineares. Porém devido à possível complexidade desses termos de interacção foram produzidos apenas modelos lineares simples. O que diferia entre cada um dos quatro modelos era a forma como os valores de b_g eram obtidos [22]:

➤ *Modelo 1 – RON dos componentes puros publicados*

O projecto API 45 [23] publicou tabelas com os números de octano para uma vasta gama de hidrocarbonetos. Estes foram usados para calcular os valores de b_g da eq.2. Nos grupos constituídos apenas por um componente o RON é igual ao apresentado pelo projecto API. Nos restantes casos os valores de RON foram estimados tendo em conta o RON dos componentes puros e a sua proporção na mistura [22].

➤ *Modelo 2 – RON das misturas publicados*

O projecto API 45 [23] também publicou o RON medido em misturas para uma elevada gama de hidrocarbonetos. Estes valores de RON foram calculados em misturas com 48%v/v de isoctano, 32%v/v de n -heptano e 20%v/v do hidrocarboneto em questão. Foi calculado o RON para cada grupo do mesmo modo mas usando agora estes novos valores de b_g [22].

➤ *Modelo 3 e 4 – RON efectivo estimado por regressão linear*

O número de octano da mistura, b_g , foi estimado por regressão linear múltipla a partir dos dados provenientes de 460 amostras, representativas de uma vasta gama de gasolinas finais e componentes de gasolina, Tabela 4 [22].

Os quatro modelos foram testados em 88 amostras cobrindo toda a gama de componentes de gasolina assim como de gasolinas finais.

Tabela 4 – Grupos usados para caracterizar as gasolinas (de acordo com a ordem de eluição) e coeficientes para estimar o RON das gasolinas [22].

Grupo	Componentes de cada grupo	Coefficiente de Regressão, b_g
1	componentes antes do <i>n</i> -butano	103,9
2	<i>n</i> -butano	88,1
3	<i>n</i> -butano a <i>i</i> -pentano (excluindo)	144,3
4	<i>i</i> -pentano	84,0
5	<i>i</i> -pentano a <i>n</i> -pentano (excluindo)	198,2
6	<i>n</i> -pentano	67,9
7	<i>n</i> -pentano a 2-metilpentano (excluindo)	95,2
8	2-metilpentano a 3-metilpentano (incluindo)	86,6
9	3-metilpentano a <i>n</i> -hexano (excluindo)	95,9
10	<i>n</i> -hexano	20,9
11	<i>n</i> -hexano a benzeno (excluindo)	94,9
12	benzeno	105,2
13	benzeno a 2-metil-hexano (excluindo)	113,6
14	2-metil-hexano a 3-metil-hexano (incluindo)	80,0
15	3-metil-hexano a <i>n</i> -heptano (excluindo)	97,8
16	<i>n</i> -heptano	-47,8
17	<i>n</i> -heptano a tolueno (excluindo)	62,3
18	tolueno	113,9
19	tolueno a 2-metilpentano (excluindo)	115,1
20	2-metil-heptano a 3-metil-heptano (incluindo)	81,7
21	3-metil-heptano a <i>n</i> -octano (excluindo)	109,7
22	<i>n</i> -octano	10,5
23	<i>n</i> -octano a etilbenzeno (excluindo)	96,1
24	etilbenzeno	122,6
25	etilbenzeno a <i>p</i> -xileno (excluindo)	45,4
26	<i>p</i> -xileno e <i>m</i> -xileno	102,0
27	<i>m</i> -xileno a <i>o</i> -xileno (excluindo)	73,3
28	<i>o</i> -xileno	123,6
29	fim <i>o</i> -xileno a fim do <i>n</i> -nonano	35,0
30	<i>n</i> -nonano a <i>n</i> -decano (excluindo)	112,0
31	<i>n</i> -decano e seguintes	85,6

Verifica-se que os Modelos 1 e 2 mostram erros sistemáticos, indicando que o RON efectivo é geralmente superior ao RON reportado pelo projecto API. Como era de esperar o RON das misturas apresentado no projecto API é

melhor do que o apresentado para os compostos puros pelo mesmo projecto. Em média o Modelo 1 apresenta resultados que são 10% abaixo do valor verdadeiro do RON, enquanto o Modelo 2 apresenta resultados apenas 2 a 3 % inferiores a este. Os Modelos 3 e 4 cujos coeficientes foram determinados por regressão linear são ambos melhores do que os Modelos 1 e 2 [22].

Verifica-se que equações lineares podem ser desenvolvidas e usadas no cálculo do RON. Este pode ser estimado a partir do RON dos compostos, do RON das misturas ou por regressão linear múltipla. O último método dá-nos valores mais realistas que embora não sejam suficientemente precisos para o controlo de qualidade das gasolinas finais, são muito úteis para cálculos processuais, controlo do processo de produção de componentes de mistura, entre outros [22].

Uma outra equação foi proposta por *Morris, et al.* [24] em 1975, que descreve o comportamento não linear, no que se refere às variáveis, em misturas de gasolinas.

$$\text{Número de Octano} = x_1a_1 + x_2a_2 + b_{12}x_1x_2 \quad (\text{eq.3})$$

Em que a_i é o número de octano do componente i , x_i a fracção em volume do componente i e b_{12} é o coeficiente de interacção dos componentes 1 e 2 [24].

No entanto existem algumas desvantagens em usar este modelo para descrever o comportamento da mistura de gasolina. Em primeiro lugar o parâmetro b_{12} é uma constante empírica que depende das características dos componentes 1 e 2, não depende só do tipo de componentes mas também do número de octano e da diferença entre o valor de octano de cada componente. Os valores de b_{12} podem variar entre o muito negativo e o muito positivo e podem alterar-se se as características dos componentes se alterarem. Os parâmetros binários obtidos pela equação só podem ser usados para os componentes a partir dos quais estes foram encontrados. É muito difícil generalizar estes parâmetros de interacção como função dessas propriedades. A segunda desvantagem é que quando novos componentes são adicionados a uma gasolina, novos parâmetros de interacção são necessários para descrever o comportamento do novo componente com os componentes presentes. Assim, é necessário um novo estudo de mistura com o novo componente de modo a determinar o termo de interacção adicional. Com n componentes, são

necessários $n(n-1)/2$ parâmetros de interação. Se quatro novos componentes forem adicionados a um mistura com 8 componentes o número de parâmetros de interação aumenta drasticamente de 28 para 66 [24].

Rusin, et al [25], em 1981 apresentaram um método de transformação complexo para estimar o RON para misturas de gasolinas a partir do RON dos seus componentes. Este método consiste em três passos:

- Transformação das propriedades dos componentes.
- Mistura linear dessas propriedades transformadas.
- Transformação inversa dos resultados.

Este método é similar ao do índice de mistura, no entanto pode causar inconsistência na transformação dos dados entre estes três passos [25].

Duran et al. [13], em 1987 desenvolveram uma técnica, por cromatografia gasosa, que lhes permitiu identificar e analisar cerca de 200 componentes de gasolina cujos números de octano eram conhecidos. Neste caso a primeira aproximação foi desenvolver um modelo linear, eq. 4 [13]:

$$RON = \sum (RON_{ppi} + K_i) \cdot C_i \quad (\text{eq. 4})$$

Em que RON é o número de octano da gasolina, RON_{ppi} o número de octano do componente i puro, C_i a concentração em % m/m e K_i o coeficiente que representa o desvio à idealidade. O valor de K_i é idêntico para componentes da mesma família de hidrocarbonetos com o mesmo grau de ramificação e com o mesmo número de átomos de carbono [13].

Zahed, et al. [26] propuseram em 1993 um modelo com 5 variáveis independentes para prever o número de octano. No entanto verifica-se que o modelo funcionava bem para as misturas usadas na regressão, mas tinha uma prestação pobre para outras misturas. As 5 variáveis derivadas deste modelo eram internamente inconsistentes. Os números de octano previstos a partir desta equação não eram muito próximos dos números de octano originais dos componentes da mistura. No caso do n -heptano o valor previsto a partir deste modelo era de 108,77 em oposição ao valor predefinido de zero. Similarmente o valor previsto para o isoctano por este modelo era de 108,95 versus o definido teoricamente de 100. O conjunto de variáveis obtidas por este modelo

pode apenas ser usado para um conjunto específico de componentes, na mesma gama de composições a partir das quais as variáveis foram definidas. Assim as vantagens do uso deste método são limitadas [26].

Vera e Luiz [27], desenvolveram um modelo para a determinação do RON e MON por regressão linear, a partir da análise cromatográfica (PIANO), usando o *Statgraphics* como ferramenta para o ajuste dos modelos propostos. Da análise cromatográfica foram detectados cerca de 300 componentes, o que tornava inviável o tratamento de todas as variáveis. Assim procederam à selecção das variáveis mais promissoras. A escolha destas variáveis foi feita com base na natureza química dos diferentes componentes e tendo em conta considerações teóricas para o seu agrupamento. O ponto de partida foi um modelo em que o número de octano da mistura era considerado o somatório dos números de octano dos diversos componentes. Devido ao grande número de componentes presentes na análise cromatográfica e tendo em conta as interacções verificou-se que é enorme o número de modelos que podem ser sugeridos [27]. Neste trabalho foram usadas 100 amostras de gasolinas provenientes de pelo menos 5 refinarias da *Petrobras* oriundas de processos de produção diferentes. Os dados foram avaliados por meio de histogramas de frequência das variáveis explicativas. Foram levados em conta os gráficos-caixa, para os valores de RON referentes a diversos hidrocarbonetos, usando-se os valores recomendados pelo *Manual de Dados Técnicos da Petrobras, Rio de Janeiro*. A divisão destes componentes em grupos foi feita de acordo com a sua natureza química: Parafinas (P), Olefinas (O), Nafténicos (N) e Aromáticos (A).

O uso de gráficos-caixa levou à escolha das variáveis explicativas do modelo, e a sua manipulação fez com que algumas delas fossem posteriormente desmembradas ou agrupadas até à obtenção do modelo considerado mais adequado (sete conjuntos diferentes de variáveis explicativas). Neste caso foram também utilizados os métodos de *stepwise* e de regressão múltipla para as tentativas de obtenção de regressão da variável RON em função dos componentes da análise cromatográfica. Foi sempre analisada a qualidade do ajuste, os gráficos de resíduos e a coerência dos valores dos coeficientes obtidos, obtendo o seguinte modelo [27]:

$$\text{RON} = 98,43 + 0,10 \cdot \text{AROM} - 0,18 \cdot \text{NAFRAR1} - 1,35 \cdot \text{NC7ENC8M} \quad (\text{eq.5})$$

Em que são usadas como variáveis explicativas para o melhor modelo: AROM, NAFFAR1 e NC7ENC8M [27].

Os resultados obtidos por este método foram comparados com os obtidos pelo método ASTM D2699 por meio do teste da média das diferenças. Os testes mostraram que os resultados obtidos pelo método cromatográfico podem ser considerados iguais aos obtidos pelo motor. Foram obtidas regressões consideradas adequadas, ou seja, foram criados modelos cujos gráficos dos resíduos não apresentavam qualquer tipo de estrutura, embora alguns dos valores absolutos fossem maiores do que a reprodutibilidade do método de obtenção do RON [27].

Recentemente, e tendo como base os contínuos avanços dos computadores, foram desenvolvidos alguns novos e promissores métodos não lineares, baseados em patentes reconhecidas, para a previsão do RON [9]. O novo método proposto por *N. Nikolaou et al*, [9] é usado para a previsão do RON em gasolinas de isomerização, eq. 6:

$$\text{RON}_G = \sum_{i=1}^N K_i \text{RON}_i y_i \quad (\text{eq.6})$$

Em que y_i é a fracção em volume do hidrocarboneto presente na gasolina, e obtido por GC, o RON_i é o RON do hidrocarboneto puro (publicado), Tabela 5, e K_i é o factor de peso que é calculado por:

$$K_i = \frac{\text{RON}_i}{\text{BRON}_i} \cdot \frac{\sum_{i=1}^N y_i \text{BRON}_i}{\sum_{i=1}^N y_i \text{RON}_i} \quad (\text{eq.7})$$

Em que BRON_i é o RON do hidrocarboneto na mistura (publicado), Tabela 5. O produto $K_i \text{RON}_i$ dá-nos um RON calculado efectivo ou da mistura, Tabela 5, característico para uma determinada gasolina, o qual também depende, do número, tipo e concentração dos hidrocarbonetos presentes na gasolina especificada [9].

Foram consideradas 20 amostras de produtos puros de isomerização de refinarias com diferente número de componentes e concentração de cada um. Foi determinado o RON pelo CFR e as mesmas amostras foram analisadas por

cromatografia. Foi detectado um número máximo de 26 hidrocarbonetos, e mais de 98% m/m da composição da gasolina foi determinada quantitativamente [9].

Tabela 5 – Números de Octano dos vários hidrocarbonetos presentes em gasolinas de isomerização [9].

Hidrocarbonetos	y_i	K_i	RON	RON	RON
			puro	mistura	calculado
<i>n</i> -butano	0,0043	0,876	93,8	113,0	82,1
2,2-dimetilpropano	0,0003	0,902	85,5	100,0	77,1
2-metilbutano	0,4507	0,974	92,3	100,0	89,9
<i>n</i> -pentano	0,1335	1,05	61,7	62,0	64,8
2,2-dimetilbutano	0,2342	1,088	91,8	89,0	99,9
ciclo-pentano	0,0116	0,758	101,3	141,0	76,8
2,3-dimetilbutano	0,0270	1,137	103,5	96,0	117,7
2-metilpentano	0,0490	0,944	73,4	82,0	69,3
3-metilpentano	0,0200	0,914	74,5	86,0	68,1
<i>n</i> -hexano	0,0260	1,377	24,8	19,0	34,1
2,2-dimetilpentano	0,0004	1,100	92,8	89,0	102,1
metil ciclo-pentano	0,0113	0,900	91,3	107,0	82,2
2,4-dimetilpentano	0,0003	1,153	83,1	76,0	95,8
2,2,3-trimetilbutano	0,0001	1,056	112,1	112,0	118,4
3,3-dimetilpentano	0,0002	1,015	80,8	84,0	82,0
ciclo-hexano	0,0249	0,796	83	110,0	66,1
2-metil-hexano	0,0006	1,118	42,4	40,0	47,4
2,3-dimetilpentano	0,0003	1,102	91,9	88,0	101,2
1,1-dimetil ciclo-pentano	0,0002	1,014	92,3	96,0	93,6
3-metil-hexano	0,0005	0,979	52	56,0	50,9
1,3 dimetil- <i>cis</i> -ciclopentano	0,0002	0,852	79,2	98,0	67,5
1,3 dimetil- <i>trans</i> -ciclopentano	0,0002	0,945	80,6	90,0	76,1
3-etilpentano	0,0002	1,071	65	69,6	69,6
<i>n</i> -heptano	0,0002		0	0	0
metil ciclo-hexano	0,0039	0,759	74,8	104	56,7
etil ciclo-hexano	0,0001	1,158	67,2	61,2	77,8

Verifica-se que os valores encontrados para o RON estão dentro do intervalo de $\pm 0,5$ RON do RON(ASTM) medido. Verifica-se que o RON(ASTM) do motor CFR é na maioria dos casos superior ao obtido a partir dos dados do GC e esta diferença pode dever-se a perda de informação do GC, uma vez que a composição determinada quantitativamente é inferior a 100% m/m. A desvantagem é que o RON da mistura calculado por este método tem uma correlação negativa com o RON das misturas apresentadas no projecto API 45 e ambos são comparados com os valores puros de RON. No entanto este é o efeito dos factores de peso K_i , que dependem não só da composição específica das gasolinas mas também da mistura dos RON's [9].

Das várias metodologias apresentadas são evidentes duas grandes desvantagens. A primeira é que embora os métodos baseados em análises de GC nos possam dar dados de composição detalhados, o agrupamento dos hidrocarbonetos analisados em grupos com estruturas similares, para a previsão do número de octano, automaticamente remove a maior vantagem oferecida por esses métodos. Este agrupamento é sempre baseado no incerto pressuposto de que os grupos estruturais se comportam do mesmo modo durante a combustão de diferentes gasolinas. A outra desvantagem dos métodos propostos é o facto dos valores de RON dos componentes puros e das misturas terem sido obtidos sob condições específicas e serem aplicados de igual modo para todos os tipos de misturas de gasolinas. Por exemplo, os números de octano determinados no projecto API 45, no caso das misturas, foram estimados a partir da análise de misturas de 48% v/v isoctano, 32% v/v *n*-heptano e 20% v/v do hidrocarboneto específico. Deve ser reconhecido que esse número de octano nos dá alguma informação qualitativa em termos do desempenho do RON, relacionado com esse hidrocarboneto específico nessa mistura. No entanto a sua extensão para qualquer número, tipo e concentração de hidrocarbonetos presentes na gasolina pode ser questionável.

5.3 APLICAÇÃO DOS MÉTODOS ALTERNATIVOS AO REFORMADO

Dos modelos apresentados nos vários trabalhos publicados, verifica-se que em alguns deles não são mencionados todos os dados necessários à sua aplicação em outras amostras. Outros por sua vez, embora apresentem os dados

necessários e bons resultados, nas amostras a partir das quais foram desenvolvidos, quando aplicados em novas amostras, já não apresentam resultados satisfatórios [20].

No trabalho publicado por *Anderson et al.* [22] os vários compostos detectados na análise cromatográfica foram agrupados em 31 grupos, de acordo com as suas estruturas químicas e ordem de eluição. Verifica-se que apresentam uma separação muito diferente da obtida com o equipamento disponível para o desenvolvimento da presente dissertação. Para além disso verifica-se que as matrizes de gasolina usadas são diferentes da matriz de Reformado. As gasolinas são constituídas apenas por *i-parafinas*, *n-parafinas* e compostos *aromáticos*, como consequência alguns dos compostos que constituem o Reformado não são apresentados no trabalho publicado. É o caso das *olefinas* e dos compostos *nafténicos*, que representam cerca de 2%*m/m* da composição das amostras de Reformado. Deste modo os modelos desenvolvidos não são aplicáveis na previsão do RON das amostras de reformado, objecto de estudo desta dissertação.

Vera e Luiz [22], desenvolveram o seu trabalho em amostras de gasolinas e classificaram inicialmente os compostos em grupos, de acordo com a sua natureza química: *Parafinas (P)*, *Olefinas (O)*, *Nafténicos (N)* e *Aromáticos (A)*. No entanto no processo de escolha das variáveis explicativas e sua manipulação, estas foram desmembradas ou agrupadas de um modo não muito explícito. No final as variáveis explicativas para o melhor modelo são: *AROM*, *NAFPAR1* e *NC7ENC8*. Tendo por base a descrição no trabalho publicado não é possível rastrear quais os componentes que fazem parte de cada uma das três variáveis explicativas, não sendo por isso possível aplicar o modelo proposto a outras amostras.

No caso do trabalho de *N. Nikolaou et al* [9], desenvolvido em amostras de gasolinas de isomerização, verifica-se que os componentes que constituem essas gasolinas são diferentes dos que constituem as amostras de Reformado. Estas gasolinas não têm compostos Aromáticos que representam em média 67%*m/m* da composição das amostras de Reformado. Por este motivo não é também possível aplicar o modelo desenvolvido neste trabalho às amostras de Reformado em estudo.

Os restantes trabalhos publicados, apresentam essencialmente os resultados finais e o modo de proceder sem que apresentem dados suficientes que permitam a aplicação dos modelos em outras amostras, nomeadamente nas amostras de Reformado. Assim, o ponto seguinte descreve o método proposto no presente trabalho para determinar o RON em amostras de Reformado, a partir da composição determinada por cromatografia.

6. METODOLOGIA PROPOSTA NESTE TRABALHO PARA A DETERMINAÇÃO DO RON

Embora o GC ofereça informação detalhada sobre a composição da gasolina, o problema mais comum para o cálculo preciso do número de octano é a contribuição não-linear das características dos componentes individuais da amostra, a qual é normalmente muito complexa. Assim, o número de octano determinado para cada componente não pode simplesmente ser combinado linearmente, de acordo com a composição da gasolina, para o cálculo do número de octano da mistura. A combinação não-linear dos componentes puros é o mais complexo para o cálculo preciso do número de octano.

O que se propõe neste trabalho é o desenvolvimento de um modelo de regressão linear múltipla para calcular os valores de RON em amostras de **Reformado**, a partir dos resultados da análise cromatográfica das mesmas. Este modelo não usa os valores de RON publicados (para hidrocarbonetos puros ou em misturas) para deste modo evitar o efeito da contribuição não-linear das características dos hidrocarbonetos individuais na mistura. Uma vez que em todas as amostras de Reformado nas quais é determinado o número de octano (RON) também é determinada a composição por cromatografia, esta nova metodologia apresenta desde já a vantagem de reduzir o número de análises necessárias a uma única - a análise cromatográfica. A esta acrescenta-se o facto da análise cromatográfica ser mais rápida, mais económica e ambientalmente preferível devido ao reduzido volume de amostra a manipular.

Esta nova metodologia poderá ser extendida a outras correntes processuais ou componentes de gasolina, se se mostrar vantajosa.

6.1 AMOSTRAS

As amostras estudadas neste trabalho, num total de 375, são de *Reformado* proveniente da Fábrica de Combustíveis da Refinaria do Porto da Galp Energia. Foram recolhidas no período compreendido entre 2003 e 2006, três vezes por semana. Todos os dados referentes a estas amostras são apresentados no Ficheiro 4: anexos, do CD-ROM anexo à dissertação. O critério de selecção das

amostras, durante este período de tempo, foi a análise simultânea da composição por cromatografia e do número de octano pelo motor CFR (método ASTM D 2699). Todo o trabalho laboratorial foi realizado em rotina, pelos analistas do laboratório, e fora do contexto desta pós-graduação.

6.2 DADOS DO MOTOR CFR

Todas as amostras consideradas neste trabalho foram analisadas de acordo com o método ASTM D 2699 – Determinação do Número de Octano (RON). Para tal foi usado um motor CFR (Cooperative Research Fuel), Figura 11, a 600 ± 6 rpm, temperatura de admissão do ar de $52 \pm 1^\circ\text{C}$, sem pré-aquecimento do combustível e usando o procedimento da *Razão de Compressão*. Para cada amostra foram feitas duas determinações que não variaram entre si mais do que 0,2 e o RON foi obtido por interpolação dessa média com as misturas padrão, como mostra o exemplo do Esquema 1 [16,17].

6.3 DADOS CROMATOGRÁFICOS

A composição das amostras de Reformado foi determinada por cromatografia gasosa multidimensional. O laboratório dispõe de um cromatógrafo, Hewlett-Packard 5890 com Tecnologia AC PIONA, que separa a amostra nas suas 5 famílias de compostos (*n*-parafinas, *i*-parafinas, olefinas, compostos nafténicos e compostos aromáticos), por número de carbonos, Figura 13.

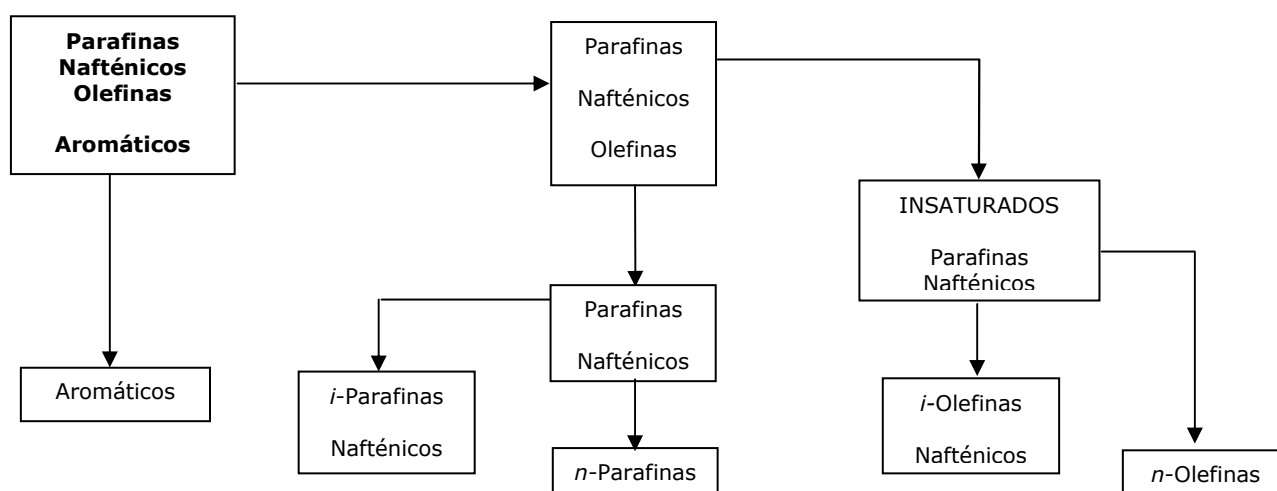


Figura 13 – Diagrama de separação do Cromatógrafo PIONA [28].

Este sistema é constituído por um conjunto de válvulas e colunas que permitem a separação desejada, sendo que cada coluna, Tabela 6, tem um propósito específico.

Tabela 6 – Descrição de cada uma das colunas do cromatógrafo PIONA [28].

Coluna	Tipo
OV275	Coluna de enchimento com 3,0mx2,0mm (d.i.), com 30% OV275 em <i>Chromosorb PAW 60-80 mesh</i>
HP-1	Coluna capilar, 15mx0,53mm (d.i.) com 5mm HP-1 (<i>Crosslinked MetilSiloxano</i>)
Trap Tenax	Coluna de enchimento, 0,15mx2,5mm (d.i.)
Trap de Olefinas	Coluna de enchimento, 0,30mx2,5mm (d.i.)
5A <i>Molecular Sieve</i>	Coluna de enchimento, 0,10mx2,5mm (d.i.)
13X <i>Molecular Sieve</i>	Coluna de enchimento, 1,80mx1,7mm (d.i.)
Pt	Catalizador Hidrogenador de Platina

A *Coluna OV-275* é usada como uma pré-coluna de separação, separa a amostra em três fracções as quais são analisadas posteriormente noutras colunas do sistema. Esta coluna retarda especialmente a eluição dos compostos aromáticos: o benzeno (ponto de ebulição de 80°C) elui depois do dodecano (ponto de ebulição 200°C). Isto significa que a fracção da amostra que elui antes do benzeno contém parafinas e compostos nafténicos com pontos de ebulição até 200°C. Estes componentes são separados sequencialmente no cromatógrafo pela *Trap de Olefinas*, 5A e 13X. O eluente proveniente da coluna OV-275 com início no benzeno é separado em duas fracções. Estas fracções são transferidas para a coluna HP-1 por via da *Trap Tenax* [28].

A *Trap Tenax* absorve, focaliza e transfere as fracções de aromáticos da coluna OV-275 para a HP-1. Esta coluna absorve essa fracção à temperatura de 70°C e desorve para a coluna HP-1 à temperatura de 280°C [28].

A *Coluna HP-1* é usada para separar os aromáticos com baixos pontos de ebulição das parafinas e compostos nafténicos com elevado ponto de ebulição. Devido ao número limitado de compostos aromáticos com ponto de ebulição abaixo dos 200°C a separação pode ser feita por diferença dos números de

carbono dos compostos aromáticos. Os componentes com ponto de ebulição abaixo dos 200°C são enviados de volta, da coluna para o detector [28].

A *Trap de Olefinas* é usada para separar as olefinas dos compostos saturados. A fracção de hidrocarbonetos com ponto de ebulição até 200°C (C₄ a C₁₁) elui da coluna OV-275 para a *Trap de Olefinas*. As olefinas são retidas enquanto os compostos saturados passam pela coluna; de seguida, e por aquecimento, a *Trap* liberta as olefinas. A sua selectividade depende da concentração e distribuição dos componentes olefínicos. Para amostras que contenham mais de 15% de olefinas a selectividade é superior ao número de carbono sete. O resultado é a separação entre saturados C₁₁ e olefinas C₄. A *Trap* tem uma temperatura de adsorção entre 90-160°C e de dessorção de 280°C [28].

A *Coluna do Catalisador Hidrogenado de Platina* hidrogena todos os componentes olefínicos à sua forma saturada antes de entrarem nas colunas 5A e 13X. Uma vez que estas colunas só actuam com compostos saturados, caso algum componente olefínico elua desta coluna irá resultar em picos fantasmas. A hidrogenação funciona a temperaturas entre os 180-220°C, geralmente é escolhida a temperatura de 220°C e reduzida a um mínimo de 180°C quando a actividade da coluna está muito elevada [28].

A *Coluna 5A Molecular Sieve* é usada para separar as *n*-parafinas das *i*-parafinas na gama do C₄ ao C₁₁. Funciona como uma *trap* reversível para as *n*-parafinas nesta gama de número de carbono. À temperatura de adsorção as *i*-parafinas e os compostos nafténicos passam pela coluna sem serem impedidos enquanto as *n*-parafinas são retidas e mais tarde são dessorvidas a uma temperatura mais elevada. A temperatura de adsorção varia entre 110-160°C e a de dessorção é de 450°C [28].

A *Coluna 13X Molecular Sieve* é usada para separar as parafinas dos compostos nafténicos por número de carbono, o ponto de ebulição dos compostos a serem analisados é de 270°C [28].

Depois de separados, os compostos são enviados para o detector de ionização de chama (FID) que os detecta e posteriormente quantifica, a partir de factores de resposta baseados na percentagem em massa de carbono. Considera-se que o metano tem um factor de resposta unitário e o factor de resposta de cada componente é calculado com base na massa atómica do carbono 12,011 e do hidrogénio 1,0080, assim:

$$\text{Factor Resposta} = \frac{0,7487x \frac{(C_{aw} \cdot C_n) + (H_{aw} \cdot H_n)}{C_n}}{C_{aw}} \quad (\text{eq.8})$$

Em que C_{aw} é a massa atómica do carbono, C_n é o número de átomos de carbono na molécula, H_{aw} é a massa atómica do hidrogénio, H_n é o número de átomos de hidrogénio na molécula e 0,7487 é a correcção do factor de resposta do metano para a unidade.

A Figura 14 mostra a separação obtida para uma das amostras de Reformado, usada para o modelo, analisada pelo cromatógrafo PIONA e, cujas concentrações estão descritas na Tabela 7.

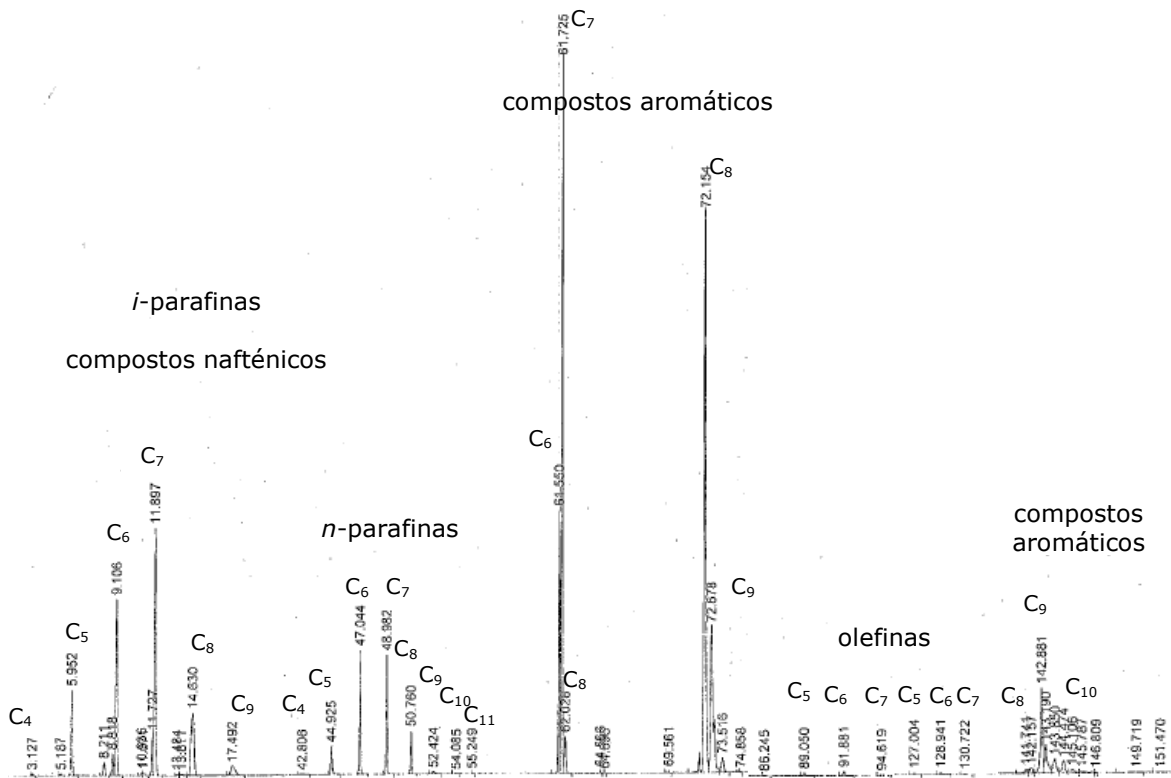


Figura 14 – Exemplo de um cromatograma de uma amostra de Reformado usada no modelo.

Tabela 7 – Exemplo da composição de uma amostra de Reformado usada no modelo.

Composto	%, m/m	Composto	%, m/m		
<i>i</i> -Parafinas	C4	0,00	C4	0,00	
	C5	2,52	C5	0,00	
	C6	8,33	C6	0,10	
	C7	8,57	C7	0,22	
	C8	2,61	C8	0,00	
	C9	0,11	C9	0,00	
	C10	0,08	C10	0,00	
	C11	0,03	Total	0,32	
Total	22,25	<i>n</i> -Olefinas	C4	0,00	
<i>n</i> -Parafinas	C4		0,00	C5	0,00
	C5		1,41	C6	0,03
	C6		3,44	C7	0,00
	C7	2,20	Total	0,03	
	C8	0,69	Nafténicos	C5	0,19
	C9	0,58		C6	0,55
	C10	0,00		C7	0,28
C11	0,00	C8		0,11	
Total	8,32	C9		0,00	
<i>i</i> -Olefinas	C4	0,00	C10	0,00	
	C5	0,00	Total	1,13	
	C6	0,21	Aromáticos	C6	5,69
	C7	0,35		C7	17,02
	C8	0,00		C8	23,08
	C9	0,00		C9	17,63
C10	0,00	C10		2,87	
Total	0,56	Total	66,29		

6.4 ANÁLISE ESTATÍSTICA DOS DADOS

6.4.1 Programa Informático para Estatística – SPSS

O SPSS (Statistical Package for the Social Sciences), é por tradição um programa usado para a análise estatística em ciências sociais [29]. No entanto este programa também é usado nas ciências exactas, como no caso da presente dissertação onde toda a análise estatística foi executada recorrendo ao SPSS.

Este programa faz a análise de dados em 4 etapas [29]:

- Introduzem-se as variáveis da amostra no editor de dados
- Seleccionam-se os procedimentos de análise de um dos menus
- Seleccionam-se as variáveis a analisar
- Examinam-se e interpretam-se os resultados

Na primeira etapa é necessário ter particular atenção ao tipo de variáveis em estudo (escala de medida). Na segunda etapa seleccionam-se os procedimentos estatísticos a utilizar. Depois de seleccionadas as variáveis e as opções dos diversos campos obtêm-se os resultados. Na interpretação dos resultados é necessário ter em conta que o programa não testa *à priori* os pressupostos que é necessário que sejam verificados para utilizar determinados testes [29].

6.4.2 O modelo de regressão linear [29]

A regressão linear é uma técnica estatística usada para modelar relações entre variáveis e prever a partir de um conjunto de variáveis independentes o valor da variável dependente. No modelo de regressão linear múltipla, a relação funcional entre uma variável dependente Y e as várias variáveis independentes X_i ($i=1, \dots, p$) é do tipo:

$$Y_j = \beta_0 + \beta_1 \cdot X_{1j} + \beta_2 \cdot X_{2j} + \dots + \beta_p \cdot X_{pj} + \varepsilon_j \quad (j=1, \dots, n) \quad (eq.9)$$

Em que:

β_i –são os coeficientes de regressão.

ε_j - erros aleatórios ou resíduos do modelo.

As hipóteses subjacentes a este modelo são:

- O termo ε_j reflecte apenas os erros da medição e a variação natural em Y_j . Os valores de X_i são encarados como constantes predeterminadas, sem erro.
- Os erros ε_j são aleatórios, independentes, têm valor esperado nulo, variância constante (σ^2) e são normalmente distribuídos, isto é, $\varepsilon_j \sim N(0, \sigma I)$.
- As variáveis independentes são ortogonais (não estão correlacionadas, ou quanto muito apresentam apenas correlações fracas).

6.4.3 Estimação dos coeficientes de regressão – método dos mínimos quadrados [29]

Os parâmetros $\beta_0, \beta_1, \dots, \beta_p$ do modelo podem ser calculados a partir de um conjunto de observações $(X_{1j}, X_{2j}, \dots, X_{pj}, Y_j)$ recorrendo ao método dos mínimos quadrados. Neste método, as estimativas dos coeficientes de regressão são obtidas de modo a que os erros ou resíduos do modelo de regressão linear, obtidos por $e_j = y_j - \hat{y}_j$ sejam mínimos. A partir do somatório do quadrado de e_j é possível avaliar as diferenças existentes entre os valores estimados pela recta (\hat{y}_j) e os valores observados (y_j). Assim a estimação dos parâmetros $\beta_0, \beta_1, \dots, \beta_p$ do modelo, resume-se à determinação do mínimo da função designada por Soma dos Quadrados dos Erros ou Resíduos do modelo (SQE), relativamente a cada um dos coeficientes de regressão, eq.10.

$$SQE = \sum_{j=1}^n e_j^2 \quad (\text{eq. 10})$$

O modelo em forma matricial é:

$$y = X\beta + \varepsilon \quad (\text{eq. 11})$$

em que:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad \text{é o vector das } n \text{ observações da variável dependente, (eq.12)}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{bmatrix} \quad \text{é o vector dos } p+1 \text{ coeficientes do modelo, (eq.13)}$$

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{12} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{pn} \end{bmatrix} \quad \text{é a matriz das } n \text{ observações de } X_i \text{ (} i=1, \dots, p \text{), (eq.14)}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \text{ é o vector dos erros com } \varepsilon \sim N(0, \sigma I), \text{ (eq.15)}$$

A minimização da soma dos quadrados dos erros é dada por $(y - X\beta)'(y - X\beta)$ e obtém-se para o vector $\hat{\beta}$ que satisfaça a condição:

$$\frac{\partial}{\partial \hat{\beta}} [(y - X\hat{\beta})'(y - X\hat{\beta})] = 0 \quad (\text{eq.16})$$

A solução desta equação diferencial, resolvendo em ordem a $\hat{\beta}$ é:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (\text{eq.17})$$

6.4.4 Inferências sobre o modelo de regressão linear

6.4.4.1 A análise de variância do modelo de regressão linear [29]

Depois de desenvolvida a regressão linear é necessário avaliar, a partir das estimativas amostrais, se alguma das variáveis independentes pode ou não influenciar a variável dependente, isto é, se o modelo ajustado é ou não significativo:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad \text{vs.}$$

$$H_1: \exists i: \beta_i \neq 0 \quad (i=1, \dots, p)$$

(H_1 lê-se como "existe pelo menos um i tal que β_i é diferente de zero")

Para o teste destas hipóteses, a variabilidade total em Y , que é expressa pela soma dos quadrados dos desvios de Y em relação a \bar{Y} (SQT), divide-se em duas componentes. Uma delas é explicada pelo modelo ajustado (SQR) e a outra não é explicada pelo modelo mas sim pelos erros (SQE), eq. 18.

$$\sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2 + \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 \quad (\text{eq. 18})$$

$$SQT = SQR + SQE$$

Se a variância total em Y , explicada pelo modelo de regressão for significativamente maior do que a variância total não explicada pelo modelo,

então o modelo ajustado é significativo. A variância explicada pela regressão é estimada dividindo a SQR pelos respectivos graus de liberdade p , e a variância dos erros pode ser estimada dividindo a SQE pelos respectivos graus de liberdade $(n-p-1)$:

$$F = \frac{SQR/p}{SQR/(n-p-1)} = \frac{QMR}{QME} \quad (\text{eq.19})$$

Em que QMR e QME são respectivamente os quadrados médios da regressão e dos erros. O SPSS cria um p -value para esta estatística de teste, em que se p -value $\leq \alpha$, rejeitamos a H_0 a favor da H_1 . Concluindo deste modo que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente, logo o modelo ajustado aos dados é significativo. É no entanto necessário analisar se todas ou apenas algumas variáveis independentes influenciam a variação da variável dependente.

6.4.4.2 Testes aos coeficientes do modelo de regressão [29]

Para saber quais dos β_i são diferentes de zero é necessário testar os β_i , de acordo com as seguintes hipóteses:

$$\begin{aligned} H_0: \beta_i &= k & \text{vs.} \\ H_1: \beta_i &\neq k & (i=1, \dots, p) \end{aligned}$$

Em que k pode tomar qualquer valor, no entanto na maioria dos *software* de análise estatística $k=0$. A estatística de teste é:

$$T_{\beta_i} = \frac{b_i - k}{\sqrt{s_{y,x}^2 \times C_{ii}}} \quad (\text{eq.20})$$

Em que b_i é a estimativa de β_i , $s_{y,x}^2 = QME$ é a estimativa da variância do modelo e C_{ii} é o elemento i da diagonal da matriz $(X'X)^{-1}$ correspondente a β_i . Para H_0 , T_{β} tem uma distribuição t -student com $(n-p-1)$ graus de liberdade. O SPSS produz um p -value e rejeita-se H_0 se p -value $\leq \alpha$.

6.4.4.3 O coeficiente de determinação

Existem vários parâmetros que ajudam a qualificar o modelo de regressão como: o coeficiente de correlação (r), coeficiente de determinação (R^2) e o coeficiente de determinação ajustado (R_a^2).

O coeficiente de correlação mede a associação linear entre duas variáveis como por exemplo os valores conhecidos e os respectivos valores previstos por um modelo de calibração (y_j, \hat{y}_j). E neste caso serve para avaliar a proximidade entre os valores observados e os valores estimados pelo modelo. Varia entre -1 e +1, indicando nestes extremos a existência de uma forte correlação entre as duas variáveis (correlação negativa e correlação positiva respectivamente). Quando não há uma relação linear entre as duas variáveis então o valor de r distancia-se de -1 e +1 e é próximo de 0 (zero) [29].

O coeficiente de determinação (geralmente representado por R^2) é uma das medidas da qualidade do ajuste mais usadas, eq. 21.

$$R^2 = \text{SQR} / \text{SQT} \quad (\text{eq.21})$$

O R^2 mede a proporção da variabilidade total que é explicada pela regressão ($0 \leq R^2 \leq 1$). Quando $R^2 = 0$ o modelo não se ajusta aos dados, pelo contrário quando $R^2 = 1$ o ajuste é perfeito. O valor de R^2 para o qual se considera um ajuste adequado é muito subjectivo; é geralmente aceite um bom ajuste, no caso das ciências exactas, se $R^2 > 0,9$. No entanto o uso do R^2 como indicador da qualidade do ajuste deve ser usado com cuidado e não deve ser usado para comparar modelos cujo número de variáveis independentes seja muito diferente entre eles. Na maioria dos casos, o aumento do número de variáveis independentes provoca um aumento no R^2 mesmo que estas tenham pouca influência sobre a variável dependente [29].

Já o coeficiente de determinação ajustado R_a^2 , eq.22, pode ser usado sem este tipo de restrições, como indicador de qualidade do ajuste [29]. A adição de uma nova variável no modelo só leva ao aumento do R_a^2 se esta conduzir a um melhor ajuste do modelo aos dados, isto é se a variância dos erros (QME) diminuir relativamente à variância total (QMT) [29].

$$R_a^2 = 1 - \frac{QME}{QMT} = 1 - \frac{n-1}{n-p-1}(1-R^2) = R^2 - \frac{p(1-R^2)}{n-p-1} \quad (\text{eq. 22})$$

6.4.5 Validação dos pressupostos do modelo de regressão linear

O modelo de regressão linear só pode ser usado, depois de validado um conjunto de pressupostos do próprio modelo. Assim após a estimação dos coeficientes da regressão é necessário proceder à validação dos pressupostos respeitantes aos erros ou resíduos e à (quasi)ortogonalidade entre as variáveis independentes.

6.4.5.1 A análise de resíduos [29]

Os erros ou resíduos do modelo de regressão linear servem para estimar os coeficientes de regressão mas também para validar os pressupostos da sua aplicação. Assim a inferência sobre o modelo só é válida quando:

- $\varepsilon_j \sim N(0, \sigma)$ - os erros possuem distribuição normal de média nula e variância constante;
- $\text{Cov}(\varepsilon_k, \varepsilon_l) = 0$ ($k \neq l$; $k, l = 1, \dots, n$) - os erros são independentes.

A normalidade da distribuição dos erros pode ser feita por um procedimento gráfico. Idealmente os erros ou resíduos quando representados graficamente em função dos valores de \hat{y}_j têm uma distribuição aleatória em torno de $e_i = 0$. O pressuposto da distribuição normal dos erros também pode ser verificado através do gráfico de probabilidade normal, onde no eixo das abcissas se representa a probabilidade observada acumulada dos erros, e no eixo das ordenadas se representa a probabilidade acumulada que se observaria se os erros possuísem distribuição normal. Quando os erros possuem distribuição normal os valores representados neste gráfico distribuem-se em torno da diagonal principal. Também é possível testar a normalidade recorrendo a testes de ajustamento, como o caso do teste de Kolmogorov-Smirnov (K-S). O pressuposto da independência dos resíduos, pode ser verificado pelo teste de Durbin-Watson, em que se testa a presença de auto-correlação entre os erros ou resíduos do modelo de regressão linear. Assim se $d \approx 2$, concluí-se que

não existe auto-correlação entre os resíduos; para d muito menor do que 2 existe auto-correlação positiva, e para valores de d muito maiores do que 2 existe auto-correlação negativa. O SPSS não produz o p -value associado a esta estatística de teste, logo de uma forma mais ou menos empírica, considera-se que não existe auto-correlação para $d \approx 2,0 (\pm 0,2)$ [29].

6.4.5.2 Multicolinearidade [29]

O modelo de regressão linear múltipla pressupõe que as variáveis explicativas são linearmente independentes, isto é, não estão correlacionadas (são ortogonais), por isso não se verifica multicolinearidade. A intensidade da multicolinearidade é analisada essencialmente através de dois pontos:

- Correlação entre as variáveis independentes [29]

A partir da matriz das correlações das variáveis independentes, é possível verificar se estas estão ou não correlacionadas entre si. No entanto não existe um valor limite a partir do qual seja possível prever problemas de estimação do modelo por causa da colinearidade entre as variáveis independentes. Considera-se que correlações elevadas $|R| > 0,75$ normalmente conduzem a problemas de multicolinearidade [29].

- Tolerância e VIF

A Tolerância mede o grau em que uma variável X_i é explicada por todas as outras variáveis independentes, eq. 23.

$$T = 1 - R_i^2 \text{ (eq.23)}$$

A Tolerância varia entre 0 e 1, e quanto mais próxima estiver de zero, maior será a multicolinearidade. Normalmente considera-se como o limite a partir do qual há multicolinearidade 0,1 [31].

O inverso da Tolerância designa-se por VIF (Variance Inflation Factor) e quanto mais próximo estiver de zero menor será a multicolinearidade. Considera-se como limite acima do qual existe multicolinearidade 10 [31].

6.4.6 Selecção de regressores

Num problema de regressão linear múltipla podem-se conhecer à partida quais as variáveis independentes (regressores) a incluir no modelo de regressão.

Porém, principalmente na fase inicial da análise de regressão pode ser desconhecida qual ou quais as variáveis que conduzem ao melhor modelo e a decisão de escolha pode ser complicada pela presença de multicolinearidade e dos seus efeitos sobre a magnitude dos coeficientes de regressão. Os procedimentos de selecção de variáveis apresentam a vantagem de indicar com base num critério exacto quais as variáveis que apresentam relações mais fortes com a variável dependente. A selecção dos regressores pode ser feita por vários métodos:

6.4.6.1 Método Exaustivo [30]

Conceptualmente muito simples consiste em construir os modelos de regressão que combinem de todas as maneiras possíveis os potenciais regressores. Ordenar os modelos de acordo com o critério de qualidade escolhido e avaliar os modelos considerados melhores.

Este método apresenta no entanto uma limitação enorme que se prende com a dificuldade associada à construção de todos os modelos de regressão possíveis. Reconhecida esta limitação, foram propostos métodos alternativos que, de alguma forma, constituem uma aproximação ao método exaustivo.

6.4.6.2 Método Progressivo (ou *Forward*) [30]

Inicialmente são ajustados tantos modelos de regressão linear simples quantos os potenciais regressores. De entre os regressores, que isoladamente explicam uma proporção significativa da variação total da variável dependente, é incluído no modelo aquele que explica maior proporção. De seguida são construídos modelos de regressão linear dupla que associem, como variáveis independentes, o regressor seleccionado no passo anterior e cada um dos potenciais regressores ainda não incluídos. Neste processo está envolvida a re-estimação de parâmetros já calculados no passo anterior. O próximo regressor a adicionar ao modelo é aquele que explica uma proporção adicional significativa da variação total (adicional relativamente aquela que era explicada no passo anterior). Prossegue-se a tentativa de construção de modelos de ordem superior. O método termina quando nenhum dos potenciais

regressores ainda não incluídos no modelo, ao entrar nele, explicar uma proporção adicional significativa da variação total ou quando todos os potenciais regressores forem incluídos no modelo.

6.4.6.3 Método Regressivo (ou *Backward*) [30]

Este método tem uma base conceptual semelhante ao método progressivo, é um procedimento de alguma forma simétrico. O modelo é iniciado com todos os potenciais regressores e posteriormente, são retirados ao modelo, um a um os regressores cuja presença não contribua para explicar uma proporção significativa da variação total da variável dependente.

6.4.6.4 Método de Regressão Passo a Passo (ou *Stepwise*) [30]

Este método é um híbrido dos dois métodos anteriores. O primeiro passo inicia-se só com um regressor, tal como no método Progressivo, mas a significância da adição de uma nova variável independente ao modelo é testada como no método Regressivo. A vantagem deste método, é que permite a remoção de uma variável cuja importância no modelo se torna reduzida pela adição de novas variáveis (o método Passo a Passo é particularmente apropriado quando existem correlações significativas entre as variáveis independentes).

6.5 ABORDAGEM AOS DADOS

Inicialmente foram consideradas as 375 amostras e foi estudada a influência da discriminação da composição da amostra, na previsão do RON pelo modelo. Foram desenvolvidos três modelos com o objectivo de verificar se o aumento da discriminação da composição da amostra, usada para a elaboração dos modelos, provoca ou não um aumento da exactidão da previsão do RON por parte dos modelos.

No primeiro modelo a composição das amostras foi agrupada, de acordo com a sua família química, em 4 variáveis independentes: parafinas, olefinas, compostos nafténicos e compostos aromáticos. No segundo modelo a composição foi agrupada em 7 variáveis independentes: *i*-parafinas, *n*-

parafinas, *i*-olefinas, *n*-olefinas, *c*-olefinas, compostos nafténicos e compostos aromáticos. Por último no terceiro modelo a discriminação da composição da amostra foi máxima, considerando-se 47 variáveis independentes.

A fim de validar os modelos e verificar a homogeneidade dos dados, estes foram divididos em três grupos. No primeiro grupo foram considerados os dados referentes a 2003, no segundo os referentes a 2003-2004 e no terceiro os de 2003-2005. Com cada um destes grupos foi desenvolvido um modelo, considerando as 47 variáveis independentes, modelo esse posteriormente aplicado aos dados não incluídos na sua elaboração com o objectivo de validar cada um dos modelos.

Em todos os modelos as variáveis independentes foram caracterizadas e foi analisada a correlação que existia entre as mesmas. Foram retiradas as amostras cujos resultados eram atípicos e foi feito o diagnóstico aos pressupostos da regressão linear para cada um dos modelos estimados.

Depois de elaborado cada um dos modelos, estes foram usados para prever o RON e os resultados obtidos foram comparados com os obtidos pelo motor CFR.

7. RESULTADOS OBTIDOS COM A METODOLOGIA PROPOSTA

7.1 MODELO COM 4 VARIÁVEIS INDEPENDENTES

7.1.1 Análise das Variáveis

Os resultados cromatográficos obtidos foram agrupados, de acordo com a sua família química, em 4 variáveis: *Parafinas*, *Olefinas*, *Nafténicos* e *Aromáticos*. Considerando os resultados das 375 amostras disponíveis foi calculada a média e o desvio padrão para cada uma das variáveis, Tabela 8.

Tabela 8- Estatística descritiva da variável RON e das 4 variáveis independentes: *Parafinas*, *Olefinas*, *Nafténicos* e *Aromáticos*.

	Média	Desvio Padrão	N
RON	97,576	1,164	375
Olefinas	0,981	0,248	375
Parafinas	29,898	2,851	375
Aromáticos	67,269	3,073	375
Nafténicos	1,180	0,416	375

De seguida foi analisada a correlação existente entre a variável dependente *RON* e cada uma das variáveis independentes, assim como as correlações entre as variáveis independentes, Tabela 9, Figura 15 e Figura 16.

Tabela 9- Correlação entre as 4 variáveis independentes (*Parafinas*, *Olefinas*, *Nafténicos* e *Aromáticos*) e com a variável dependente *RON*.

	RON	Olefinas	Parafinas	Aromáticos	Nafténicos
RON	1,000	-0,244	-0,676	0,743	-0,601
Olefinas	-0,244	1,000	-0,069	-0,117	0,534
Parafinas	-0,676	-0,069	1,000	-0,969	0,307
Aromáticos	0,743	-0,117	-0,969	1,000	-0,497
Nafténicos	-0,601	0,534	0,307	-0,497	1,000

Verifica-se que a variável independente *Aromáticos* é a que mais se correlaciona com a variável dependente *RON* (0,743), e é a única que apresenta uma correlação positiva. A correlação positiva justifica-se pelo facto dos compostos aromáticos terem um RON muito elevado, sempre superior a

100 [13], assim quanto maior for a percentagem de compostos aromáticos presentes na amostra, mais elevado será o RON da amostra.

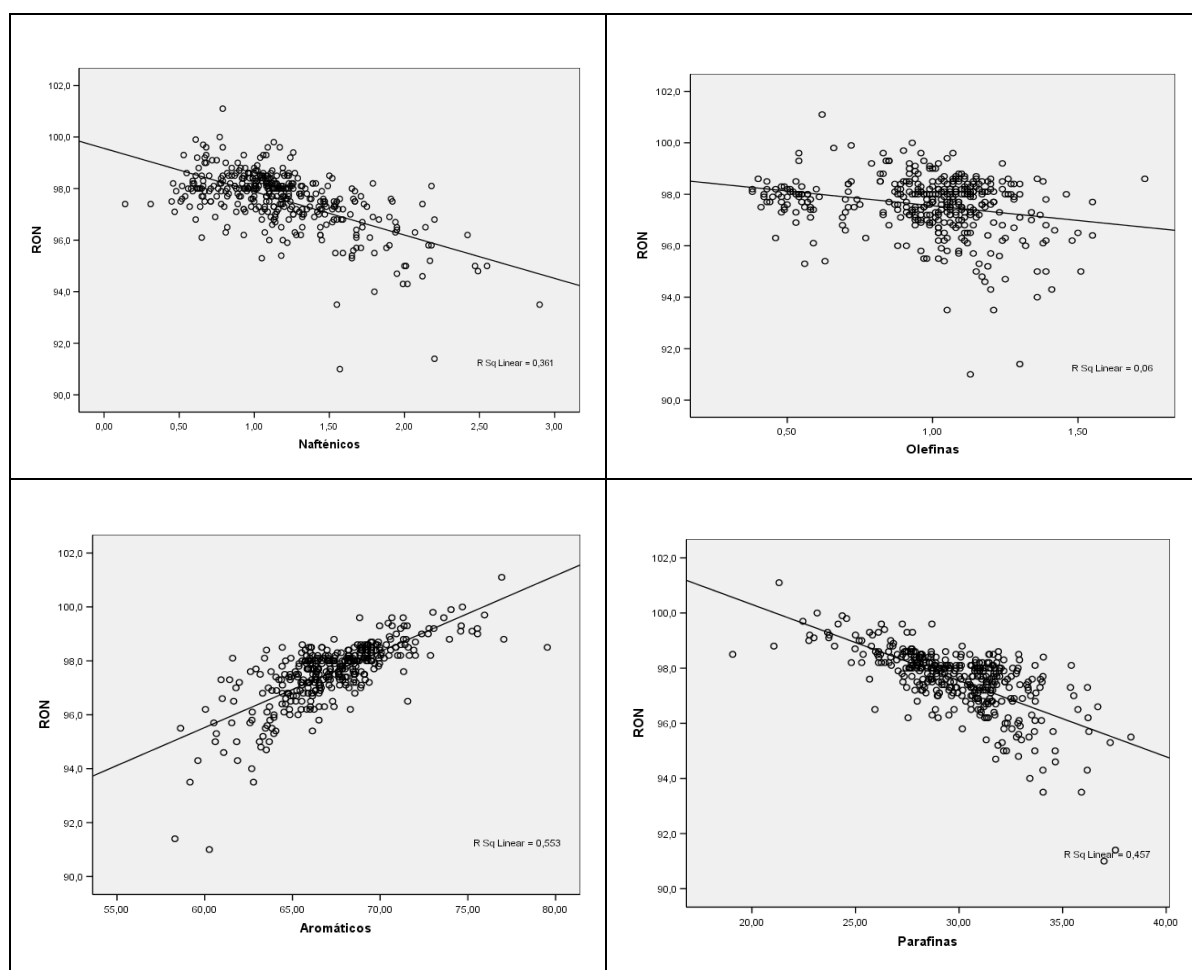


Figura 15- Representação gráfica da correlação entre cada uma das 4 variáveis independentes com a variável dependente RON.

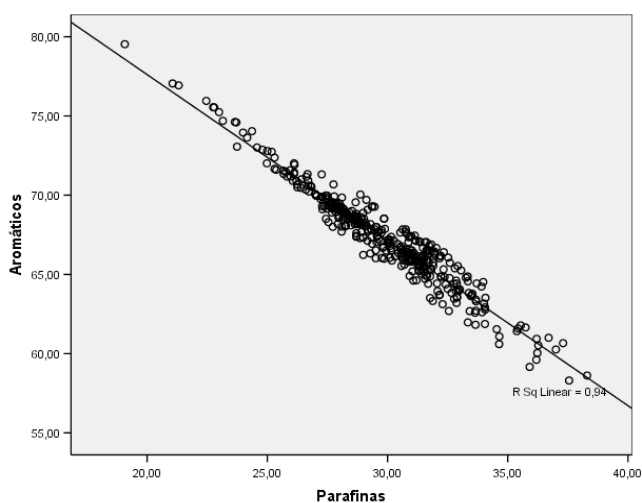


Figura 16- Representação gráfica da correlação entre as variáveis independentes *Parafinas* e *Aromáticos*.

A forte correlação que existe entre as duas variáveis deve-se a que mais de 60% da amostra é constituída por compostos aromáticos, logo é normal que esta influencie e muito o valor de RON da amostra.

As restantes variáveis independentes apresentam uma correlação mais fraca com a variável dependente RON, *Parafinas* (0,676), *Nafténicos* (0,601) e por fim a variável *Olefinas* (0,244), e em todos os casos a correlação é negativa. Verifica-se que quanto menor é a percentagem dessa variável independente na mistura menor é a correlação que esta tem com o RON, o que se percebe pelo facto da influência no valor final do RON na mistura ser menor. Por outro lado as correlações negativas são justificadas pelo facto do RON das parafinas, compostos nafténicos e olefinas variar com o tamanho da cadeira e com a complexidade das ramificações [13]. Assim um aumento da percentagem destas variáveis na mistura não corresponde a um aumento do valor de RON na mistura, depende sempre do composto cuja percentagem aumenta e do seu valor de RON.

Quanto às variáveis independentes, apresentam uma correlação fraca entre si à excepção das variáveis *Aromáticos* e *Parafinas*. A correlação apresentada por estas duas variáveis é negativa e muito forte (0,969), é superior à apresentada por qualquer das variáveis independentes relativamente ao RON. Este valor de correlação resulta simplesmente das amostras serem constituídas essencialmente por Parafinas e Aromáticos, a soma das duas é em média ~98% da composição das amostras. Assim quando a percentagem de uma delas aumenta na mistura a outra diminui e vice-versa, pelo que fatalmente estarão fortemente correlacionadas.

7.1.2 Estimação do modelo

Uma vez que é desconhecida qual ou quais as variáveis independentes a incluir no modelo de regressão, a selecção de regressores foi feita por três métodos diferentes: *Backward*, *Forward* e *Stepwise*. No entanto verifica-se que o ajuste do modelo obtido pelos métodos *Backward* e *Forward* não é melhor do que o desenvolvido pelo método *Stepwise*. Assim, e uma vez que este último é um híbrido dos outros dois, foi considerado apenas este método

de selecção de regressor para a estimação deste e dos próximos modelos de regressão linear.

O método de selecção de regressores *Stepwise*, quando aplicado às 375 amostras e considerando as quatro variáveis independentes, desenvolveu quatro modelos (M_1 , M_2 , M_3 e M_4), cujos coeficientes e regressores são apresentadas na Tabela 12.

No primeiro modelo (M_1) foi incluída apenas a variável independente *Aromáticos*, uma vez que é a que apresenta maior correlação em valor absoluto com a variável RON. A Tabela 10 mostra ainda que 55,3% da variação do RON no modelo M_1 é explicada pela variável *Aromáticos*. Este resultado é confirmado com a análise da ANOVA, Tabela 11, em que a soma total dos quadrados (507,019), na sua maior parte (280,224), o que corresponde a 55,3%, é explicada pela variável *Aromáticos*.

Tabela 10- Características dos modelos com 4 variáveis.

Modelo	r	R ²	R _a ²
M ₁	0,743	0,553	0,551
M ₂	0,790	0,624	0,622
M ₃	0,793	0,628	0,625
M ₄	0,792	0,628	0,626

Tabela 11- Resultados da análise ANOVA aos modelos com 4 variáveis.

Modelo		Soma dos Quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	p-value
M ₁	Regressão	280,224	1	280,224	460,871	0,000
	Residual	226,795	373	0,608		
	Total	507,019	374			
M ₂	Regressão	316,220	2	158,110	308,267	0,000
	Residual	190,799	372	0,513		
	Total	507,019	374			
M ₃	Regressão	318,438	3	106,146	208,824	0,000
	Residual	188,581	371	0,508		
	Total	507,019	374			
M ₄	Regressão	318,388	2	159,194	313,947	0,000
	Residual	188,631	372	0,507		
	Total	507,019	374			

Uma vez que antes da entrada da variável *Aromáticos* no modelo (M_1) o R^2 era nulo, a variação do coeficiente de determinação com a entrada da variável *Aromáticos* é igual ao valor desse coeficiente ($R^2=0,553$).

Do mesmo modo, a alteração do valor do teste F usado na ANOVA, Tabela 11, aquando da introdução da variável *Aromáticos* é igual ao valor do teste F (460,871).

A variável seguinte a entrar no modelo (M_2) é a variável independente que apresentar a maior correlação com a variável dependente RON depois de ajustados os efeitos da variável *Aromáticos* sobre a variável RON [29]. A adição da nova variável conduz a um maior acréscimo da estatística F -parcial da ANOVA da regressão de RON nessa variável, depois de considerado o efeito da variável *Aromáticos* [29]. Assim ao contrário do previsto na Tabela 9, a próxima variável independente a entrar no modelo é a variável *Nafténicos*, e só no terceiro modelo (M_3) é incluída a variável *Parafinas*.

De acordo com o teste F , Tabela 11, verifica-se que os quatro modelos ajustados em termos globais são significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Assim rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

A Tabela 12 apresenta os coeficientes de cada modelo, o seu desvio padrão, os testes t , e as tolerâncias de cada variável, para os quatro modelos desenvolvidos pelo método *Stepwise*.

O teste t permite testar a hipótese nula de inexistência de uma relação linear entre a variável dependente RON com cada uma das variáveis independentes, bem como com as constantes consideradas nos quatro modelos [31]. Em todos os modelos, à excepção do M_3 , os testes t têm associado um nível de significância inferior a 0,05, o que nos permite concluir que todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes de cada variável independente são diferentes de zero [31]. No caso do modelo M_3 , verifica-se que a variável *Aromáticos* não tem capacidade explicativa significativa do RON e por isso essa variável foi eliminada quando da elaboração do modelo seguinte, M_4 .

Tabela 12- Coeficientes dos modelos com 4 variáveis.

Modelo	Coeficientes		t	p-value	Estatística de Colinearidade		
	B	Desvio Padrão			Tolerância	VIF	
M ₁	(Constante)	78,627	0,884	88,985	0,000		
	Aromáticos	0,282	0,013	21,468	0,000	1,000	1,000
M ₂	(Constante)	83,534	1,001	83,466	0,000		
	Aromáticos	0,224	0,014	16,113	0,000	0,753	1,329
	Nafténicos	-0,859	0,103	-8,378	0,000	0,753	1,329
M ₃	(Constante)	102,733	9,245	11,112	0,000		
	Aromáticos	0,030	0,094	0,314	0,753	0,016	61,410
	Nafténicos	-1,166	0,179	-6,515	0,000	0,245	4,086
	Parafinas	-0,193	0,092	-2,089	0,037	0,020	51,038
M ₄	(Constante)	105,638	0,388	272,137	0,000		
	Nafténicos	-1,214	0,093	-13,064	0,000	0,906	1,104
	Parafinas	-0,222	0,014	-16,337	0,000	0,906	1,104

A variável independente *Olefinas* apresenta uma correlação muito fraca com a variável RON, não tem uma capacidade explicativa significativa e por isso não foi incluída em nenhum dos modelos. O método de selecção de regressores *Stepwise* criou estes quatro modelos, até conseguir que as variáveis não introduzidas no modelo não tenham capacidade explicativa da variável dependente e todas as que estão no modelo o tenham [29]. O modelo final estimado corresponde ao modelo M₄ e é dado por:

$$RON = 105,638 - 1,214 * Nafténicos - 0,222 * Parafinas$$

Depois de estimado o modelo de regressão, é necessário analisar as hipóteses que estão na sua base.

7.1.3 Validação dos pressupostos do modelo estimado

7.1.3.1 Homocedasticidade

Quando não se verifica a hipótese de homocedasticidade, os parâmetros do modelo estimados embora sejam centrados, não são, eficientes [31]. Para avaliar a homocedasticidade pode-se observar as relações que existem, por um lado entre os resíduos normalizados e os valores estimados de Y, e por outro lado entre os resíduos normalizados e os valores estimados de Y também normalizados [31], Figura 17.

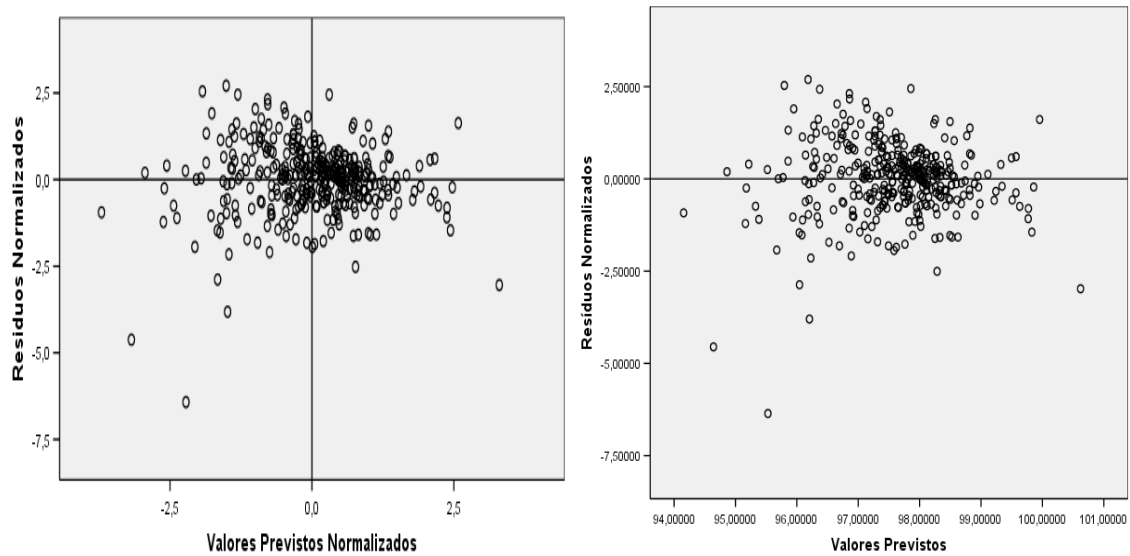


Figura 17- Avaliação da Homocedasticidade para o modelo M_4 , com 4 variáveis.

Da análise dos dois gráficos é visível que os resíduos têm uma amplitude aproximadamente constante em torno do eixo horizontal zero, ou seja, não existe uma tendência crescente ou decrescente, logo não se rejeita a hipótese da homocedasticidade.

7.1.3.2 Covariância Nula

A covariância é analisada pelo teste de Durbin-Watson em que para o modelo M_4 $d=1,102$, logo existe auto-correlação entre os resíduos.

7.1.3.3 Normalidade dos Resíduos

A Normalidade é testada usando o teste de Kolmogorov-Smirnov, Tabela 13. Os desvios à normalidade podem também ser observados graficamente, Figura 18.

Tabela 13- Teste à Normalidade dos resíduos, modelo M_4 , com 4 variáveis.

	Kolmogorov-Smirnov		
	Estadística	Graus Liberdade	Significância
Resíduos Padrão	0,069	375	0,000

De acordo com o teste Kolmogorov-Smirnov, verifica-se que os resíduos não seguem uma distribuição normal para um nível de significância de 0,05.

Graficamente também é visível a não normalidade, uma vez que nem todas as observações se dispõem ao longo das rectas oblíqua e horizontal.

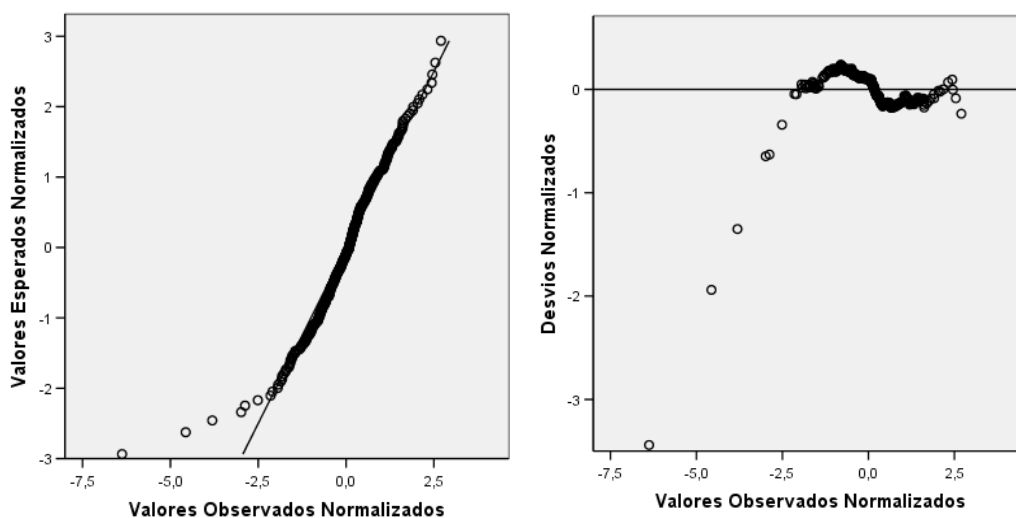


Figura 18- Análise gráfica da normalidade dos resíduos para o modelo M_4 , com 4 variáveis.

7.1.3.4 Multicolinearidade

O modelo de regressão linear múltipla pressupõe que não se verifique multicolinearidade, ou seja, que as variáveis explicativas sejam linearmente independentes [31]. A intensidade da multicolinearidade é analisada através da correlação entre as variáveis independentes e dos valores da Tolerância e do VIF. Relativamente à correlação entre as variáveis independentes já foi visto no ponto 7.1.1 que é muito forte entre as variáveis *Aromáticos* e *Parafinas* mas apenas pelo facto do somatório das duas representar em média ~98% da composição das amostras.

Os valores obtidos para a Tolerância e para o VIF, [Tabela 12](#), confirmam a multicolinearidade entre as variáveis independentes *Aromáticos* e *Parafinas*, no modelo M_3 . No entanto esta situação é ultrapassada uma vez que a variável *Aromáticos* é excluída no modelo M_4 .

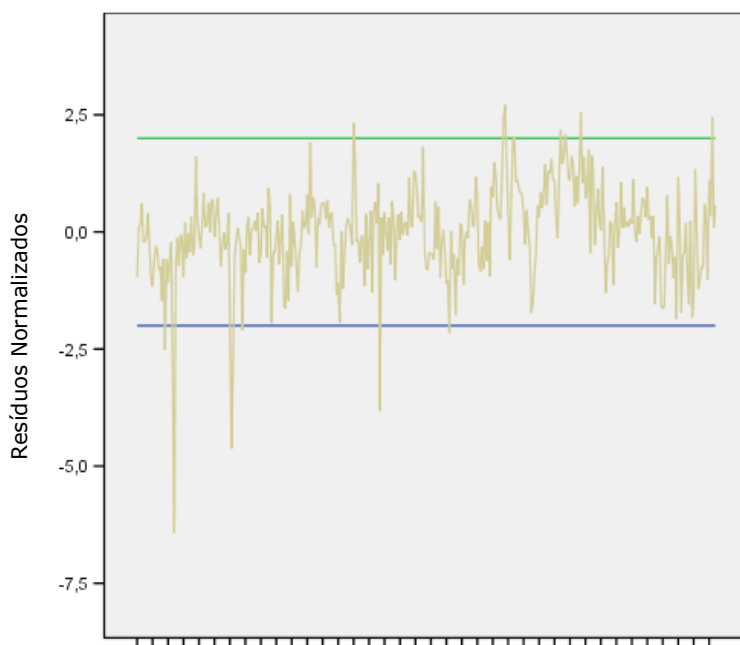
7.1.4 Análise dos Valores Atípicos

Da análise gráfica dos resíduos é possível identificar eventuais valores atípicos ou aberrantes. Estes valores são observações extremas, que apresentam

resíduos bastante superiores, em valor absoluto, aos resíduos das restantes observações, por exemplo um valor superior a dois desvios-padrão relativamente à média [29]. Assim é necessário proceder à análise dos resíduos para identificar e eliminar os valores atípicos, [Figura 19](#).

Neste caso são considerados atípicos todos os resíduos normalizados com valores superiores a 2 em valor absoluto. Verifica-se a existência de alguns valores atípicos. As amostras correspondentes a estes valores devem ser retiradas e o modelo deve ser estimado novamente sem a sua presença.

Estes valores atípicos podem ser consequência de um erro analítico, de um erro de transcrição de resultados, ou mesmo de um dia de produção anómalo que provocou alterações na composição da amostra desse dia. No entanto a esta distância não é possível rastrear qual o motivo destes valores atípicos.



[Figura 19](#)- Análise dos Resíduos, modelo M_4 , com 4 variáveis.

Verifica-se que este modelo M_4 não cumpre os requisitos necessários para um Modelo de Regressão Linear Multipla. Os resíduos são auto-correlacionados e não seguem uma distribuição normal, e confirma-se a presença de valores atípicos. Assim serão excluídos os valores atípicos, encontrados a partir da análise de resíduos, e desenvolvido um novo modelo considerando as mesmas 4 variáveis independentes (*Parafinas, Olefinas, Nafténicos e Aromáticos*)

7.2 MODELO COM 4 VARIÁVEIS INDEPENDENTES EXCLUÍDO OS VALORES ATÍPICOS

7.2.1 Análise das Variáveis

Retirando as amostras que apresentavam valores atípicos, o número de amostras passa a ser de 359 e cada variável apresenta uma nova média e desvio padrão, Tabela 14.

Tabela 14- Estatística descritiva do RON e das 4 variáveis independentes: *Parafinas, Olefinas, Nafténicos e Aromáticos*, sem as amostras atípicas.

	Média	Desvio Padrão	N
RON	97,630	1,0203	359
Nafténicos	1,1675	0,41138	359
Aromáticos	67,3736	2,92615	359
Parafinas	29,8039	2,72542	359
Olefinas	0,9757	0,24814	359

As correlações entre a variável dependente *RON* e cada uma das variáveis independentes, assim como as correlações entre as variáveis independentes, são idênticas às obtidas antes da exclusão das amostras atípicas e apresentadas na Tabela 9.

7.2.2 Estimação do Modelo

O método de selecção de regressores *Stepwise*, quando aplicado às 359 amostras, considerando as 4 variáveis independentes, desenvolveu 5 modelos (M_1 , M_2 , M_3 , M_4 e M_5), os seus coeficientes e regressores estão apresentados na Tabela 15. A primeira variável independente a entrar no modelo M_1 foi novamente a *Aromáticos*, uma vez que continua a apresentar o maior coeficiente de correlação em valor absoluto [29]. A Tabela 16 mostra ainda que 62,1% da variação do RON no modelo M_1 é explicada pela variável *Aromáticos*.

Tal como no ponto 7.1, e pelos mesmos motivos, mesmo depois da exclusão das amostras atípicas, a segunda variável a entrar no modelo M_2 foi a variável *Nafténicos*, a terceira variável a ser incluída no modelo M_3 foi a variável *Parafinas* e no modelo M_4 foi novamente excluída a variável *Aromáticos*. No

entanto neste caso o método *Stepwise* desenvolveu mais um modelo M_5 no qual incluí a variável *Olefinas*, o que não aconteceu no ponto 7.1. Assim a variável *Olefinas*, depois da exclusão das amostras atípicas, passa a ter uma capacidade explicativa significativa e por isso é incluída no último modelo (M_5) [29]. Verifica-se ainda que os coeficientes estimados agora para os modelos de M_1 a M_4 são idênticos aos obtidos no ponto 7.1 antes da exclusão das amostras atípicas.

Tabela 15- Coeficientes dos modelos, com 4 variáveis, sem as amostras atípicas.

Modelo		Coeficientes		t	p-value	Estatística da Colinearidade	
		B	Desvio Padrão			Tolerance	VIF
M ₁	(Constante)	79,120	0,766	103,246	0,000	1,000	1,000
	Aromáticos	0,275	0,011	24,176	0,000		
M ₂	(Constante)	83,651	0,829	100,870	0,000	0,769	1,300
	Aromáticos	0,221	0,012	19,147	0,000		
	Nafténicos	-0,792	0,082	-9,643	0,000		
M ₃	(Constante)	98,283	7,412	13,260	0,000	0,018	55,887
	Aromáticos	0,073	0,075	0,969	0,333		
	Nafténicos	-1,028	0,144	-7,136	0,000		
	Parafinas	-0,147	0,074	-1,986	0,048		
M ₄	(Constante)	105,458	0,324	325,203	0,000	0,923	1,084
	Nafténicos	-1,147	0,075	-15,355	0,000		
	Parafinas	-0,218	0,011	-19,305	0,000		
M ₅	(Constante)	105,932	0,377	280,812	0,000	0,593	1,685
	Nafténicos	-1,014	0,093	-10,956	0,000		
	Parafinas	-0,227	0,012	-19,149	0,000		
	Olefinas	-0,358	0,148	-2,411	0,016		

Tabela 16- Características dos modelos com 4 variáveis, sem as amostras atípicas.

Modelo	r	R ²	R _a ²
M ₁	0,788	0,621	0,620
M ₂	0,836	0,699	0,698
M ₃	0,838	0,703	0,700
M ₄	0,838	0,702	0,700
M ₅	0,841	0,707	0,704

Verifica-se que o melhor ajuste é o do modelo M₅ com $r=0,841$; $R^2=0,707$ e $R_a^2=0,704$; a exclusão das amostras atípicas foi uma boa decisão uma vez que aumentou o coeficiente de determinação dos modelos. O R_a^2 passou de 0,626 para 0,704, Tabela 16.

De acordo com o teste *F*, Tabela 17, todos os cinco modelos ajustados são globalmente significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) em cada modelo, e verifica-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

Tabela 17- Resultados da análise ANOVA aos modelos com 4 variáveis, sem amostras atípicas.

Modelo		Soma dos quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	<i>p-value</i>
M ₁	Regressão	231,360	1	231,360	584,493	0,000
	Residual	141,311	357	0,396		
	Total	372,671	358			
M ₂	Regressão	260,626	2	130,313	414,042	0,000
	Residual	112,045	356	0,315		
	Total	372,671	358			
M ₃	Regressão	261,858	3	87,286	279,627	0,000
	Residual	110,814	355	0,312		
	Total	372,671	358			
M ₄	Regressão	261,564	2	130,782	419,043	0,000
	Residual	111,107	356	0,312		
	Total	372,671	358			
M ₅	Regressão	263,355	3	87,785	285,077	0,000
	Residual	109,316	355	0,308		
	Total	372,671	358			

A Tabela 15 apresenta os coeficientes dos vários modelos desenvolvidos e o teste *t* aos coeficientes dos regressores e tal como antes da exclusão das amostras atípicas, verifica-se no modelo M₃ que a variável *Aromáticos* deixa de ser significativa e é excluída logo no modelo seguinte. O modelo que apresenta o melhor ajuste é modelo M₅:

$$RON = 105,932 - 1,014 * Nafténicos - 0,227 * Parafinas - 0,358 * Olefinas$$

7.2.3 Validação dos pressupostos do modelo estimado

7.2.3.1 Homocedasticidade

Tal como no modelo anterior, os resíduos apresentam uma amplitude aproximadamente constante em torno do eixo horizontal zero, e não se verificam tendências crescentes ou decrescentes. Assim não se rejeita a hipótese da homocedasticidade.

7.2.3.2 Covariância Nula

Para o modelo M_5 , o teste de Durbin-Watson tem um valor $d=1,192$, de acordo com os critérios deste teste não se pode dizer que não existe auto-correlação entre os resíduos. No entanto de seguida é possível verificar que apesar da auto-correlação os resíduos seguem uma distribuição normal.

7.2.3.3 Normalidade dos Resíduos

Depois de retiradas as amostras atípicas e de acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, para um nível de significância de 0,05, Tabela 18. Graficamente verifica-se que os valores esperados e observados normalizados se dispõem ao longo da recta oblíqua; os desvio normalizados também se dispõem ao longo da recta horizontal mas com uma grande amplitude e revelando uma certa tendência oscilatória, Figura 20.

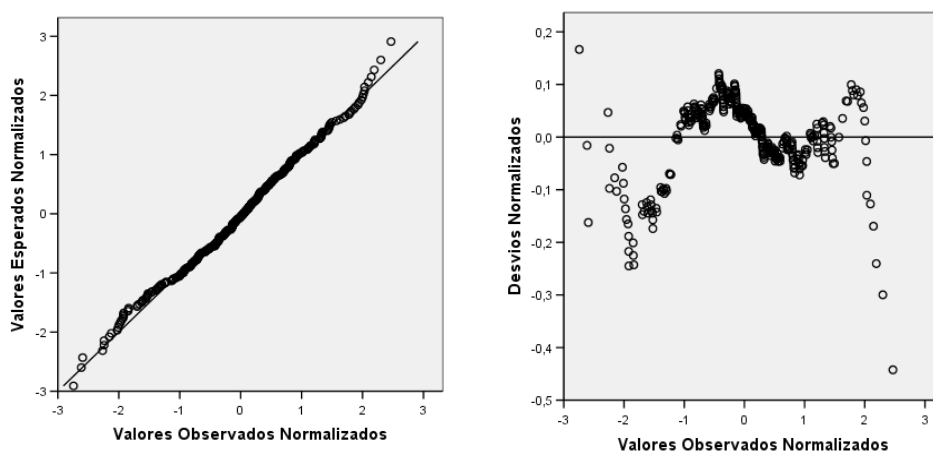


Figura 20- Análise gráfica da normalidade dos resíduos para o modelo M_5 com 4 variáveis, sem amostras atípicas.

Tabela 18- Teste à normalidade dos resíduos, modelo M₅ com 4 variáveis, sem amostras atípicas.

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,038	359	0,200

7.2.3.4 Multicolinearidade

Neste modelo M₃, tal como no anterior modelo M₃ de 7.1, e pelos mesmos motivos verifica-se a existência de multicolinearidade entre as variáveis *Aromáticos* e *Parafinas*, Tabela 15. Esta multicolinearidade é confirmada pelo baixo valor da Tolerância e elevado valor do VIF apresentados no modelo M₃, Tabela 15.

Com a exclusão dos valores atípicos verifica-se que o coeficiente de correlação do modelo M₅ é melhor, a homocedasticidade mantém-se e os resíduos passam a ter uma distribuição normal, de acordo com o teste de Kolmogorov-Smirnov. Continua a existir multicolinearidade entre as variáveis *Aromáticos* e *Parafinas*, no entanto este facto não afecta o modelo seleccionado M₅ uma vez que a variável *Aromáticos* não foi incluída neste modelo.

7.2.4 Previsão do RON usando o modelo

Considerando o modelo desenvolvido é possível prever o RON para as amostras usadas na sua elaboração, a partir da sua composição obtida por cromatografia.

Os valores obtidos a partir do modelo matemático para o RON podem ser comparados com os valores obtidos para o RON pelo motor CFR, Figura 21. Idealmente a relação entre as duas variáveis deve ser uma recta com declive um e ordenada na origem igual a zero, no entanto a relação obtida foi um pouco diferente:

$$\text{RON}_{\text{modelo}} = 28,66 + 0,71 * \text{RON}_{\text{CFR}} \text{ com } R^2 = 0,71.$$

Outra forma de comparar os dois métodos de determinação do RON, e avaliar a exactidão dos resultados obtidos pelo modelo, face aos obtidos pelo CFR, é o

teste *t* de Student das médias [32]. Neste caso é necessário calcular as médias dos valores obtidos por cada um dos métodos, \bar{x}_1 e \bar{x}_2 , e os respectivos desvios padrão s_1 e s_2 , Tabela 19.

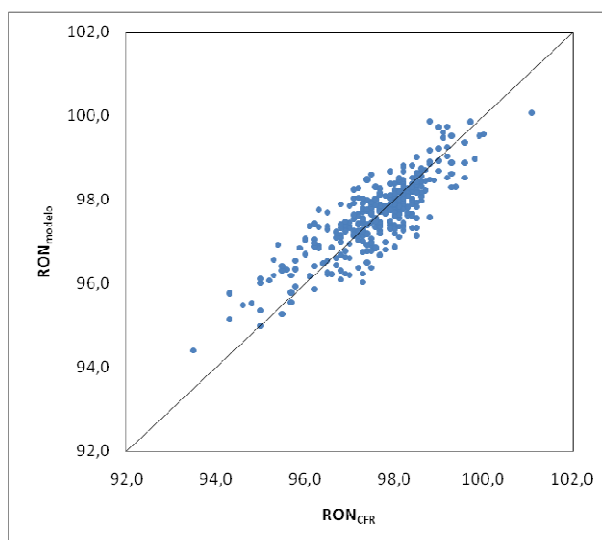


Figura 21- Representação gráfica do valor do RON_{CFR} e do RON_{modelo}, com 4 variáveis, sem as amostras atípicas.

Tabela 19- Resultados da comparação do RON_{CFR} e do RON_{modelo}, com 4 variáveis, sem amostras atípicas.

	RON _{CFR}	RON _{modelo}
Média	97,63	97,63
Desvio Padrão	1,02	0,86
Variância	1,04	0,73
n	359	359
<hr/>		
F _{calc}		1,42
F _{tab}		0,83
F _{calc} > F _{tab}		
<hr/>		
t _{calc}		0,05
t _{tab}		1,96
t _{calc} < t _{tab}		

Considera-se como hipótese nula que os dois métodos originam os mesmos resultados, $H_0: \mu_1 = \mu_2$, e é necessário avaliar se $(\bar{x}_1 - \bar{x}_2)$ é significativamente diferente de zero. Para decidir se a diferença entre as duas médias é significativa, e testar a hipótese nula, surge o teste *t de Student*, eq. 24 [32].

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (\text{eq.24})$$

em que:

$$s = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \quad (\text{eq.25})$$

O valor de *t* calculado (t_{calc}) é comparado com o valor de *t* tabelado (t_{tab}) para $(n_1 + n_2 - 1)$ graus de liberdade, Tabela 19. Como $t_{\text{calc}} < t_{\text{tab}}$, para um nível de significância de 0,05, não se rejeita a hipótese nula, considera-se que as médias dos resultados obtidos pelos dois métodos não são significativamente diferentes.

A fim de verificar se as variâncias obtidas pelos dois métodos são significativamente diferentes ou não, aplica-se o teste *F*, em que:

$$F = \frac{s_1^2}{s_2^2} \quad F \geq 1 \quad (\text{eq. 26})$$

Considera-se como hipótese nula que os dois métodos têm variâncias semelhantes, $H_0: \sigma_1 = \sigma_2$, e compara-se o valor de *F calculado* (F_{calc}) com o valor de *F tabelado* (F_{tab}) com $(n_1 - 1)$ e $(n_2 - 1)$ graus de liberdade, Tabela 19.

Como neste caso $F_{\text{calc}} > F_{\text{tab}}$, para um nível de significância de 0,05, rejeita-se a hipótese nula, as variâncias obtidas pelos dois métodos não são semelhantes; sendo que os resultados obtidos a partir do motor CFR apresentam uma variância superior aos obtidos pelo modelo.

É também importante avaliar as diferenças obtidas entre os resultados provenientes pelos dois métodos, Figura 22. Verifica-se que estas diferenças não apresentam uma tendência crescente ou decrescente no entanto em 22% dos casos o RON obtido pelo modelo difere do RON obtido pelo motor CFR mais do que $|0,7|$ (reprodutibilidade do método de referência ASTM D 2699).

Embora estes dados não sirvam para comparar os dois métodos são fundamentais para o uso do modelo como uma alternativa ao motor CFR, uma vez que este é o desvio máximo permitido pelo método de referência.

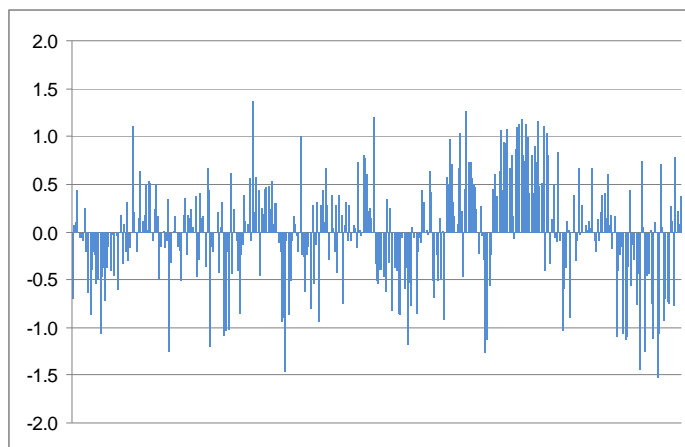


Figura 22- Diferença entre os resultados obtidos para o RON pelo modelo, com 4 variáveis sem as amostras atípicas, e pelo motor CFR.

7.3 MODELO COM 7 VARIÁVEIS INDEPENDENTES

7.3.1 Análise das Variáveis

Neste novo modelo foram consideradas 7 variáveis: *i-Parafinas*, *n-Parafinas*, *i-Olefinas*, *n-Olefinas*, *c-Olefinas*, *Nafténicos* e *Aromáticos*. Ao conjunto total das 375 amostras foram retirados as amostras atípicas, após análise de resíduos, e foram consideradas 355 amostras para desenvolver o modelo, Tabela 20.

Tabela 20- Estatística descritiva do RON e das 7 variáveis independentes: *i-Parafinas*, *n-Parafinas*, *i-Olefinas*, *n-Olefinas*, *c-Olefinas*, *Nafténicos* e *Aromáticos*.

	Média	Desvio Padrão	N
RON	97,653	1,0032	355
<i>i-Parafinas</i>	21,7637	2,11107	355
<i>n-Parafinas</i>	8,0492	0,86507	355
<i>i-Olefinas</i>	0,6799	0,17177	355
<i>n-Olefinas</i>	0,2499	0,08361	355
<i>c-Olefinas</i>	0,0457	0,03269	355
Nafténicos	1,1643	0,41075	355
Aromáticos	67,3710	2,94347	355

Da análise da correlação entre a variável dependente *RON* e cada uma das variáveis independentes verifica-se que a variável independente *Aromáticos* continua a ser a que apresenta uma correlação mais forte (0,786) com o *RON*, Tabela 21,

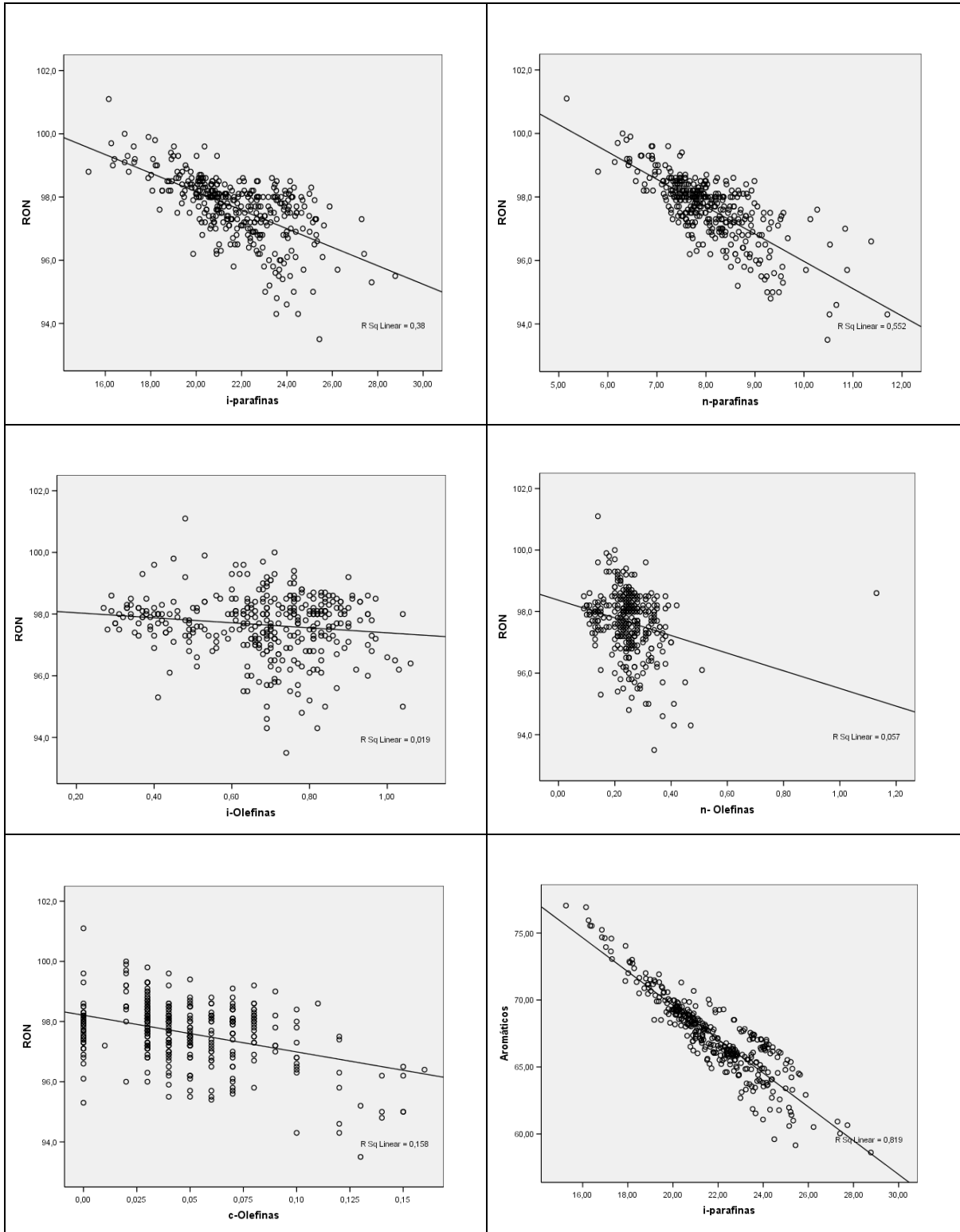


Figura 20- Representação gráfica da correlação entre as variáveis independentes e a variável dependente RON, e das variáveis independentes Aromáticos e *i*-Parafinas.

As correlações das restantes variáveis independentes com o RON não diferem das já apresentadas no ponto 7.1.1. Quanto às variáveis independentes correlacionam-se pouco entre si à excepção da variável *Aromáticos* com as variáveis *i-Parafinas* e *n-Parafinas*. A correlação apresentada por estas duas variáveis relativamente ao RON é muito forte, tal como já explicado em pontos anteriores, Tabela 21, Figura 20.

	RON	<i>i</i> -parafinas	<i>n</i> -parafinas	<i>i</i> -olefinas	<i>n</i> -olefinas	<i>c</i> -olefinas	nafténicos	aromáticos
RON	1,000	-0,617	-0,743	-0,139	-0,239	-0,398	-0,616	0,786
<i>i</i> -parafinas	-0,617	1,000	0,634	-0,349	0,018	-0,175	0,157	-0,905
<i>n</i> -parafinas	-0,743	0,634	1,000	0,165	0,434	0,290	0,502	-0,858
<i>i</i> -olefinas	-0,139	-0,349	0,165	1,000	0,451	0,673	0,388	0,044
<i>n</i> -olefinas	-0,239	0,018	0,434	0,451	1,000	0,454	0,465	-0,279
<i>c</i> -olefinas	-0,398	-0,175	0,290	0,673	0,454	1,000	0,738	-0,160
nafténicos	-0,616	0,157	0,502	0,388	0,465	0,738	1,000	-0,478
aromáticos	0,786	-0,905	-0,858	0,044	-0,279	-0,160	-0,478	1,000

Tabela 21- Correlação entre as 7 variáveis independentes (*i-Parafinas*, *n-Parafinas*, *i-Olefinas*, *n-Olefinas*, *c-Olefinas*, *Nafténicos* e *Aromáticos*) e com a variável dependente RON.

7.3.2 Estimação do Modelo

O método de selecção de regressores *Stepwise*, quando aplicado às 7 variáveis independentes desenvolveu cinco modelos (M_1 , M_2 , M_3 , M_4 , M_5), cujos coeficientes e regressores estão apresentados na Tabela 24.

A primeira variável independente a entrar no modelo (M_1) foi novamente a *Aromáticos*. Esta variável continua a ser aquela que apresenta maior coeficiente de correlação em valor absoluto com a variável RON. A Tabela 22 mostra ainda que 61,8% da variação do RON no modelo M_1 é explicada pela variável *Aromáticos*.

A variável adicionada ao modelo (M_2) é a *c-olefinas*, ao modelo (M_3) é a *n-olefinas*, ao modelo (M_4) os *nafténicos* e por fim as *n-parafinas* ao último modelo (M_5). A variável *i-parafinas* não é adicionada a nenhum dos modelos,

sendo por isso implícito que não apresenta uma capacidade explicativa significativa das variações da variável dependente RON. A ordem de entrada das variáveis no modelo é bastante diferente da prevista quando analisadas inicialmente as correlações entre a variável RON e cada uma das variáveis independentes, Tabela 21. No entanto como é sabido o método *Stepwise* recalcula o efeito de cada uma das restantes variáveis independentes na variável dependente RON após a adição de cada variável independente ao modelo [29]. E só termina a elaboração do modelo quando todas as variáveis não introduzidas no modelo não tiverem capacidade explicativa da variável RON e todas as que estejam no modelo tenham [29].

Tabela 22- Características dos modelos com 7 variáveis.

Modelo	r	R ²	R _a ²
M ₁	0,786	0,618	0,617
M ₂	0,833	0,694	0,692
M ₃	0,840	0,706	0,704
M ₄	0,848	0,718	0,715
M ₅	0,853	0,728	0,724

Tabela 23- Resultados da análise ANOVA aos modelos com 7 variáveis.

Modelo	Soma dos quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	p-value
M ₁ Regressão	220,158	1	220,158	570,991	0,000
Residual	136,107	353	0,386		
Total	356,265	354			
M ₂ Regressão	247,255	2	123,627	399,199	0,000
Residual	109,010	352	0,310		
Total	356,265	354			
M ₃ Regressão	251,549	3	83,850	281,059	0,000
Residual	104,716	351	0,298		
Total	356,265	354			
M ₄ Regressão	255,924	4	63,981	223,173	0,000
Residual	100,341	350	0,287		
Total	356,265	354			
M ₅ Regressão	259,463	5	51,893	187,088	0,000
Residual	96,802	349	0,277		
Total	356,265	354			

De acordo com o teste *F*, Tabela 23, verifica-se que todos os modelos ajustados em termos globais são significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Assim rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

De acordo com o teste *t*, Tabela 24, todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes das variáveis independente são diferentes de zero, em todos os modelos [31].

Tabela 24- Coeficientes dos modelos com 7 variáveis.

Modelo	Coeficientes		t	p-value	Estatísticas da Colinearidade		
	B	Desvio Padrão			Tolerância	VIF	
M ₁	(Constante)	79,603	0,756	105,281	0,000		
	Aromáticos	0,268	0,011	23,895	0,000	1,000	1,000
M ₂	(Constante)	81,018	0,694	116,688	0,000		
	Aromáticos	0,253	0,010	24,828	0,000	0,975	1,026
	c-Olefinas	-8,573	0,917	-9,354	0,000	0,975	1,026
M ₃	(Constante)	80,094	0,724	110,679	0,000		
	Aromáticos	0,262	0,010	25,478	0,000	0,921	1,086
	c-Olefinas	-10,208	0,997	-10,234	0,000	0,793	1,262
	n- Olefinas	1,521	0,401	3,794	0,000	0,750	1,334
M ₄	(Constante)	82,029	0,865	94,813	0,000		
	Aromáticos	0,238	0,012	20,249	0,000	0,676	1,480
	c-Olefinas	-6,304	1,398	-4,508	0,000	0,388	2,580
	n- Olefinas	1,689	0,395	4,272	0,000	0,741	1,350
	Nafténicos	-0,479	0,123	-3,906	0,000	0,319	3,134
M ₅	(Constante)	87,959	1,866	47,149	0,000		
	Aromáticos	0,178	0,020	8,743	0,000	0,218	4,596
	c-Olefinas	-5,470	1,395	-3,921	0,000	0,377	2,654
	n- Olefinas	2,147	0,409	5,243	0,000	0,668	1,496
	Nafténicos	-0,514	0,121	-4,243	0,000	0,317	3,154
	n-parafinas	-0,249	0,070	-3,572	0,000	0,216	4,636

O melhor ajuste é o apresentado pelo modelo M₅ com $r=0,853$; $R^2=0,728$ e $R_a^2=0,724$, Tabela 22. O uso de 7 variáveis em vez de 4 aumentou o R_a^2 do modelo, no entanto verifica-se que o melhor modelo obtido não usa as 7 variáveis, modelo M₅:

$$RON = 87,959 + 0,178 * \text{Aromáticos} - 5,470 * \text{c-Olefinas} + 2,147 * \text{n-Olefinas} - 0,514 * \text{Nafténicos} - 0,249 * \text{n-Parafinas}$$

Depois de estimado o modelo de regressão é necessário analisar as hipóteses que estão na sua base.

7.3.3 Validação dos pressupostos do modelo estimado

7.3.3.1 Homocedasticidade

Também neste modelo os resíduos apresentam uma amplitude aproximadamente constante em volta do eixo horizontal zero, e não se verificam tendências crescentes ou decrescentes, Figura 21, pelo que não se rejeita a hipótese da homocedasticidade.

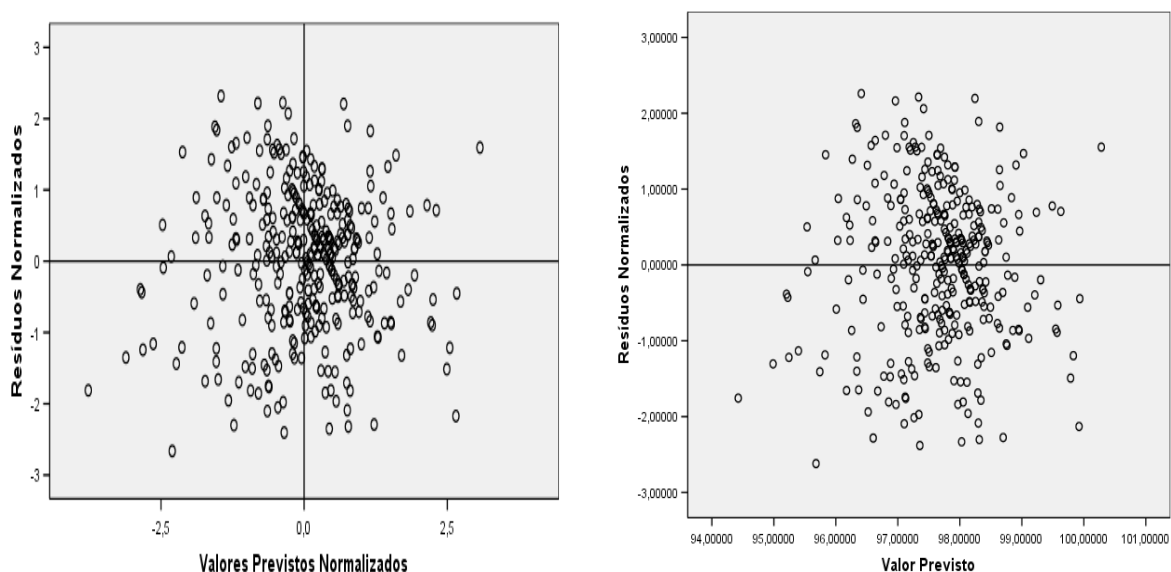


Figura 21- Avaliação da Homocedasticidade para o modelo M_5 com 7 variáveis.

7.3.3.2 Covariância Nula

Para o modelo M_5 , o teste de Durbin-Watson tem um valor de $d=1,360$, o que de acordo com os critérios deste teste não se pode dizer que não existe autocorrelação entre os resíduos.

7.3.3.3 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, para um nível de significância de 0,05, Tabela 25.

Graficamente os valores esperados normalizados e os valores observados normalizados dispõem-se ao longo da recta oblíqua, Figura 22.

Os desvios normalizados também se dispõem ao longo da recta horizontal, sendo que apresentam uma menor amplitude que no modelo anterior e revelam uma tendência oscilatória menor, Figura 22.

Tabela 25- Teste à Normalidade dos resíduos, para o modelo M₅ com 7 variáveis.

	Kolmogorov-Smirnov		
	Estadística	Graus Liberdade	Significância
Resíduos Padrão	0,051	355	0,200

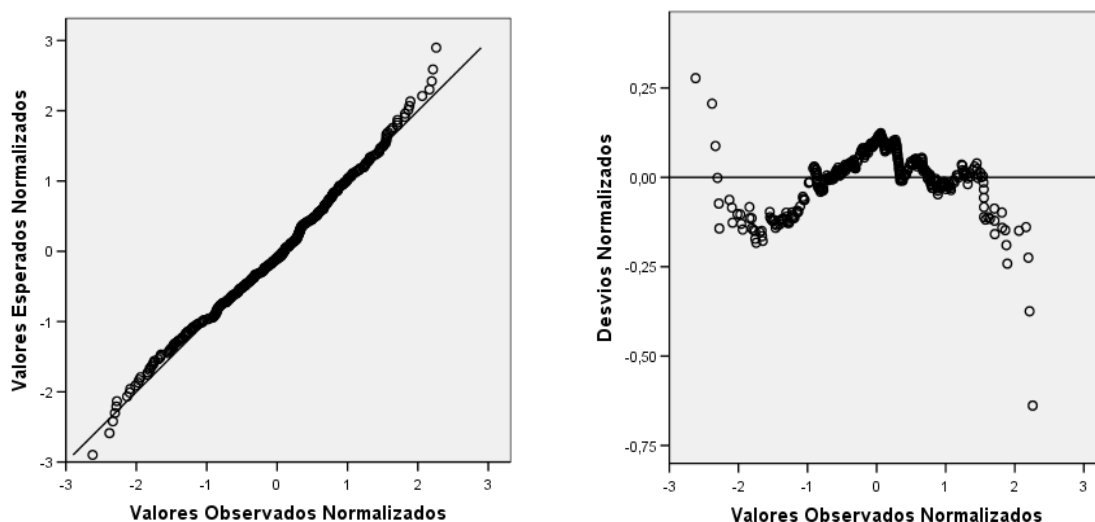


Figura 22- Análise gráfica da normalidade dos resíduos, para o modelo M₅, com 7 variáveis.

7.3.3.4 Multicolinearidade

Apesar de se verificar uma correlação forte entre as variáveis *Aromáticos* e *i-Parafinas* e *Aromáticos* e *n-Parafinas*, Tabela 21, esta não se confirma quando da elaboração dos modelos. A variável *i-Parafinas* não é incluída em nenhum dos modelos desenvolvidos logo o problema da multicolinearidade não se põe.

A variável *n-Parafinas* é introduzida apenas no modelo M₅ e não é verificada multicolinearidade uma vez que os valores da Tolerância são todos superiores a 0,1 e os valores de VIF são inferiores a 10, Tabela 24.

7.3.4 Previsão do RON usando o modelo

Considerando os coeficientes obtidos no modelo M₅, foi previsto o valor do RON para as amostras usadas na sua elaboração. Estes valores foram comparados com os valores do RON obtidos pelo motor CFR, Figura 23. Verifica-se que a recta obtida se aproxima pouco mais da relação ideal (declive=1 e ordenada na origem=0) e o coeficiente de determinação melhora:

$$\text{RON}_{\text{modelo}} = 26,55 + 0,73 * \text{RON}_{\text{CFR}} \text{ com } R^2 = 0,73.$$

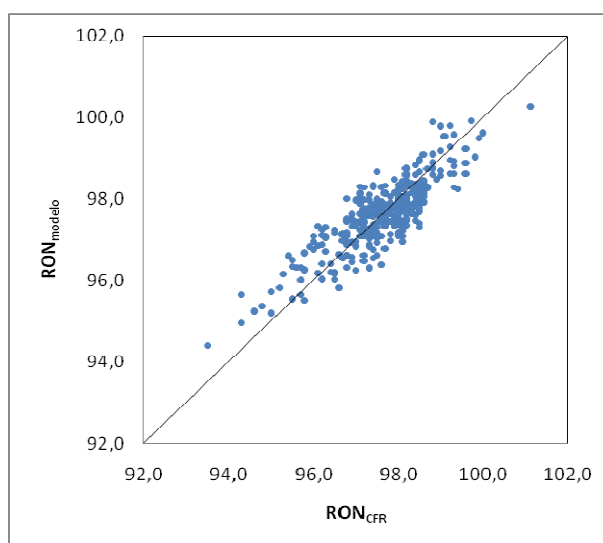


Figura 23- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 7 variáveis.

Na comparação das médias obtidas pelos dois métodos, de acordo com o teste *t* não se rejeita a hipótese nula, para um nível de significância de 0,05, uma vez que $t_{\text{calc}} < t_{\text{tab}}$, Tabela 26. Assim considera-se que as médias dos resultados obtidos pelos dois métodos não são significativamente diferentes.

Na comparação das variâncias, como $F_{\text{calc}} > F_{\text{tab}}$, para um nível de significância de 0,05, rejeita-se a hipótese nula, as variâncias obtidas pelos dois métodos

não são semelhantes. No entanto os resultados obtidos a partir do motor CFR apresentam uma variância superior aos obtidos pelo modelo, Tabela 26. Com este novo modelo as diferenças entre os resultados obtidos pelos dois métodos diminuem, incluindo a percentagem de diferenças superiores a $|0,7|$ que diminui para 20%, Figura 24.

Tabela 26- Resultados da comparação do RON_{CFR} e RON_{modelo} com 7 variáveis.

	RON_{CFR}	RON_{modelo}
Média	97,7	97,6
Desvio Padrão	1,00	0,86
Variância	1,01	0,73
n	355	355
F_{calc}		1,37
F_{tab}		0,89
$F_{calc} > F_{tab}$		
t_{calc}		0,25
t_{tab}		1,96
$t_{calc} < t_{tab}$		

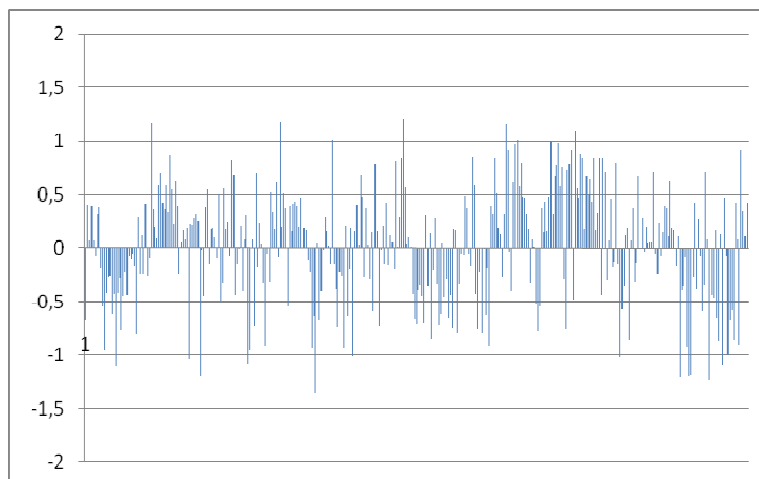


Figura 24- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} com 7 variáveis.

7.4 MODELO COM 47 VARIÁVEIS INDEPENDENTES

7.4.1 Análise das Variáveis

Neste novo modelo foram consideradas 47 variáveis, descritas na Tabela 27, sobre as quais foi calculada a média e o desvio padrão. No entanto as variáveis *n-Olefinas C₄* e *c-Olefinas C₁₀*, não foram consideradas uma vez que eram constantes com média e desvio padrão igual a zero. Assim o número de variáveis foi reduzido para 45. Do conjunto total das 375 amostras foram retiradas aquelas que apresentavam um comportamento atípico, depois da análise de resíduos, e consideradas 347 amostras para desenvolver o modelo pelo método *Stepwise*.

Tabela 27- Estatística descritiva da variável RON e das 47 variáveis independentes.

Variável	Média	Desvio Padrão	N	Variável	Média	Desvio Padrão	N
RON	97,614	1,0634	347	C nr 5 n olef	0,0204	0,01336	347
C nr 4 i par	0,1013	0,33766	347	C nr 6 n olef	0,1002	0,02660	347
C nr 5 i par	2,4249	0,68934	347	C nr 7 n olef	0,0892	0,05666	347
C nr 6 i par	8,2524	1,81958	347	C nr 8 n olef	0,0327	0,03003	347
C nr 7 i par	7,8521	0,88534	347	C nr 9 n olef	0,0085	0,01816	347
C nr 8 i par	2,6129	0,69266	347	C nr 10 n olef	0,0010	0,00585	347
C nr 9 i par	0,4509	0,31231	347	C nr 4 c olef	0,0003	0,00288	347
C nr 10i par	0,0719	0,08038	347	C nr 5 c olef	0,0029	0,01086	347
C nr 11 i par	0,0171	0,01846	347	C nr 6 c olef	0,0342	0,02240	347
C nr 4 n par	0,1386	0,33774	347	C nr 7 c olef	0,0090	0,01677	347
C nr 5 n par	1,4246	0,30561	347	C nr 8 c olef	0,0008	0,00688	347
C nr 6 n par	3,4793	0,56830	347	C nr 9 c olef	0,0001	0,00268	347
C nr 7 n par	2,1453	0,29535	347	C nr 10 c olef	0,0000	0,00000	347
C nr 8 n par	0,6313	0,22005	347	C nr 5 Naft	0,2029	0,10330	347
C nr 9 npar	0,2469	0,15880	347	C nr 6 Naft	0,5639	0,23788	347
C nr 10 npar	0,0076	0,02920	347	C nr 7 Naft	0,2743	0,09908	347
C nr 11 npar	0,0004	0,00282	347	C nr 8 Naft	0,1203	0,05644	347
C nr 4 i olef	0,0008	0,00970	347	C nr 9 Naft	0,0041	0,02267	347
C nr 5 i olef	0,0231	0,03687	347	C nr 10 Naft	0,0003	0,00591	347
C nr 6 i olef	0,2252	0,07468	347	C nr 6 Arom	6,3548	1,13838	347
C nr 7 i olef	0,3226	0,08254	347	C nr 7 Arom	18,0727	1,67272	347
C nr 8 i olef	0,1078	0,06537	347	C nr 8 Arom	23,6677	1,61909	347
C nr 9 i olef	0,0016	0,01399	347	C nr 9 Arom	16,4759	1,66360	347
C nr 10 i olef	0,0002	0,00322	347	C nr 10 Arom	2,7454	0,87844	347
C nr 4 n olef	0,0000	0,00000	347	C nr 10 Arom	2,7454	0,87844	347

De seguida foi analisada a correlação existente entre a variável dependente *RON* e cada uma das variáveis independentes assim como as correlações entre as variáveis independentes.

Verifica-se que a variável independente *Nafténicos C₆* é a que mais se correlaciona com a variável dependente *RON* (0,599), e as variáveis independentes têm uma correlação fraca entre si.

7.4.2 Estimação do Modelo

Os modelos foram desenvolvidos pelo método de selecção de regressores *Stepwise*, num total de 12. A primeira variável a ser introduzida no modelo (M_1) foi a *Nafténicos C₆*, uma vez que é a que apresenta maior correlação com a variável dependente *RON*. As restantes variáveis foram sendo introduzidas progressivamente no modelo de acordo com os princípios deste método de selecção de regressores.

A Tabela 28 mostra os ajustes obtidos para os vários modelos, sendo que o melhor ajuste é obtido para o modelo M_{12} com $r=0,933$; $R^2=0,870$ e $R_a^2=0,766$.

Tabela 28- Características dos modelos com 47 variáveis.

Modelo	r	R ²	R _a ²
M ₁	0,599	0,359	0,357
M ₂	0,716	0,513	0,510
M ₃	0,776	0,602	0,598
M ₄	0,860	0,739	0,736
M ₅	0,887	0,786	0,783
M ₆	0,906	0,822	0,818
M ₇	0,906	0,822	0,819
M ₈	0,916	0,839	0,837
M ₉	0,923	0,852	0,849
M ₁₀	0,929	0,862	0,859
M ₁₁	0,931	0,867	0,864
M ₁₂	0,933	0,870	0,866

O aproveitamento das potencialidades da cromatografia e conseqüente uso do número máximo de variáveis disponíveis (47) melhorou a qualidade do ajuste, o R_a^2 passou de 0,724 para 0,866, para o modelo com melhor ajuste, modelo M_{12} :

$$\begin{aligned}
 \text{RON} = & 93,190 + 0,174 * \text{AromáticosC}_8 - 1,217 * n - \text{ParafinasC}_7 - 0,551 * n - \\
 & \text{ParafinaC}_6 - 1,236 * i - \text{ParafinaC}_9 + 0,226 * \text{AromáticosC}_9 - 0,346 * n - \\
 & \text{ParafinasC}_5 - 50,624 * c - \text{OlefinasC}_9 + 2,222 * \\
 & i - \text{OlefinasC}_5 + 0,083 * \text{AromáticosC}_7 + 0,094 * \text{AromáticosC}_6
 \end{aligned}$$

De acordo com o teste *F*, Tabela 29, todos os modelos ajustados são globalmente significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Assim rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) [31].

Tabela 29- Resultados da ANOVA aos modelos com 47 variáveis.

Modelo	Soma dos quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	<i>p-value</i>	
M ₁	Regressão	140,305	1	140,305	192,874	0,000
	Residual	250,969	345	0,727		
	Total	391,274	346			
M ₂	Regressão	200,638	2	100,319	181,025	0,000
	Residual	190,635	344	0,554		
	Total	391,274	346			
M ₃	Regressão	235,446	3	78,482	172,751	0,000
	Residual	155,828	343	0,454		
	Total	391,274	346			
M ₄	Regressão	289,113	4	72,278	241,964	0,000
	Residual	102,160	342	0,299		
	Total	391,274	346			
M ₅	Regressão	307,676	5	61,535	251,006	0,000
	Residual	83,598	341	0,245		
	Total	391,274	346			
M ₆	Regressão	321,474	6	53,579	260,987	0,000
	Residual	69,800	340	0,205		
	Total	391,274	346			
M ₇	Regressão	321,459	5	64,292	314,026	0,000
	Residual	69,814	341	0,205		
	Total	391,274	346			
M ₈	Regressão	328,447	6	54,741	296,242	0,000
	Residual	62,827	340	0,185		
	Total	391,274	346			
M ₉	Regressão	333,485	7	47,641	279,469	0,000
	Residual	57,789	339	0,170		
	Total	391,274	346			
M ₁₀	Regressão	337,384	8	42,173	264,513	0,000
	Residual	53,890	338	0,159		
	Total	391,274	346			
M ₁₁	Regressão	339,376	9	37,708	244,862	0,000

M ₁₂	Residual	51,898	337	0,154	225,535	0,000
	Total	391,274	346			
	Regressão	340,540	10	34,054		
	Residual	50,733	336	0,151		
	Total	391,274	346			

Apesar de estarem disponíveis as 47 variáveis, este método de selecção de regressores seleccionou apenas 10 variáveis como explicativas do modelo que apresenta melhor ajuste M₁₂, Tabela 30:

Tabela 30- Coeficientes do modelo com melhor ajuste (M₁₂), com 47 variáveis.

Modelo	Coeficientes		t	p-value	Estatísticas de Colinearidade	
	B	Erro Padrão			Tolerância	VIF
M ₁₂ (Constante)	93,190	1,651	56,455	0,000		
C nr 8 Arom	0,174	0,021	8,198	0,000	0,368	2,715
C nr 7 n par	-1,217	0,150	-8,086	0,000	0,221	4,526
C nr 6 n par	-0,551	0,073	-7,557	0,000	0,255	3,928
C nr 9 i par	-1,236	0,108	-11,420	0,000	0,382	2,616
C nr 9 Arom	0,226	0,025	9,022	0,000	0,250	3,996
C nr 5 n par	-0,346	0,103	-3,378	0,001	0,444	2,251
C nr 9 c olef	-50,624	8,201	-6,173	0,000	0,901	1,110
C nr 5 i olef	2,222	0,616	3,610	0,000	0,847	1,180
C nr 7 Arom	0,083	0,026	3,236	0,001	0,236	4,243
C nr 6 Arom	0,094	0,034	2,777	0,006	0,296	3,374

7.4.3 Validação dos pressupostos do modelo estimado

7.4.3.1 Homocedasticidade

Os resíduos mantêm uma amplitude constante relativamente ao eixo horizontal zero, Figura 25, logo não se rejeita a hipótese da homocedasticidade.

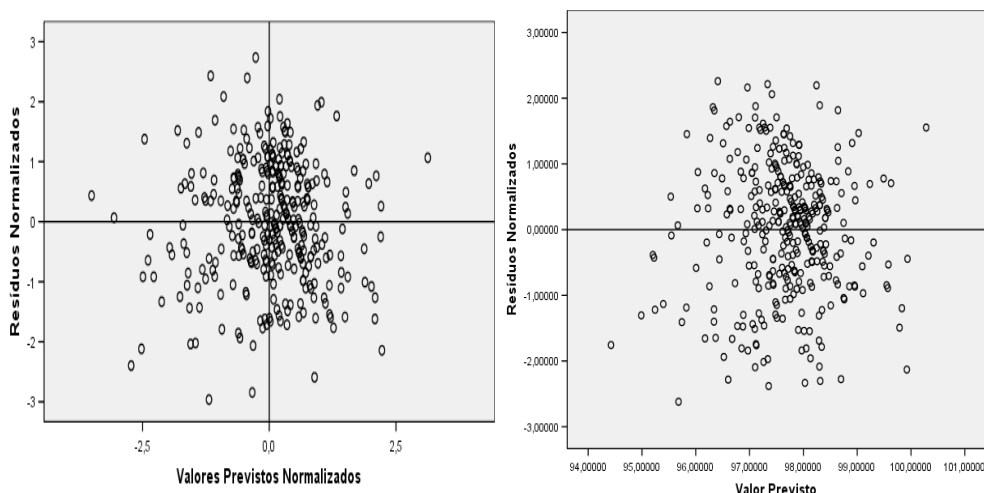


Figura 25- Avaliação da Homocedasticidade para o modelo M_{12} com 47 variáveis.

7.4.3.2 Covariância Nula

De acordo com o teste de Durbin-Watson o modelo M_{12} apresenta um valor de *teste* $d=1,438$, logo existe auto-correlação entre os resíduos.

7.4.3.3 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, Tabela 31.

Graficamente os valores esperados e observados normalizados dispõem-se ao longo da recta oblíqua; os desvios normalizados também se dispõem ao longo da recta horizontal, sendo que apresentam uma menor amplitude que nos modelos anteriores e uma menor tendência oscilatória, Figura 26.

Tabela 31- Teste à Normalidade dos resíduos, modelo M_{12} com 47 variáveis.

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,034	347	0,200

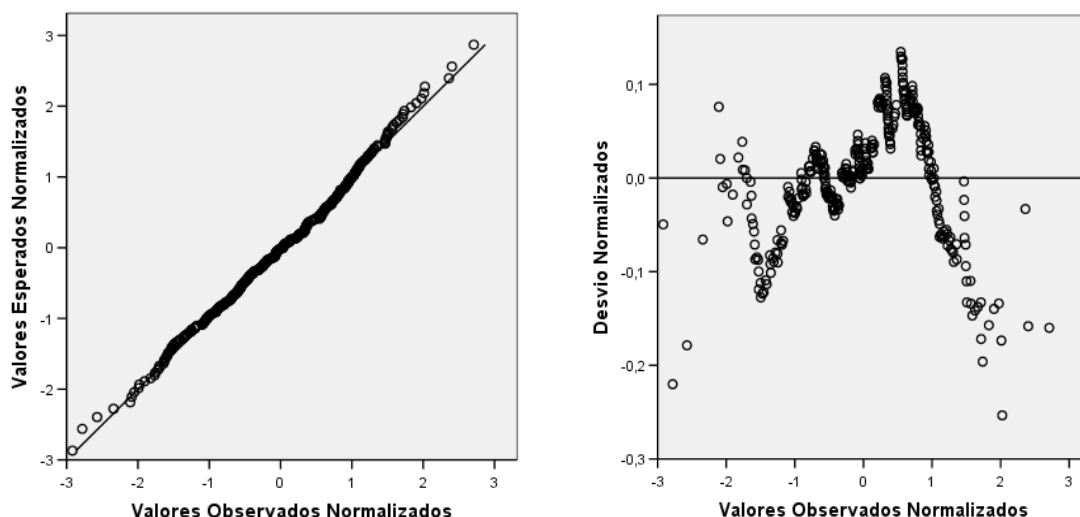


Figura 26- Análise gráfica da normalidade dos resíduos, para o modelo M_{12} com 47 variáveis.

7.4.3.4 Multicolinearidade

Não se verifica a existência de multicolinearidade entre as variáveis independentes. Esta análise é feita com base na matriz das correlações e nos valores da Tolerância e VIF, Tabela 30.

7.4.4 Previsão do RON usando o modelo

Com base no modelo desenvolvido é possível calcular o valor de RON, a partir da composição cromatográfica, para as amostras usadas na sua elaboração. Os valores obtidos são comparados com os obtidos pelo CFR através de uma regressão linear, Figura 27, e pelos *testes t e F*, Tabela 32.

Da comparação dos resultados por regressão linear, verifica-se que os resultados para o RON obtidos por este modelo são mais próximos dos obtidos pelo método do motor CFR, uma vez que a recta obtida se aproxima mais da relação ideal e apresenta um coeficiente de determinação superior:

$$RON_{\text{modelo}} = 12,67 + 0,87RON_{\text{CFR}} \text{ com } R^2 = 0,87.$$

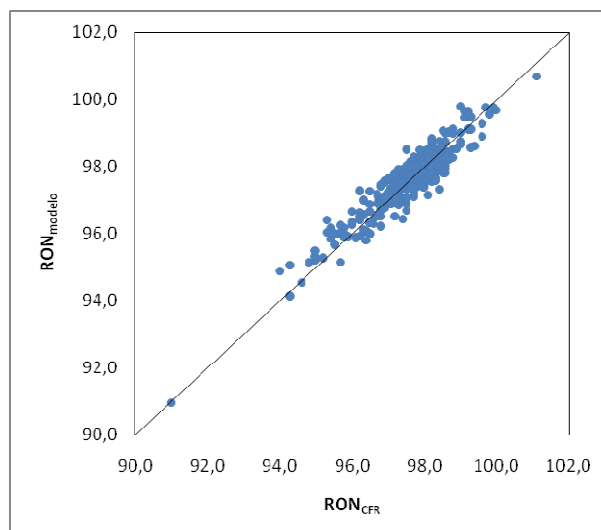


Figura 27- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis.

Dos *testes t e F* concluí-se que não existem desvios significativos entre os resultados obtidos por um ou outro método.

Relativamente às diferenças obtidas entre os dois métodos estas também diminuíram e a percentagem de resultados cuja diferença é superior a $|0,7|$ diminuiu drasticamente de 20% para 5% dos resultados, Figura 28.

Tabela 32- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis.

	RON_{CFR}	RON_{modelo}
Média	97,6	97,6
Desvio Padrão	1,06	0,99
Variância	1,13	0,98
n	347	347
F_{calc}		1,15
F_{tab}		1,27
$F_{calc} < F_{tab}$		
t_{calc}		0,24
t_{tab}		1,96
$t_{calc} < t_{tab}$		

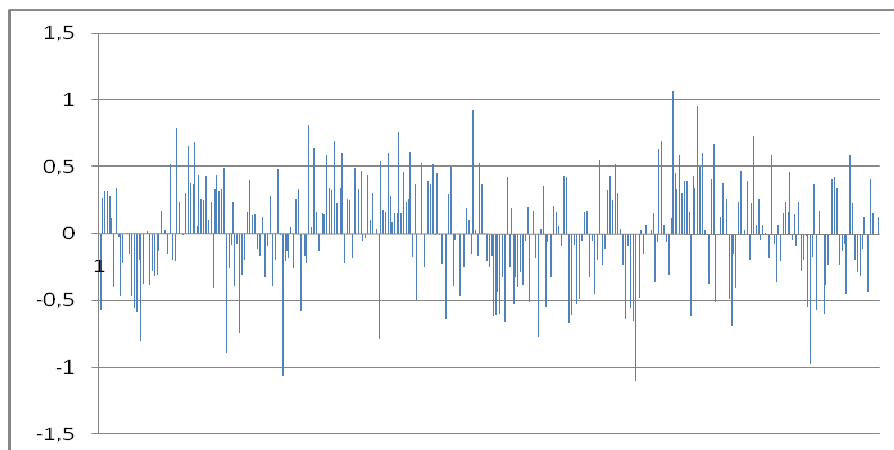


Figura 28- Diferença entre os resultados obtidos com o modelo e com o motor CFR, com 47 variáveis.

7.5 MODELO COM 47 VARIÁVEIS NORMALIZADAS

Uma vez que as concentrações médias de cada uma das 47 variáveis apresentam uma vasta amplitude, estas foram normalizadas antes de desenvolver um novo modelo.

Para isso foi calculada a média e o desvio padrão de cada uma das variáveis e estas foram transformadas em variáveis normalizadas Z_i , eq.27:

$$Z_i = \frac{X_i - \bar{X}}{\text{desv.pad}} \quad (\text{eq.27})$$

Foram excluídas as variáveis *n-Olefinas* C_4 e *n-Olefinas* C_{10} pelo facto de apresentarem desvio padrão igual a zero e não ser possível a sua normalização.

No entanto os resultados obtidos pelo modelo com as variáveis normalizadas foram iguais aos obtidos no ponto 7.4, sem as variáveis normalizadas. Uma vez que não se obtêm melhorias esta hipótese foi abandonada.

7.6 MODELO COM 47 VARIÁVEIS INDEPENDENTES – DADOS DE 2003

De acordo com os resultados obtidos nos modelos anteriores verifica-se que o melhor modelo é aquele para o qual foram consideradas as 47 variáveis, desenvolvido pelo método de selecção de regressores *Stepwise*.

A fim de validar os modelos e verificar se existem variações ao longo do tempo, os dados foram divididos em grupos e criados modelos a partir desses grupos. O primeiro modelo foi desenvolvido a partir dos dados do ano de 2003, e aplicado às restantes amostras (2004 a 2006), a fim de verificar a resposta do modelo a amostras que não foram usadas na sua elaboração.

7.6.1 Análise das Variáveis

Neste novo modelo foram consideradas 47 variáveis, sobre as quais foi calculada a média e o desvio padrão, Tabela 33. As variáveis *n-Olefinas C₄* e *c-Olefinas C₁₀*, foram rejeitadas uma vez que eram constantes com média e desvio padrão igual a zero, logo o número de variáveis passou para 45. Do conjunto total de 123 amostras, foram retiradas as amostras atípicas e consideradas 112 amostras para desenvolver o modelo.

Neste caso, considerando as amostras apenas de 2003, a variável *n-Parafinas C₉* é aquela que apresenta uma correlação mais forte com o RON e por isso a primeira a entrar no modelo.

Tabela 33- Estatística descritiva da variável RON e das 47 variáveis independentes (2003).

	Média	Desvio Padrão	N		Média	Desvio Padrão	N
RON	97,904	1,1387	112	C nr 5 n olef	0,0273	0,00968	112
C nr 4 i par	0,0584	0,19876	112	C nr 6 n olef	0,1038	0,02476	112
C nr 5 i par	2,2688	0,36628	112	C nr 7 n olef	0,0920	0,02760	112
C nr 6 i par	7,3221	1,03483	112	C nr 8 n olef	0,0326	0,02003	112
C nr 7 i par	7,4436	0,79744	112	C nr 9 n olef	0,0049	0,01230	112
C nr 8 i par	2,6046	0,69533	112	C nr 10 n olef	0,0007	0,00439	112
C nr 9 i par	0,4732	0,26530	112	C nr 4 c olef	0,0004	0,00378	112
C nr 10i par	0,0818	0,06900	112	C nr 5 c olef	0,0032	0,01015	112
C nr 11 i par	0,0231	0,01617	112	C nr 6 c olef	0,0388	0,01693	112
C nr 4 n par	0,1504	0,34447	112	C nr 7 c olef	0,0192	0,02023	112
C nr 5 n par	1,3689	0,22436	112	C nr 8 c olef	0,0023	0,01185	112
C nr 6 n par	3,2740	0,46438	112	C nr 9 c olef	0,0004	0,00472	112
C nr 7 n par	2,1475	0,30998	112	C nr 10 c olef	0,0000	0,00000	112
C nr 8 n par	0,6240	0,21690	112	C nr 5 Naft	0,1703	0,07399	112
C nr 9 npar	0,2589	0,09137	112	C nr 6 Naft	0,5188	0,14579	112
C nr 10 npar	0,0080	0,02650	112	C nr 7 Naft	0,2579	0,06738	112

C nr 11 npar	0,0014	0,00517	112	C nr 8 Naft	0,1249	0,04289	112
C nr 4 i olef	0,0011	0,00798	112	C nr 9 Naft	0,0046	0,02620	112
C nr 5 i olef	0,0390	0,04118	112	C nr 10 Naft	0,0010	0,01039	112
C nr 6 i olef	0,2711	0,05582	112	C nr 6 Arom	6,7641	1,13211	112
C nr 7 i olef	0,3721	0,06739	112	C nr 7 Arom	18,2303	1,77145	112
C nr 8 i olef	0,1395	0,04891	112	C nr 8 Arom	23,4730	1,32037	112
C nr 9 i olef	0,0035	0,02249	112	C nr 9 Arom	17,1788	1,48207	112
C nr 10 i olef	0,0005	0,00567	112	C nr 10 Arom	3,3575	0,86683	112
Cnr 4 n olef	0,0000	0,0000	112				

7.6.2 Estimação do Modelo

O método *Stepwise* desenvolveu um total de 17 modelos até que todas as variáveis não introduzidas no modelo não tivessem capacidade explicativa da variável dependente RON e todas as que estão no modelo o tivessem [29].

A Tabela 34 mostra os ajustes obtidos para os vários modelos; o melhor ajuste corresponde ao modelo M_{17} , com $r=0,974$; $R^2=0,949$ e $R_a^2=0,942$.

Embora se parta de um conjunto inicial de 45 variáveis o modelo obtido por este método usa apenas 13, Tabela 35. Algumas destas variáveis são comuns às consideradas como significativas para o modelo anterior (ponto 7.4) no entanto verifica-se uma forte presença das variáveis Nafténicos, não incluídas no último modelo M_{17} :

$$\begin{aligned} \text{RON} = & 98,292 - 4,249 * i\text{-Parafinas}C_8 - 0,313 * i\text{-} \\ & \text{Parafinas}C_6 + 0,241 * \text{Aromáticos}C_9 + 0,089 * \text{Aromáticos}C_7 + 6,232 * i\text{-} \\ & \text{Parafinas}C_{10} + 1,533 * \text{Nafténicos}C_5 - 0,970 * n\text{-Parafinas}C_5 - 1,838 * n\text{-} \\ & \text{Parafinas}C_7 - 24,662 * \text{Nafténicos}C_{10} + 7,362 * \text{Nafténicos}C_9 + 0,320 * i\text{-} \\ & \text{Parafinas}C_7 + 9,864 * n\text{-Olefinas}C_5 - 0,326 * i\text{-Parafinas}C_4 \end{aligned}$$

Tabela 34- Características dos modelos com 47 variáveis (2003).

Modelo	r	R ²	R _a ²
M ₁	0,778	0,606	0,602
M ₂	0,886	0,785	0,781
M ₃	0,928	0,860	0,857
M ₄	0,940	0,884	0,880
M ₅	0,948	0,899	0,894
M ₆	0,955	0,912	0,907
M ₇	0,958	0,917	0,912
M ₈	0,957	0,915	0,910

M ₉	0,959	0,919	0,913
M ₁₀	0,960	0,922	0,916
M ₁₁	0,963	0,928	0,922
M ₁₂	0,966	0,934	0,927
M ₁₃	0,968	0,938	0,931
M ₁₄	0,968	0,937	0,931
M ₁₅	0,970	0,941	0,935
M ₁₆	0,973	0,946	0,940
M ₁₇	0,974	0,949	0,942

O modelo M₁₇ é significativo na globalidade, logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para o modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31]. A Tabela 35 apresenta os coeficientes do modelo M₁₇, o seu desvio padrão, os testes *t*, e as tolerâncias de cada variável. Os testes *t* permitem testar a hipótese nula de inexistência de uma relação linear entre a variável dependente RON com cada uma das variáveis independentes, bem como com as constantes consideradas no modelo [31]. Verifica-se que todos os testes *t* têm associado um nível de significância inferior a 0,05, excepto o do coeficiente associado ao regressor *Aromáticos C₇*; este último é justificado pelo facto deste coeficiente apresentar um valor muito próximo de zero (0,089). O resultado destes testes permitem-nos concluir que todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes de cada variável independente são diferentes de zero, mesmo o associado ao regressor *Aromáticos C₇* [31].

Tabela 35- Coeficientes do modelo com melhor ajuste M₁₇, com 47 variáveis (2003).

Modelo	Coeficientes		t	p-value	Estatísticas de Colinearidade	
	B	Erro Padrão			Tolerância	VIF
M ₁₇ (Constante)	98,292	2,312	42,508	0,000		
C nr 9 i par	-4,249	0,246	-	0,000	0,159	6,293
C nr 6 i par	-0,313	0,055	-5,716	0,000	0,210	4,756
C nr 9 Arom	0,241	0,046	5,212	0,000	0,143	6,989
C nr 7 Arom	0,089	0,048	1,870	0,065	0,094	10,641
C nr 10i par	6,232	0,756	8,243	0,000	0,248	4,034
C nr 5 Naft	1,533	0,454	3,377	0,001	0,598	1,673
C nr 5 n par	-0,970	0,161	-6,030	0,000	0,518	1,930

C nr 7 n par	-1,838	0,351	-5,237	0,000	0,057	17,555
C nr 10 Naft	-	4,534	-5,440	0,000	0,304	3,292
C nr 9 Naft	7,362	1,923	3,829	0,000	0,266	3,761
C nr 7 i par	0,320	0,100	3,201	0,002	0,106	9,399
C nr 5 n olef	9,864	3,004	3,284	0,001	0,798	1,253
C nr 4 i par	-0,326	0,141	-2,317	0,023	0,863	1,158

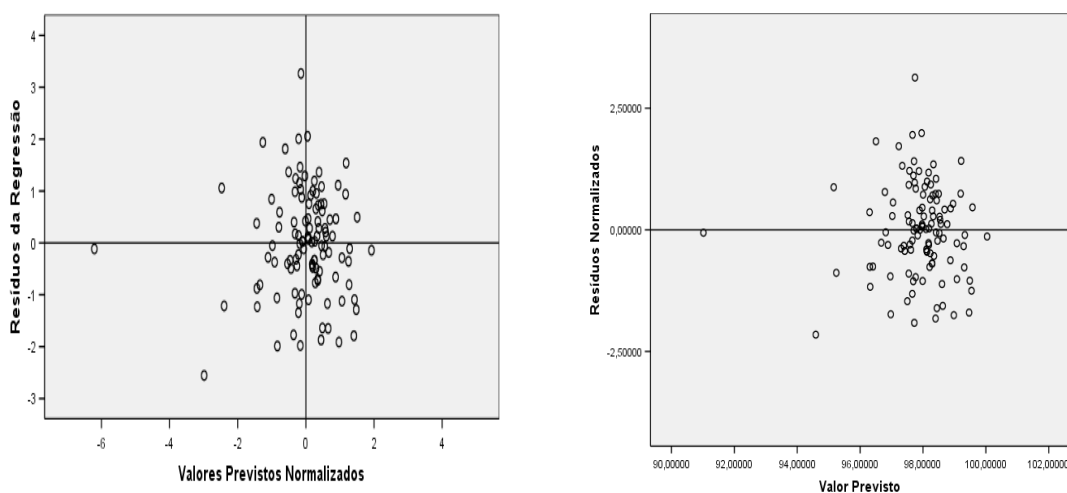
7.6.3 Validação dos pressupostos do modelo

7.6.3 Homocedasticidade

Os resíduos não apresentam tendências crescentes ou decrescentes, pelo que não se rejeita a hipótese da homocedasticidade, [Figura 29](#).

7.6.3.1 Covariância Nula

De acordo com o teste de Durbin-Watson, $d=2,184$, não existe auto-correlação entre os resíduos.



[Figura 29](#)- Avaliação da Homocedasticidade para o modelo M_{17} com 47 variáveis (2003).

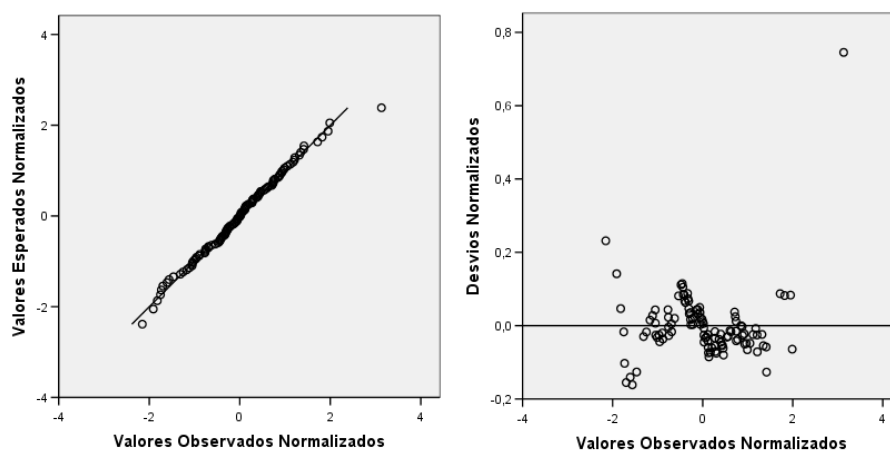
7.6.3.2 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, [Tabela 36](#); esta hipótese é

confirmada graficamente uma vez que as observações se dispõem ao longo das rectas oblíqua e horizontal, [Figura 30](#).

[Tabela 36](#)- Teste à Normalidade dos resíduos, modelo M_{17} com 47 variáveis (2003).

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,047	112	0,200



[Figura 30](#)- Análise gráfica da normalidade dos resíduos, para o modelo M_{17} , com 47 variáveis (2003).

7.6.3.3 Multicolinearidade

Neste modelo verifica-se a existência de multicolinearidade para as variáveis independentes com 7 átomos de carbono (*i-Parafinas C₇*, *n-Parafinas C₇* e *Aromáticos C₇*), confirmado pelos valores da Tolerância e do VIF, [Tabela 35](#).

7.6.4 Previsão do RON usando o modelo

O modelo desenvolvido foi aplicado nas amostras de 2004 a 2006. O objectivo é verificar a resposta do mesmo quando aplicado a amostras que não contribuíram para o seu desenvolvimento, [Figura 31](#).

Embora este modelo seja o que o apresenta melhor ajuste, dos desenvolvidos até aqui, verifica-se que quando aplicado a outras amostras existe uma grande dispersão de resultados e uma diferença muito elevada entre o RON_{CFR} e RON_{modelo} , [Figura 32](#); em 43% dessas amostras a diferença é superior a $|0,7|$.

No entanto quando comparados os resultados obtidos pelos dois métodos pelos testes F e t , não são encontradas diferenças significativas nas variâncias, nem nas médias, respectivamente. Concluí-se portanto que não existem desvios significativos entre os resultados obtidos pelos dois métodos, Tabela 37.

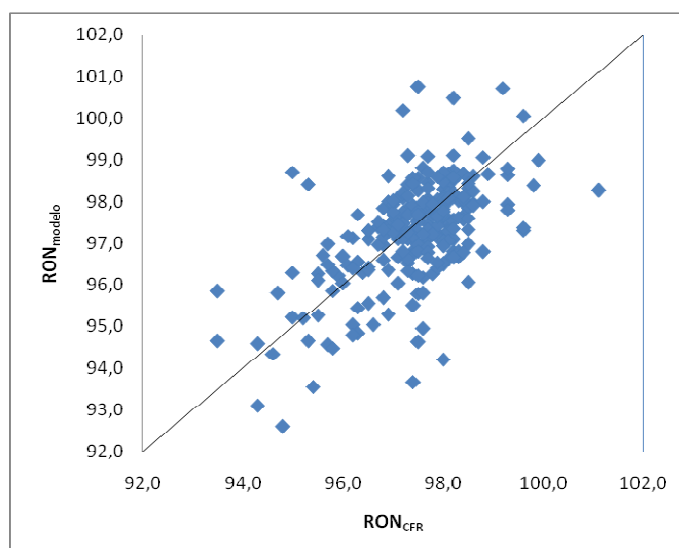


Figura 31- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis (2003).

Tabela 37- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis (2003).

	RON_{CFR}	RON_{modelo}
Média	97,4	97,3
Desvio Padrão	1,09	1,24
Variância	1,19	1,53
n	252	252
F_{calc}		0,78
F_{tab}		1,18
$F_{calc} < F_{tab}$		
t_{calc}		0,98
t_{tab}		1,96
$t_{calc} < t_{tab}$		

É de salientar que a diferença aumenta com o afastamento temporal das amostras, às quais é aplicado o modelo, relativamente às que foram usadas na sua elaboração (amostras de 2003). O que pode ser resultado de pequenas variação da matriz ao longo do tempo, Figura 32.

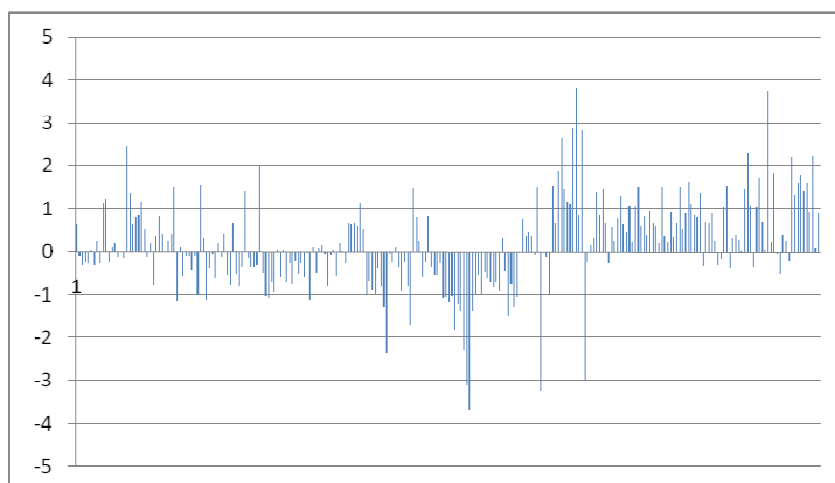


Figura 32- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} , quando aplicado às amostras de 2004 a 2006.

7.7 MODELO COM 47 VARIÁVEIS COM OS DADOS DE 2003 A 2004

7.7.1 Análise das Variáveis

Neste modelo foram consideradas novamente 47 variáveis, descritas na Tabela 38, sobre as quais foi calculada a média e o desvio padrão. As variáveis *n-Olefinas* C_4 , *c-Olefinas* C_9 e *c-Olefinas* C_{10} , foram excluídas por serem constantes com média e desvio padrão igual a zero. Do conjunto total foram retiradas as amostras atípicas e consideradas 252 amostras para desenvolver o modelo.

7.7.2 Estimação do Modelo

O modelo foi desenvolvido novamente pelo método de selecção de regressores *Stepwise*, em que a primeira variável a ser introduzida foi a *Nafténicos* C_6 ,

devido à forte correlação com a variável dependente RON. É de notar que esta também foi a primeira variável a ser introduzida, quando do desenvolvimento do primeiro modelo com 47 variáveis.

O melhor ajuste corresponde ao modelo M_{15} , com $r=0,925$, $R^2=0,856$ e $R_a^2=0,849$, Tabela 39. Este modelo escolheu como regressores 11 das 45 variáveis independentes inicialmente consideradas, modelo M_{15} :

$$RON = 106,119 - 1,071 * i\text{-Nafténicos}C_6 - 0,247 * i\text{-Parafinas}C_7 - 0,517 * n\text{-Parafinas}C_5 - 0,829 * n\text{-Parafinas}C_7 - 0,930 * n\text{-Parafinas}C_6 - 7,897 * n\text{-Parafinas}C_{10} + 0,056 * \text{Aromáticos}C_9 - 2,755 * i\text{-Olefinas}C_8 - 0,279 * n\text{-Parafinas}C_4 - 1,088 * n\text{-Parafinas}C_8 - 19,420 * n\text{-Parafinas}C_{11}$$

O número de regressores e os próprios regressores são diferentes dos usados nos modelos anteriores, também com 47 variáveis, Tabela 40.

O modelo é significativo em termos globais, logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

Tabela 38- Estatística descritiva das 47 variáveis (2003 a 2004).

	Média	Desvio Padrão	N		Média	Desvio Padrão	N
RON	97,608	1,0381	252	C nr 4 n olef	0,0000	0,00000	252
C nr 4 i par	0,0856	0,30456	252	C nr 5 n olef	0,0235	0,01233	252
C nr 5 i par	2,2975	0,48411	252	C nr 6 n olef	0,1043	0,02474	252
C nr 6 i par	7,7050	1,61151	252	C nr 7 n olef	0,0934	0,03094	252
C nr 7 i par	7,8151	0,82991	252	C nr 8 n olef	0,0346	0,02508	252
C nr 8 i par	2,6777	0,64169	252	C nr 9 n olef	0,0112	0,02093	252
C nr 9 i par	0,4806	0,29410	252	C nr 10 n olef	0,0013	0,00670	252
C nr 10i par	0,0908	0,07705	252	C nr 4 c olef	0,0004	0,00338	252
C nr 11 i par	0,0228	0,01826	252	C nr 5 c olef	0,0035	0,01110	252
C nr 4 n par	0,1454	0,35733	252	C nr 6 c olef	0,0403	0,01902	252
C nr 5 n par	1,4328	0,31781	252	C nr 7 c olef	0,0117	0,01787	252
C nr 6 n par	3,3933	0,57146	252	C nr 8 c olef	0,0007	0,00621	252
C nr 7 n par	2,1642	0,27820	252	C nr 9 c olef	0,0000	0,00000	252
C nr 8 n par	0,6483	0,19463	252	C nr 10 c olef	0,0000	0,00000	252
C nr 9 npar	0,2937	0,13861	252	C nr 5 Naft	0,2042	0,11099	252
C nr 10 npar	0,0077	0,01873	252	C nr 6 Naft	0,6150	0,21978	252
C nr 11 npar	0,0006	0,00351	252	C nr 7 Naft	0,2885	0,09135	252

C nr 4 i olef	0,0011	0,01138	252	C nr 8 Naft	0,1294	0,04294	252
C nr 5 i olef	0,0292	0,04344	252	C nr 9 Naft	0,0040	0,02216	252
C nr 6 i olef	0,2409	0,06500	252	C nr 10 Naft	0,0004	0,00693	252
C nr 7 i olef	0,3526	0,06000	252	C nr 6 Arom	6,3700	1,15233	252
C nr 8 i olef	0,1278	0,05759	252	C nr 7 Arom	18,1644	1,70049	252
C nr 9 i olef	0,0023	0,01665	252	C nr 8 Arom	23,3187	1,51487	252
C nr 10 i olef	0,0002	0,00378	252	C nr 9 Arom	16,7770	1,67006	252
				C nr 10 Arom	3,0210	0,85587	252

Tabela 39- Características dos modelos com 47 variáveis (2003 a 2004).

Modelo	r	R ²	R _a ²
M ₁	0,741	0,548	0,547
M ₂	0,816	0,665	0,662
M ₃	0,848	0,720	0,716
M ₄	0,861	0,741	0,737
M ₅	0,878	0,772	0,767
M ₆	0,903	0,815	0,811
M ₇	0,902	0,814	0,811
M ₈	0,908	0,825	0,820
M ₉	0,907	0,823	0,820
M ₁₀	0,912	0,832	0,828
M ₁₁	0,915	0,837	0,832
M ₁₂	0,918	0,843	0,838
M ₁₃	0,921	0,849	0,843
M ₁₄	0,923	0,853	0,846
M ₁₅	0,925	0,856	0,849

De acordo com o teste *t*, Tabela 40, todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes das variáveis independente são diferentes de zero, para o modelo M₁₅ [31].

Tabela 40- Coeficientes do modelo com melhor ajuste M₁₅, com 47 variáveis (2003 a 2004).

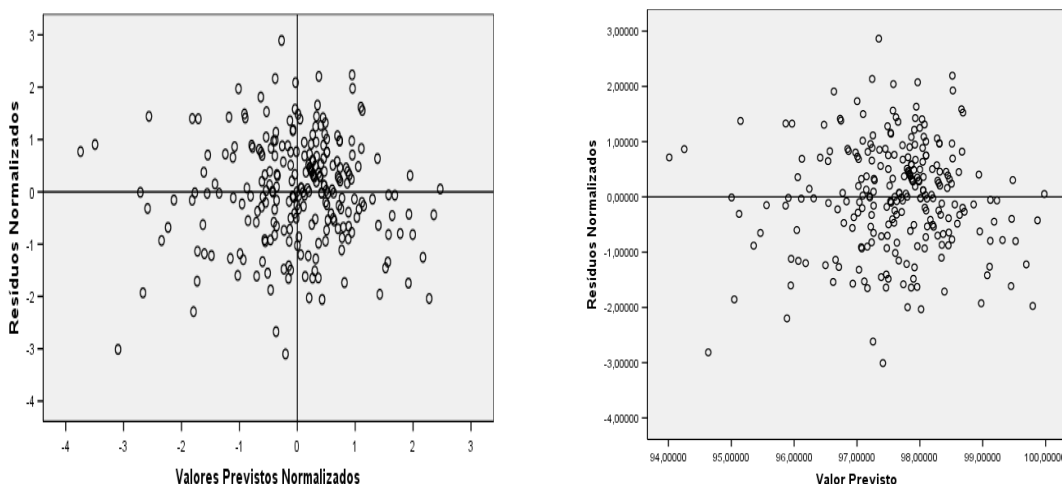
Modelo	Coeficientes		t	p-value	Estatísticas de Colinearidade		
	B	Erro Padrão			Tolerancia	VIF	
M ₁₅	Constante	106,119	0,506	209,839	0,000		
	C nr 6 Naft	-1,071	0,185	-5,795	0,000	0,392	2,551
	C nr 7 i par	-0,247	0,055	-4,525	0,000	0,314	3,186
	C nr 5 n par	-0,517	0,097	-5,334	0,000	0,680	1,471
	C nr 7 n par	-0,829	0,300	-2,768	0,006	0,093	10,744
	C nr 6 n par	-0,930	0,065	-14,387	0,000	0,474	2,112
	C nr 10 npar	-7,897	1,846	-4,278	0,000	0,541	1,849
	C nr 9 Arom	0,056	0,020	2,804	0,005	0,576	1,736
	C nr 8 i olef	-2,755	0,617	-4,467	0,000	0,512	1,952
	C nr 4 n par	-0,279	0,076	-3,651	0,000	0,869	1,151

C nr 8 n par	-1,088	0,394	-2,759	0,006	0,110	9,115
C nr 11 npar	-19,420	7,994	-2,429	0,016	0,819	1,220

7.7.3 Diagnóstico do Modelo Estimado

7.7.3.1 Homocedasticidade

De acordo com a distribuição dos resíduos, [Figura 33](#), não se rejeita a hipótese da homocedasticidade.



[Figura 33](#) – Avaliação da Homocedasticidade para o modelo M_{15} , com 47 variáveis (2003 a 2004).

7.7.3.1 Covariância Nula

De acordo com o teste de Durbin-Watson o modelo M_{15} apresenta um valor de teste $d=1,646$, logo existe auto-correlação entre os resíduos.

7.7.3.2 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, [Tabela 41](#); esta hipótese é confirmada graficamente uma vez que as observações se dispõem ao longo das rectas oblíqua e horizontal, [Figura 34](#).

[Tabela 41](#)- Teste à Normalidade dos resíduos, modelo M_{15} , com 47 variáveis (2003 a 2004).

	Kolmogorov-Smirnov		
	Estatística	Graus	Significância

		Liberdade	
Resíduos Padrão	0,047	252	0,200

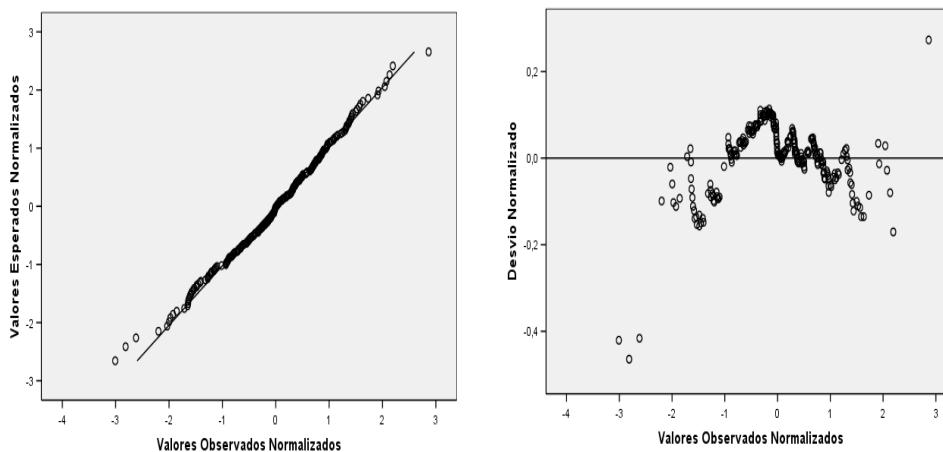


Figura 34- Análise gráfica da normalidade dos resíduos, para o modelo M_{15} com 47 variáveis (2003 a 2004).

7.7.3.4 Multicolinearidade

Não se verifica a existência de multicolinearidade entre as variáveis independentes. Esta análise é feita com base na matriz das correlações e nos valores da Tolerância e VIF, Tabela 40.

7.7.4 Previsão do RON usando o modelo

O modelo desenvolvido neste ponto foi aplicado nas amostras de 2005 a 2006, com o objectivo de verificar a resposta do modelo quando aplicado em amostras independentes ao seu desenvolvimento.

Embora o modelo apresente um bom ajuste quando aplicado a outras amostras apresenta uma grande dispersão de resultados, tal como no modelo anterior. Há uma diferença grande entre o RON_{CFR} e RON_{modelo} , Figura 35.

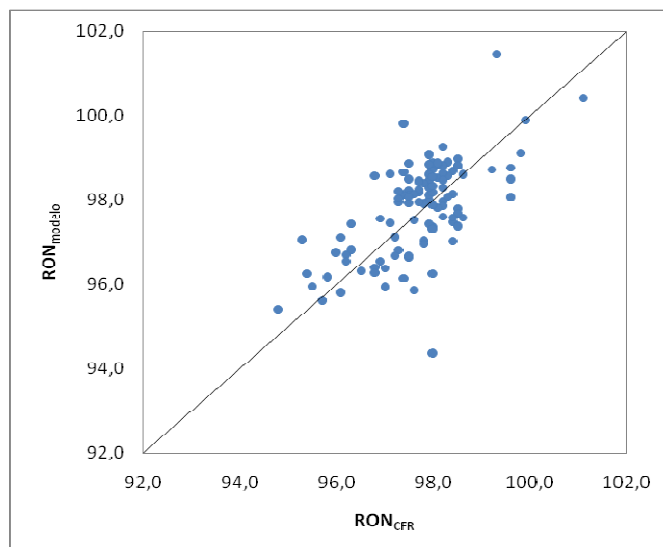


Figura 35- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis (2003 a 2004).

Tabela 42- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis (2003 a 2004).

	RON_{CFR}	RON_{modelo}
Média	97,7	97,7
Desvio Padrão	0,98	1,72
Variância	0,97	2,95
n	112	112
F _{calc}		1,15
F _{tab}		1,15
$F_{calc} = F_{tab}$		
t _{calc}		0,003
t _{tab}		1,96
$T_{calc} < t_{tab}$		

No entanto de acordo com o teste *t* das médias concluí-se que não existem desvios significativos entre os resultados obtidos por um ou outro método, Tabela 42. Relativamente às diferenças obtidas entre os dois métodos mantêm-se relativamente ao método desenvolvido apenas com os dados de 2003. A percentagem de resultados cuja diferença é superior a |0,7| manteve-se em 43% dos resultados, Figura 36.

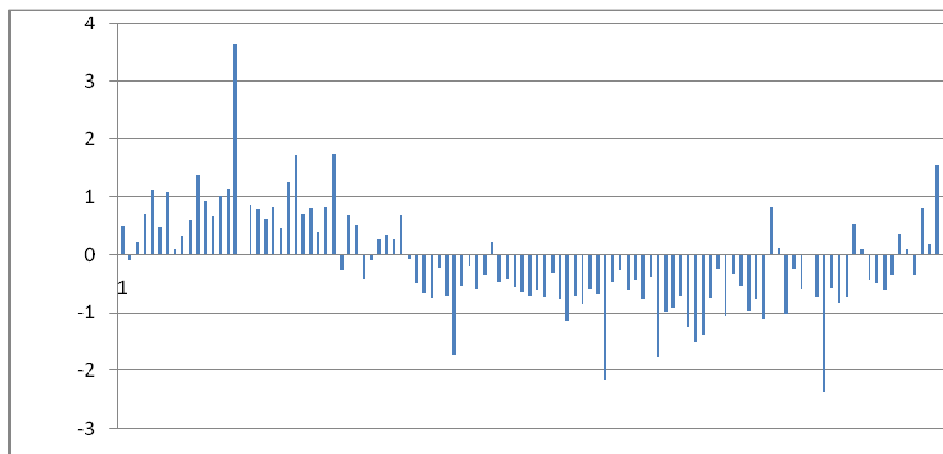


Figura 36- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} com 47 variáveis (2003 a 2004).

7.8 MODELO COM 47 VARIÁVEIS COM OS DADOS DE 2003 A 2005

7.8.1 Análise das Variáveis Independentes

Neste modelo foram consideradas novamente 47 variáveis, descritas na [Tabela 43](#), sobre as quais foi calculada a média e o desvio padrão. As variáveis *n-Olefinas* C_4 , *c-Olefinas* C_9 e *c-Olefinas* C_{10} , foram excluídas por serem constantes com média e desvio padrão igual a zero. Do conjunto total foram retiradas as amostras atípicas e consideradas 334 amostras para desenvolver o modelo.

7.8.2 Estimação do Modelo

O modelo foi desenvolvido novamente pelo método de selecção de regressores *Stepwise*, em que a primeira variável a ser introduzida foi a *Nafténicos* C_6 , devido à forte correlação com a variável dependente RON.

Tabela 43- Estatística descritiva das 47 variáveis (2003 a 2005).

	Média	Desvio Padrão	N		Média	Desvio Padrão	N
RON	97,667	0,9895	334	C nr 5 n olef	0,0203	0,01368	334
C nr 4 i par	0,0895	0,31036	334	C nr 6 n olef	0,1007	0,02626	334
C nr 5 i par	2,4061	0,62214	334	C nr 7 n olef	0,0887	0,03739	334
C nr 6 i par	8,2501	1,86453	334	C nr 8 n olef	0,0332	0,03026	334
C nr 7 i par	7,8428	0,80981	334	C nr 9 n olef	0,0093	0,01896	334
C nr 8 i par	2,5697	0,64926	334	C nr 10 n olef	0,0011	0,00596	334
C nr 9 i par	0,4323	0,29325	334	C nr 4 c olef	0,0003	0,00294	334

C nr 10i par	0,0742	0,08071	334	C nr 5 c olef	0,0026	0,00975	334
C nr 11 i par	0,0176	0,01859	334	C nr 6 c olef	0,0346	0,02197	334
C nr 4 n par	0,1435	0,34985	334	C nr 7 c olef	0,0089	0,01635	334
C nr 5 n par	1,4469	0,35722	334	C nr 8 c olef	0,0006	0,00550	334
C nr 6 n par	3,4405	0,60413	334	C nr 9 c olef	0,0000	0,00000	334
C nr 7 n par	2,1210	0,30593	334	C nr 10 c olef	0,0000	0,00000	334
C nr 8 n par	0,6327	0,29128	334	C nr 5 Naft	0,2067	0,10584	334
C nr 9 npar	0,2561	0,19210	334	C nr 6 Naft	0,5636	0,23654	334
C nr 10 npar	0,0090	0,04254	334	C nr 7 Naft	0,2745	0,09488	334
C nr 11 npar	0,0019	0,02696	334	C nr 8 Naft	0,1175	0,05043	334
C nr 4 i olef	0,0008	0,00989	334	C nr 9 Naft	0,0030	0,01932	334
C nr 5 i olef	0,0222	0,03973	334	C nr 10 Naft	0,0003	0,00602	334
C nr 6 i olef	0,2263	0,07250	334	C nr 6 Arom	6,3404	1,14254	334
C nr 7 i olef	0,3282	0,08070	334	C nr 7 Arom	18,0660	1,71770	334
C nr 8 i olef	0,1093	0,06567	334	C nr 8 Arom	23,6535	1,61433	334
C nr 9 i olef	0,0019	0,01455	334	C nr 9 Arom	16,5595	1,64405	334
C nr 10 i olef	0,0002	0,00328	334	C nr 10 Arom	2,7854	0,88023	334
C nr 4 n olef	0,0000	0,00000	334				

O melhor ajuste corresponde ao modelo M_{20} , com $r=0,900$, $R^2=0,811$ e $R_a^2=0,802$, Tabela 44. Este modelo escolheu como regressores 14 das 45 variáveis inicialmente consideradas. O número de regressores e os próprios regressores são diferentes dos usados nos modelos anteriores, também com 47 variáveis, Tabela 45.

O modelo é significativo em termos globais, logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

De acordo com o *teste t*, Tabela 45, todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes das variáveis independentes são diferentes de zero, para o modelo M_{20} [31].

Tabela 44- Características dos modelos com 47 variáveis (2003 a 2005).

Modelo	r	R^2	R^2 Ajustado
M_1	0,622	0,387	0,385
M_2	0,731	0,534	0,531
M_3	0,778	0,606	0,602
M_4	0,834	0,696	0,692
M_5	0,848	0,719	0,715

M ₆	0,866	0,750	0,745
M ₇	0,873	0,762	0,757
M ₈	0,880	0,774	0,769
M ₉	0,883	0,779	0,773
M ₁₀	0,885	0,783	0,776
M ₁₁	0,888	0,788	0,781
M ₁₂	0,892	0,795	0,787
M ₁₃	0,891	0,794	0,787
M ₁₄	0,894	0,800	0,792
M ₁₅	0,897	0,804	0,796
M ₁₆	0,898	0,807	0,798
M ₁₇	0,900	0,809	0,800
M ₁₈	0,899	0,809	0,800
M ₁₉	0,899	0,808	0,800
M ₂₀	0,900	0,811	0,802

Tabela 45- Coeficientes do modelo com melhor ajuste M₂₀, com 47 variáveis (2003 a 2005).

Modelo	Coeficientes		t	p-value	Estatística da Colinearidade		
	B	Erro Padrão			Tolerância	VIF	
M ₂₀	Constante	93,740	1,607	58,324	0,000		
	C nr 6 Naft	-0,600	0,201	-2,987	0,003	0,258	3,881
	C nr 7 n par	-1,010	0,179	-5,647	0,000	0,194	5,152
	C nr 6 n par	-0,248	0,086	-2,863	0,004	0,213	4,697
	C nr 11 npar	-5,040	1,276	-3,950	0,000	0,491	2,036
	C nr 9 i par	-0,748	0,207	-3,609	0,000	0,158	6,349
	C nr 5 n par	-0,494	0,102	-4,860	0,000	0,440	2,271
	C nr 5 i olef	2,915	0,690	4,225	0,000	0,774	1,292
	C nr 8 Arom	0,135	0,021	6,307	0,000	0,485	2,063
	C nr 9 Arom	0,217	0,030	7,134	0,000	0,232	4,314
	C nr 8 i par	-0,342	0,118	-2,894	0,004	0,099	10,138
	C nr 6 n olef	-5,101	1,639	-3,112	0,002	0,314	3,187
	C nr 7 i olef	1,218	0,476	2,558	0,011	0,394	2,538
	C nr 7 Arom	0,122	0,029	4,221	0,000	0,235	4,256
	C nr 8 Naft	1,959	0,978	2,002	0,046	0,239	4,186

$$\begin{aligned}
 \text{RON} = & 106,119 - 1,071 * i\text{-Nafténicos} C_6 - 0,247 * i\text{-Parafinas} C_7 - 0,517 * n\text{-} \\
 & \text{Parafinas} C_5 - 0,829 * n\text{-Parafinas} C_7 - 0,930 * n\text{-Parafinas} C_6 - 7,897 * n\text{-} \\
 & \text{Parafinas} C_{10} + 0,056 * \text{Aromáticos} C_9 - 2,755 * i\text{-Olefinas} C_8 - 0,279 * n\text{-} \\
 & \text{Parafinas} C_4 - 1,088 * n\text{-Parafinas} C_8 - 19,420 * n\text{-Parafinas} C_{11}
 \end{aligned}$$

7.8.3 Diagnóstico do Modelo Estimado

7.8.3.1 Homocedasticidade

De acordo com a distribuição dos resíduos, Figura 37, não se rejeita a hipótese da homocedasticidade.

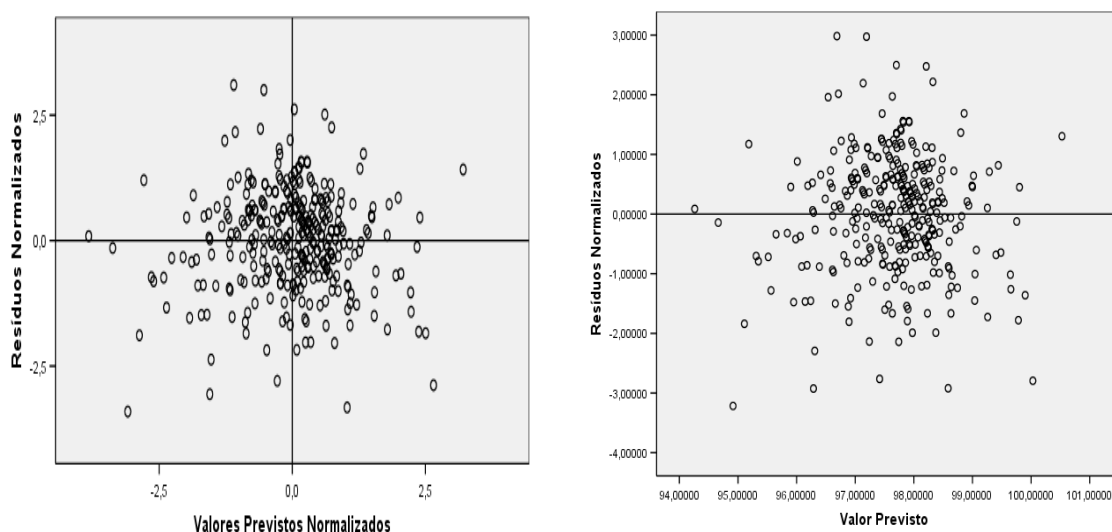


Figura 37– Avaliação da Homocedasticidade para o modelo M_{20} , com 47 variáveis (2003 a 2005).

7.8.3.2 Covariância Nula

De acordo com o teste de Durbin-Watson o modelo M_{20} apresenta um valor de teste $d=1,523$, logo existe auto-correlação entre os resíduos.

7.8.3.3 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, Tabela 46. Esta hipótese é confirmada graficamente uma vez que as observações se dispõem ao longo das rectas oblíqua e horizontal, Figura 38.

Tabela 46- Teste à Normalidade dos resíduos, modelo M_{20} com 47 variáveis (2003 a 2005).

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,040	334	0,200

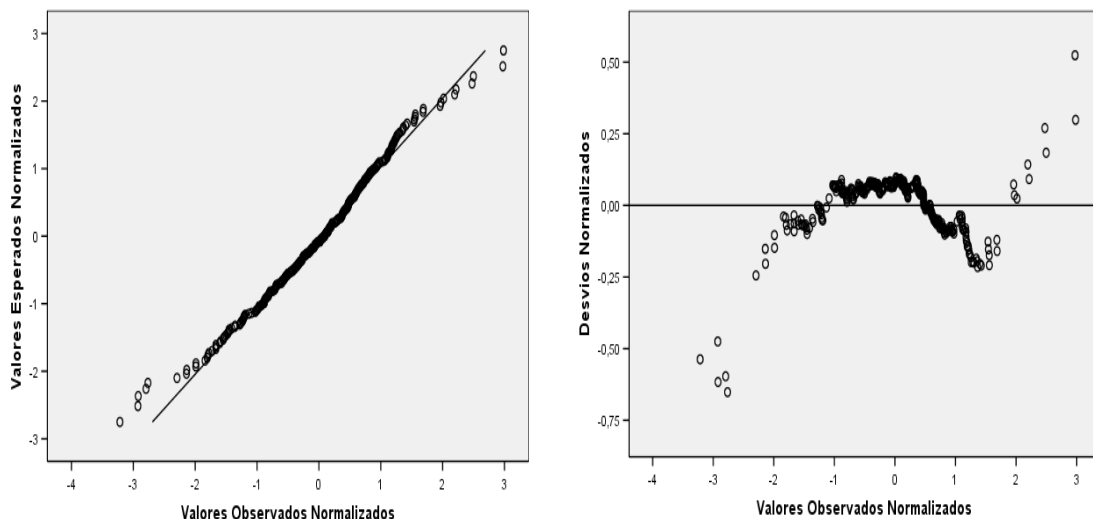


Figura 38- Análise gráfica da normalidade dos resíduos, para o modelo M₂₀, com 47 variáveis (2003 a 2005).

7.8.3.4 Multicolinearidade

Não se verifica a existência de multicolinearidade entre as variáveis independentes. Esta análise é feita com base na matriz das correlações e nos valores da tolerância e VIF, [Tabela 46](#).

7.8.4 Previsão do RON usando o modelo

Com base no modelo desenvolvido é possível calcular o valor de RON, a partir da composição cromatográfica, para as amostras de 2006 não usadas na sua elaboração. Os valores obtidos são comparados com os obtidos pelo CFR através de uma regressão linear, [Figura 39](#), e pelo teste *t de Student* das médias, [Tabela 47](#).

Da comparação dos resultados por regressão linear, verifica-se que os resultados para o RON obtidos por este modelo são mais próximos dos obtidos pelo método do motor CFR, do que os do modelo anterior. O coeficiente de correlação é superior e a recta obtida aproxima-se mais da relação ideal:

$$\text{RON}_{\text{modelo}} = 29,69 + 0,69\text{RON}_{\text{CFR}} \text{ com } R^2 = 0,85.$$

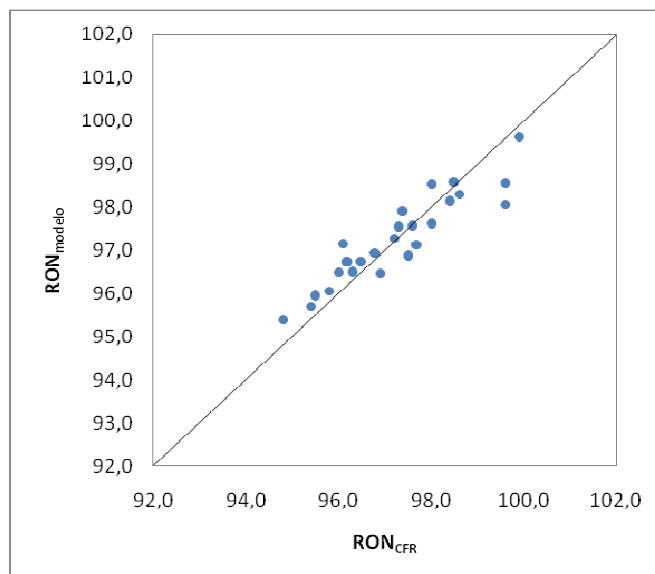


Figura 39- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis (2003 a 2005).

Tabela 47- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis (2003 a 2005).

	RON_{CFR}	RON_{modelo}
Média	97,2	97,3
Desvio Padrão	1,35	1,02
Variância	1,83	1,04
n	26	26
F _{calc}		1,76
F _{tab}		1,96
$F_{calc} < F_{tab}$		
t _{calc}		0,09
t _{tab}		2,01
$t_{calc} < t_{tab}$		

Do teste *t* concluí-se que não existem desvios significativos entre os resultados obtidos por um ou outro método, Tabela 47.

Relativamente às diferenças obtidas entre os dois métodos estas também diminuíram e a percentagem de resultados cuja diferença é superior a $|0,7|$ diminuiu drasticamente de 43% para 12% dos resultados, Figura 40.

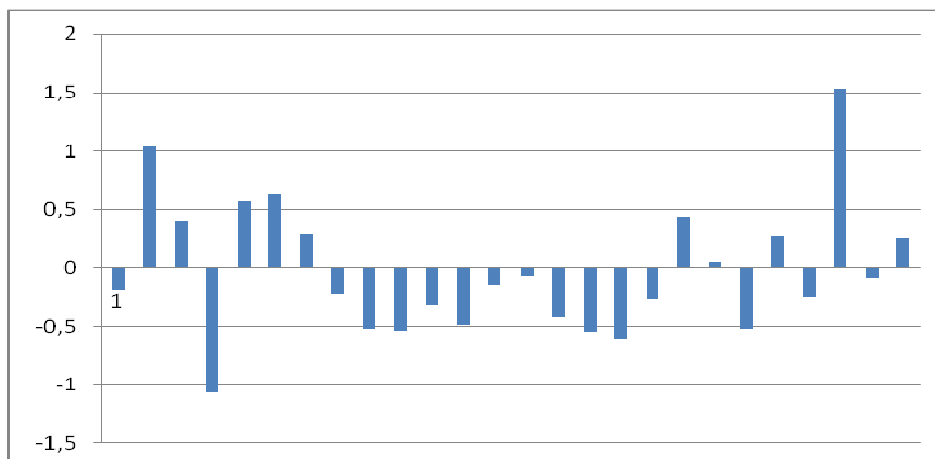


Figura 40- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} com 47 variáveis 2003 a 2005.

8. DISCUSSÃO E CONCLUSÃO

O Número de Octano (RON), à semelhança de outras propriedades físicas e químicas das gasolinas e dos componentes de gasolina, está directamente relacionado com a sua composição [9].

Esta relação é confirmada pelos inúmeros estudos realizados com o objectivo de desenvolver métodos alternativos, na maioria dos casos modelos matemáticos, para a determinação do RON a partir da composição da amostra. De entre os vários métodos alternativos publicados salienta-se que os melhores resultados foram sempre obtidos para aqueles cuja composição foi determinada por cromatografia gasosa (GC) [9]. No entanto apesar desta técnica analítica permitir uma maior discriminação da composição da amostra, esta vantagem nem sempre foi usada uma vez que na maior parte dos modelos desenvolvidos a composição é agrupada, tendo em conta as estruturas dos vários compostos. Estes métodos para além da composição também usam os valores do RON dos compostos puros que constituem a mistura. Estes por sua vez são determinados em condições muito específicas, no entanto quando presentes numa mistura, com outros componentes, apresentam um comportamento bastante diferente [23]. Verifica-se ainda que cada um dos diversos métodos foi sempre desenvolvido para uma matriz específica sendo difícil aplicá-los a outras matrizes. Assim os vários métodos alternativos publicados apresentam um carácter muito específico no que diz respeito à sua aplicação, ao contrário do método de referência ASTM D 2699 que é universal.

A dificuldade de aplicar estes métodos a matrizes diferentes, daquelas para os quais foram desenvolvidos, foi sentida quando se tentaram aplicar alguns dos métodos às amostras de Reformado, matriz objecto de estudo nesta dissertação. Assim surgiu a necessidade de desenvolver um método para a determinação do RON, a partir da composição, direccionado para o Reformado. O método desenvolvido é também muito específico, uma vez que foi desenvolvido para o Reformado e o seu comportamento é ainda desconhecido para outras matrizes.

Inicialmente foi desenvolvida uma série de 3 modelos em que se aumentou gradualmente a discriminação da composição da amostra (4, 7 e 47 variáveis

independentes) no sentido de confirmar se o aumento da discriminação da composição da amostra melhora a capacidade de previsão do RON por parte dos modelos.

Confirma-se que com o aumento da discriminação da composição da amostra o R_a^2 do modelo melhora, passa de 0,704 do primeiro modelo (com 4 variáveis) para 0,866 no terceiro modelo (com 47 variáveis).

A percentagem de amostras em que a diferença entre os valores do RON determinado pelo motor CFR e pelo modelo é superior a $|0,7|$ (reprodutibilidade do método) passa de 22% no primeiro modelo (com 4 variáveis) para apenas 5% no terceiro modelo (com 47 variáveis).

Perante estes resultados conclui-se que o aumento da discriminação da composição da amostra melhora a capacidade de previsão do modelo e por isso os modelos seguintes foram desenvolvidos com as 47 variáveis.

Da análise de resíduos foram encontradas algumas amostras com valores atípicos. Estas observações extremas que apresentam resíduos superiores aos resíduos das restantes amostras [29], podem ser consequência de erros analíticos, de erros de transcrição de resultados, ou mesmo dias de produção anómalos que provocaram alterações na composição das amostras. No entanto como a esta distância não é possível rastrear qual o verdadeiro motivo destes valores, estas amostras não foram consideradas para a elaboração dos modelos.

Uma vez que é sempre indicado validar o modelo com outro conjunto de dados [29], os dados globais foram divididos em 3 grupos, que continham as amostras referentes aos períodos: 2003, 2003 a 2004 e 2003 a 2005.

Assim foi desenvolvido um primeiro modelo com os dados de 2003 e validado com os restantes dados 2004 a 2006, um segundo com os dados de 2003 a 2004 e validado com os dados de 2005 a 2006, e um terceiro modelo com os dados de 2003 a 2005 e validado com os dados de 2006. Os pressupostos do modelo respeitantes aos erros ou resíduos e à quasi(ortogonalidade) entre as variáveis independentes estão validados para os vários modelos [29].

Apesar de disponíveis as 47 variáveis, em nenhum dos modelos foram usadas todas as variáveis disponíveis, o número máximo de variáveis usado foi de 14.

A análise do coeficiente de determinação ajustado e a diferença entre o valor obtido para o RON pelo motor CFR e pelo modelo foram os critérios de qualidade do modelo. Assim verifica-se que o coeficiente de determinação é elevado nos três modelos no entanto diminui com o aumento do número de amostras usadas na sua elaboração: 0,942 no primeiro modelo (2003) e 0,802 no terceiro modelo (2003 a 2005).

Embora o coeficiente de determinação diminua, a capacidade de previsão do RON melhora bastante com a inclusão de mais amostras, uma vez que a diferença entre os valores obtidos pelo motor CFR e pelo modelo passa de 43% no primeiro modelo (2003) para 12% no terceiro modelo (2003 a 2005). No entanto de acordo com o definido no método de referência ASTM D 2699 a reprodutibilidade só deve ser ultrapassada 1 vez em cada 20 ensaios, isto é em 5% dos resultados, o que não acontece neste caso [16]. Este critério tem que ser cumprido para que o modelo possa ser usado como um método alternativo ao método de referência ASTM D 2699. Apesar disto, e de acordo com o teste *t* para as médias, concluímos que em todos os modelos os resultados obtidos por um e outro método não são estatisticamente diferentes, podem ser considerados como pertencentes à mesma população.

Deste modo, é de todo vantajoso num futuro trabalho prosseguir com a inclusão das amostras de 2006 no modelo e a validação do mesmo com as amostras de 2007, no sentido de confirmar a melhoria da capacidade de previsão do modelo. E também com o objectivo de alcançar diferenças de resultados para amostras analisadas pelos dois métodos inferiores à reprodutibilidade do método [0,7].

Nessa nova fase e uma vez que ainda se desconhece qual ou quais as variáveis que conduzem ao melhor modelo, o procedimento aconselhável é novamente a utilização dos três métodos (*Forward*, *Backward* e *Stepwise*) e identificação de quais as variáveis que são deixadas de fora pelos três métodos em simultâneo. Estas variáveis podem então ser eliminadas da análise e o modelo pode ser aplicado às outras variáveis [29].

Outro ponto que pode ser explorado é variar o nível de significância (α) usado como critério de saída e entrada das variáveis no modelo. Neste trabalho foram usados os valores por defeito do SPSS $\alpha_{\text{entrada}}=0,05$ e $\alpha_{\text{saída}}=0,10$. Note-se que a selecção final depende dos valores de α para a selecção de variáveis

e que diferentes valores de α podem conduzir a modelos completamente diferentes [29].

No caso de serem obtidos bons resultados seria interessante testar a aplicação do modelo desenvolvido noutras matrizes, como outros *Componentes de Gasolina*.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] – J. G. Pulido, P. Fonseca, *O Petróleo e Portugal - O Mundo do Petróleo e o seu impacto no nosso país*, Tribuna, 2004.
- [2] - www.galpennergia.com
- [3] - Soares Mota, *Refinação de Petróleos*, apresentação na Galp Energia em Lisboa a 11 de Março de 2008;
- [4] - www.setlaboratories.com
- [5] - www.intertek-cb.com
- [6] - Concawe- *The Oil Companies European Organization for Environmental and Health Protection*, Product. Dossier no. 92/103 gasolines, 1992.
- [7] – www.faqs.org, Bruce Hamilton, *Automotive Gasoline*.
- [8] - N. Pasadakis, V. Gaganis, C. Foteinopoulos, *Octane number prediction for gasoline blends*, Fuel Processing Technology 87 (2006) 505-509.
- [9] – N. Nikolaou, C. E. Papadopoulos, I. A. Gaglias, K. G. Pitarakis, *A new-linear calculation method of isomerisation gasoline research octane number based on gas chromatographic data*, Fuel 83 (2004) 517-523.
- [10] – R. L. Furey, A. M. Horowitz, N. J. Schroeder, *Manual on significance of Tests for Petroleum Products*, Automotive Gasoline, 6ª edição, 1993.
- [11] - Directiva Europeia 2003/17/CE, transposta pelo DL 235/2004.
- [12] – *Fabricar uma Gasolina Moderna*, Shell Ciência & Tecnologia, edição especial, Março de 1987.
- [13] – J. P. Wauquier, *Petroleum Refining – Crude Oil, Petroleum Products, Process Flowsheets*, Institut Français du Pétrole Publications, Éditions Technip, 1995.
- [14] – www.asifunciona.com/mecanica
- [15] – www.bp.com
- [16] - ASTM D 2699-06a, IP 237/87, *Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel*, 2006.
- [17]- ASTM D 2700-06. IP 236/87, *Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel*, 2006.
- [18]- W. F. Schoen, A. V. Mrstik, *Calculating gasoline blend octane ratings*, Ind. and Engr. Chem., 47(9), 1955, pp 1740-1742.
- [19]- W. E. Stewart, *Predict octanes for gasoline blends*, Petroleum Refiner, 38(12), 1959, pp. 135-139.

- [20]- W.C. Healy, C. W. Jr Maassen, R. T. Peterson, *A new approach to blending octanes*, API Division of Refining, 24th midyear meeting, New York, May 27, 1959.
- [21]- M. H. T. Auckland, D. J. Charnock, *The development of linear blending indices for petroleum properties*, J. of the Inst. Of Petroleum, 55(545), 1969, pp. 322-329.
- [22]- P. C. Anderson, J. M. Sharkey, R. P. Walsh, *Calculation of Research Octane Number of Motor Gasolines from Gas Chromatographic Data and New Approach to Motor Gasoline Quality Control*, Journal of the Institute of Petroleum, Volume 58, N.º 560, Março 1972.
- [23]- *Knocking Characteristics of Pure Hydrocarbons*, API Research Project 45. American Society for Testing Materials, 1958.
- [24]- W. E. Morris, *Interaction approach to gasoline blending*, NPRA Paper AM-75-30, NPRA annual meeting, March 1975.
- [25]- M. H. Rusin, H. S. Chung, J. F. Marshall, *A transformation method for calculating the research and motor octane numbers of gasoline blends*, Ind. Eng. Chem. Fundam., 20(3), 1981, pp. 195-204.
- [26]- A. H. Zahed, S. A. Mullah, M. D. Bashir, *Predict octane number for gasoline bends*, Hydrocarbon Processing, May 1993, pp. 85-87.
- [27] – Luiz Alberto da Silva, Vera Maria Moura Torres da Costa, *Correlações de Octanagem por Cromatografia Gasosa e Análise PIANIO*, Petrobras.
- [28] - Manual do equipamento HP 5890- Tecnologia AC- Analisador PIONA;
- [29] - João Maroco, *Análise Estatística – com utilização do SPSS*, Edições Sílabo, 3ª edição, 2007.
- [30] – R. C. Guimarães, J. A. Sarsfield Cabral, *Estatística*, McGraw-Hill Lda, 1997.
- [31] – M. H. Pestana, J. N. Gageiro, *Análise de Dados para Ciências Sociais- A Complementaridade do SPSS*, Lisboa, Edições Silabo, 2003.
- [32] – James N. Miller & Jane C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Pearson Education Limited, 4th edition, 2000.

1. INTRODUÇÃO

O petróleo é uma mistura complexa de hidrocarbonetos constituídos essencialmente por átomos de carbono e hidrogénio. A sua formação teve origem há centenas de milhões de anos, a partir de restos vegetais e animais e encontra-se depositado no solo a várias profundidades. A sua composição exacta varia consoante o local de extracção [1].

Apesar de existirem referências históricas sobre a sua existência cerca de 3000 anos antes de Cristo, o petróleo foi olhado ao longo da maior parte da história da humanidade como uma mera curiosidade. A sua emergência como recurso estratégico de fundamental importância económica é uma realidade com pouco mais de um século, e associada ao desenvolvimento tecnológico que marcou os processos de extracção e refinação, assim como ao alargamento da sua utilização aos mais diversos sectores da actividade humana [1].

Nos dias de hoje, o petróleo assume uma importância ímpar no âmbito dos recursos estratégicos fundamentais. Os choques petrolíferos de 1973 e 1979 colocaram a descoberto a verdadeira dependência das economias mais avançadas do mundo em relação a este recurso, bem como reforçaram o sentimento de poder dos países detentores das maiores reservas [1].

A indústria petrolífera está claramente dividida em duas grandes áreas: uma que compreende a pesquisa, exploração e produção e outra que engloba o transporte, refinação e logística. Na primeira importa avaliar parâmetros tais como o domínio mineiro sobre a zona de interesse, o risco político e económico subjacente e a logística, antes de iniciar esta etapa. Na segunda, e uma vez que o petróleo bruto contém centenas de diferentes tipos de hidrocarbonetos, é necessário separá-los e transformá-los a fim de se obterem os produtos pretendidos. A diversidade desses produtos, assim como as suas características implicam um conjunto complexo de operações de refinação, as quais se podem dividir em 3 grupos fundamentais [2]:

- *Operações físicas de separação* - incluem o fraccionamento por destilação e a extracção por solvente. São operações que permitem separar vários produtos utilizando as diferenças de pontos de ebulição, a densidade ou a solubilidade em solventes selectivos [2].

- *Operações químicas de conversão molecular* - reúnem todas as operações que se destinam a alterar a estrutura molecular dos componentes do petróleo de forma a obter produtos com determinadas características. Estas operações efectuam-se em condições variáveis de pressão e temperatura, podendo utilizar-se ou não um catalisador e decorrerem em atmosfera de hidrogénio [2].
- *Operações físicas ou químicas de tratamento* - destinadas a melhorar a qualidade dos vários derivados do petróleo por remoção de certas impurezas [2].

Das áreas da indústria petrolífera, em Portugal apenas existe a área da refinação, uma vez que a pesquisa de reservas de hidrocarbonetos ainda não detectou petróleo em quantidade que permita a sua extracção no nosso país, em condições económicas favoráveis [1].

No âmbito da refinação, Portugal possui duas refinarias, uma situada no Porto e outra em Sines. Ambas são propriedade da Galp Energia, detentora do monopólio da refinação em Portugal. A Refinaria de Sines possui uma capacidade de destilação instalada de 10,4 milhões de toneladas/ano e está orientada para maximizar a produção de gasolinas. A Refinaria do Porto tem uma capacidade de destilação instalada de 4,8 milhões de toneladas/ano e está orientada para a produção de combustíveis, lubrificantes, ceras de petróleo, produtos aromáticos de base e solventes industriais. As duas refinarias dispõem de instalações tecnologicamente actualizadas sendo que as suas valências não se sobrepõem, mas complementam-se relativamente às necessidades do país [2].

A Refinaria do Porto iniciou a sua laboração em 1969 e conjuga as Fábricas de Combustíveis, Óleos Base, Lubrificantes e Aromáticos [2].

No caso da Fábrica de Combustíveis, é constituída por 14 Unidades Processuais, Figura 1, que utilizando o petróleo bruto como matéria-prima obtêm toda a gama de produtos combustíveis incluindo grande parte da matéria-prima que alimenta a Fábrica de Aromáticos. Dada a flexibilidade e interligação das suas instalações, trata uma grande variedade de petróleos brutos, escolhidos caso a caso, de acordo com as suas características e as necessidades do mercado [2].

Assim, o petróleo bruto alimenta a Unidade de Destilação Atmosférica da Fábrica de Combustíveis (Unidade 3000), que o separa em cinco grandes fracções mais simples que serão tratadas em unidades processuais a jusante, passando a constituir componentes de produtos comerciais. Estas cinco fracções são designadas por: *Produtos de Topo* (leves), *Gasolina Pesada*, *Petróleo*, *Gasóleo* e *Resíduo Atmosférico* [2].

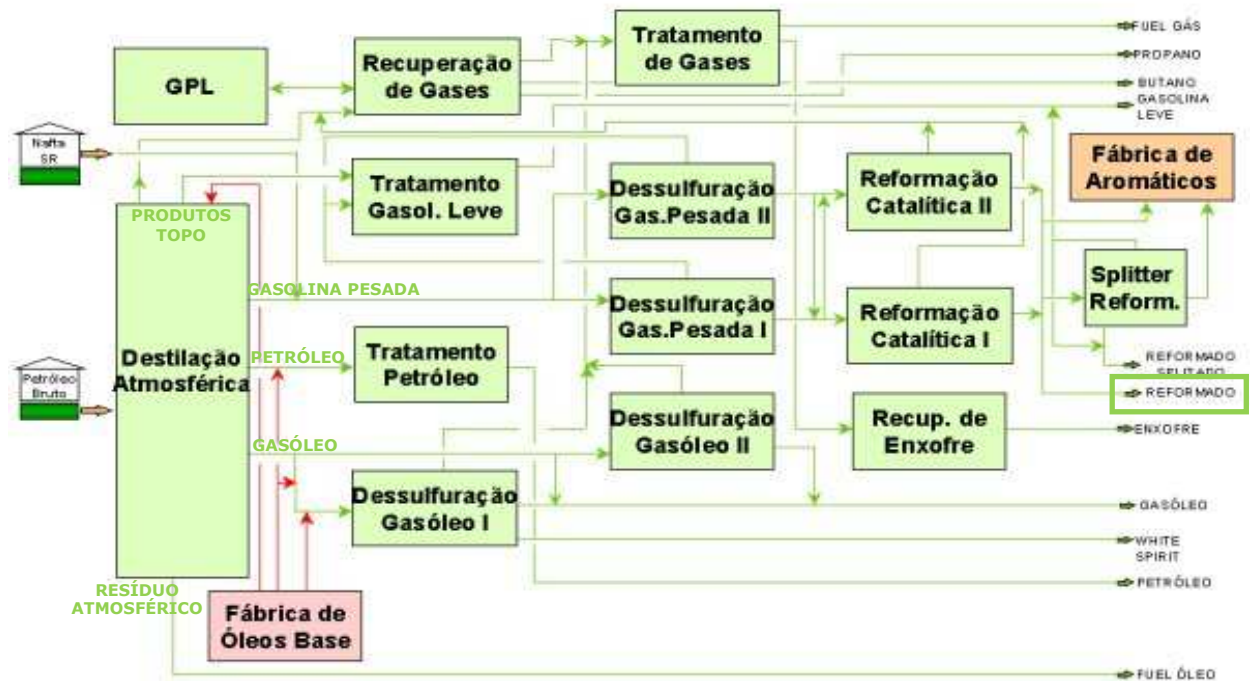


Figura 1- Esquema das várias unidades e correntes processuais que constituem a Fábrica de Combustíveis da Refinaria do Porto da Galp Energia [2].

Ao petróleo separado na destilação à pressão atmosférica são removidos alguns contaminantes e é usado como combustível para aviação designado por *Jet-Fuel*. O gasóleo proveniente da destilação à pressão atmosférica é dessulfurado, reduzindo drasticamente o seu teor em enxofre. O produto de topo é redestilado, dando origem a gases incondensáveis (*Fuel Gás*), gases de petróleo liquefeitos (*GPL*), naftas leve e pesada aos quais são posteriormente removidos os compostos de enxofre. A gasolina pesada é enviada para a Unidade de *Platforming semi-regenerativo e de regeneração contínua* a qual converte os hidrocarbonetos nafténicos e parafínicos em aromáticos (*reformação catalítica*). O produto obtido é designado por *Reformado*,

apresenta um elevado teor de hidrocarbonetos aromáticos e um número de octano também bastante elevado. Este produto é usado como componente de gasolinas ou como matéria-prima para a Fábrica de Aromáticos [2].

As correntes processuais das várias unidades produtivas de cada uma das fábricas acima referidas, são analisadas de acordo com o plano de inspecção e ensaios (PIE) de cada fábrica. Este plano foi elaborado por cada uma das fábricas em conjunto com o laboratório e o seu objectivo é definir quais as características analisadas em cada uma das correntes processuais e qual a periodicidade dessas análises. O cumprimento deste plano permite um controlo diário do processo de produção, a partir dos resultados obtidos para cada uma das características analisadas nas várias unidades. Nas fases de movimentação de produtos e expedição, todos os produtos finais são também analisados a fim de verificar a conformidade com as especificações em vigor antes da sua expedição. Neste sentido o Laboratório da Refinaria do Porto está equipado com todos os equipamentos necessários à realização dos ensaios químicos, físico-químicos e mecânicos que garantam o controlo processual e a comprovação da qualidade dos produtos [2].

A presente dissertação procura contribuir para a melhoria contínua do desempenho do Laboratório. Assim o que se propõe neste trabalho é o desenvolvimento de um método alternativo, mais rápido e económico para a determinação do *Número de Octano*, no *Reformado*, a partir da sua composição cromatográfica. Este método tem por base um modelo de Regressão Linear Múltipla que permite a partir de um único ensaio determinar duas características, reduzindo assim o tempo e os recursos humanos necessários e melhorando com isso a capacidade de resposta do Laboratório.

Face ao objectivo do trabalho, e uma vez que mais à frente será usada a composição do *Reformado* para o desenvolvimento do modelo, no ponto seguinte será feita uma caracterização da composição do petróleo bruto uma vez que está na origem de todo o processo de refinação.

De seguida será feita também uma caracterização da gasolina, uma vez que o *Reformado*, corrente processual em estudo neste trabalho, é um dos componentes usado na produção de gasolina. Assim será apresentada a sua forma de produção, composição, características de interesse, e a importância de algumas delas no funcionamento dos motores automóveis.

2. CARACTERIZAÇÃO DE PETRÓLEOS BRUTOS E PRODUTOS DE REFINAÇÃO

2.1 COMPOSIÇÃO DOS PETRÓLEOS BRUTOS

Os petróleos brutos são misturas naturais de hidrocarbonetos e de derivados orgânicos de enxofre, azoto e oxigénio, existentes no subsolo, no estado líquido. A estas misturas juntam-se normalmente outras substâncias, designadamente água, matéria inorgânica e gás sulfídrico, cuja posterior separação não altera a composição orgânica da mistura original [3].

Do ponto de vista físico, o petróleo bruto varia desde líquidos relativamente pouco viscosos, de cor palha escura, até fluídos muito viscosos, semi-sólidos. A sua massa volúmica oscila entre 790 e 1010 kg/m³ e a viscosidade varia muito desde 0,7 até 42.000 cP. Os petróleos brutos são substâncias inflamáveis em condições ambientais e exalam cheiros que variam dum quase fragância até um odor repulsivo (característico dos compostos orgânicos de enxofre) [3].

Apesar da grande diversidade observada nas características físicas, tal não se verifica no que diz respeito à composição elementar, que varia entre limites relativamente estreitos, Tabela 1. A explicação para esta particularidade é o facto dos diferentes petróleos brutos serem constituídos por hidrocarbonetos pertencentes a reduzidas séries ou famílias, Tabela 2 [3].

Tabela 1- Composição elementar média dos petróleos brutos [3].

Elemento	%m/m
Carbono	83,9-86,8
Hidrogénio	11,0-14,0
Enxofre	0,04-6,0
Azoto	0,10-1,50
Oxigénio	0,10-0,50
Metais	0,005-0,015

Tabela 2-Composição típica de alguns petróleos brutos de diferentes origens [4].

Fonte de Petróleo Bruto	Parafinas (%vol)	Aromáticos (%vol)	Naftênicos (%m/m)	Enxofre (%m/m)	Densidade API	Número de Octano
Nigéria (Leve)	37	9	54	0,2	36	60
A.Saudita (Leve)	63	19	18	2	34	40
A.Saudita (Pesado)	60	15	25	2,1	28	35
Venezuela (Pesado)	35	12	53	2,3	30	60
Venezuela (Leve)	52	14	34	1,5	24	50
USA- Midcont.	-	-	-	0,4	40	-
USA- Texas	46	22	32	1,9	32	55
Brent- Mar do Norte	50	16	34	0,4	37	50

2.2 HIDROCARBONETOS PRESENTES NOS PETRÓLEOS BRUTOS

Os petróleos brutos são constituídos por moléculas de hidrocarbonetos que podem ter desde 1 a 60 átomos de carbono. As propriedades desses hidrocarbonetos dependem do número e dos diferentes arranjos dos átomos de carbono e hidrogênio nas moléculas [3].

Os hidrocarbonetos que contêm até 4 átomos de carbono estão normalmente no estado gasoso, os que têm entre 5 e 19 átomos de carbono estão no estado líquido e os que têm 20 átomos de carbono, ou mais, encontram-se no estado sólido. Estes hidrocarbonetos agrupam-se da seguinte forma [4]:

- **Parafinas**, também designadas por alcanos, são hidrocarbonetos lineares, as *n*-parafinas, ou ramificados, as *i*-parafinas, de fórmula C_nH_{2n+2} normalmente com 1 a 20 átomos de carbono. Exemplos de parafinas: metano, etano, propano, *n*-butano, isobutano, pentano, *n*-hexano, *n*-heptano, *n*-octano, isooctano, etc, Figura 2 [4].

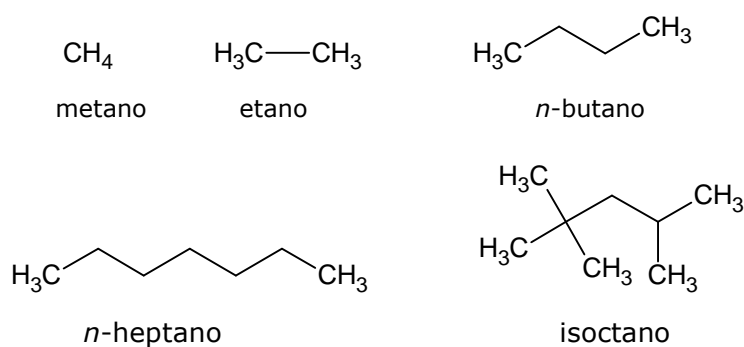
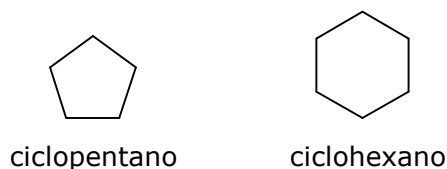


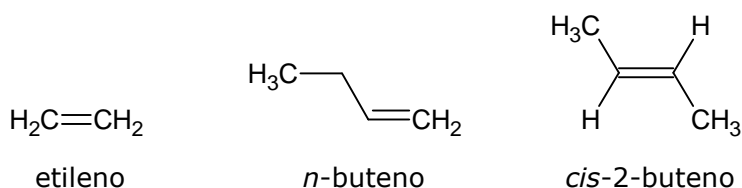
Figura 2- Estrutura química de algumas parafinas.

- **Nafténicos**, também designados por cicloparafinas ou cicloalcanos, são hidrocarbonetos com estrutura em anel mas sem ligações duplas, de fórmula C_nH_{2n} e com um ou mais anéis por molécula. Exemplos de hidrocarbonetos nafténicos: ciclopentano, ciclohexano, etc, [Figura 3](#) [4].



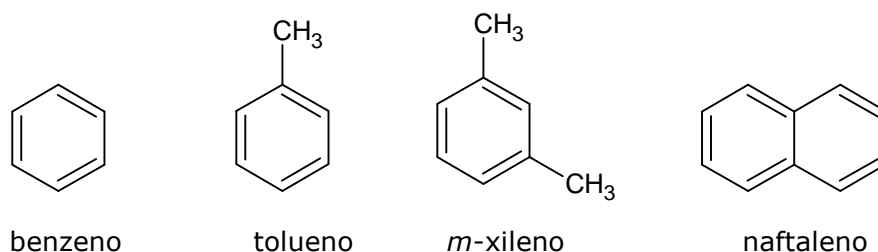
[Figura 3](#)- Estrutura química de alguns compostos nafténicos.

- **Olefinas**, também designadas por alcenos, são hidrocarbonetos lineares, ramificados ou cíclicos, com ligações duplas de fórmula C_nH_{2n} . Exemplos de olefinas: etileno, buteno, isobuteno, etc, [Figura 4](#). Estes compostos raramente estão presentes no petróleo bruto, e surgem normalmente durante o processo de refinação [4].



[Figura 4](#)- Estrutura química de algumas olefinas.

- **Aromáticos**, são hidrocarbonetos com estruturas em anel (anel de benzeno com 6 átomos de carbono e ligações duplas alternadas), com um ou mais anéis por molécula. Exemplos de hidrocarbonetos aromáticos: benzeno, tolueno, *m*-xileno, naftaleno, etc, [Figura 5](#) [4].



[Figura 5](#)- Estrutura química de alguns compostos aromáticos.

2.3 CLASSIFICAÇÃO DOS DIFERENTES PETRÓLEOS BRUTOS

Dada a diversidade de características dos petróleos brutos é frequente agrupá-los segundo uma propriedade física ou química comum [3].

A densidade API (American Petroleum Institute) é uma característica física usada para agrupar os petróleos brutos e classifica-os em quatro grandes categorias: Leves, Médios, Pesados e Muito Pesados. Quanto maior for a densidade API mais leve é o petróleo e mais baixa é a massa volúmica [3].

Os petróleos brutos também são agrupados por famílias, de acordo com as famílias predominantes na sua composição química. Designam-se por: *Petróleos Brutos Parafínicos*, *Nafténico-Parafínicos*, *Nafténicos*, *Aromáticos*, *Aromático-Asfálticos* e *Aromático-Nafténicos* [3].

Uma outra forma de agrupar os petróleos brutos considera como característica distintiva o teor em enxofre total. De acordo com este critério os petróleos brutos dividem-se em 'sweet' e 'sour' consoante o teor em enxofre é menor que 1% m/m ou maior e igual a 1% m/m respectivamente [3].

2.4 PRODUTOS DA REFINAÇÃO DO PETRÓLEO BRUTO

Os diferentes componentes do petróleo bruto têm pesos moleculares diferentes e moléculas com diferentes configurações e, como tal, diferentes pontos de ebulição. Assim é possível separá-los por destilação em que no topo da coluna se recolhem os hidrocarbonetos mais leves (essencialmente GPL). Na base da coluna, no resíduo, recolhem-se os hidrocarbonetos mais pesados e normalmente recolhem-se ao longo da coluna de destilação produtos intermédios tais como nafta, gasolina e gasóleo, por ordem crescente de pontos de ebulição [5]. Tal como esquematizado na [Figura 6](#) e já mencionado no ponto 1 quando foi apresentada a Unidade 3000 da Fábrica de Combustíveis da Refinaria do Porto da Galp Energia.

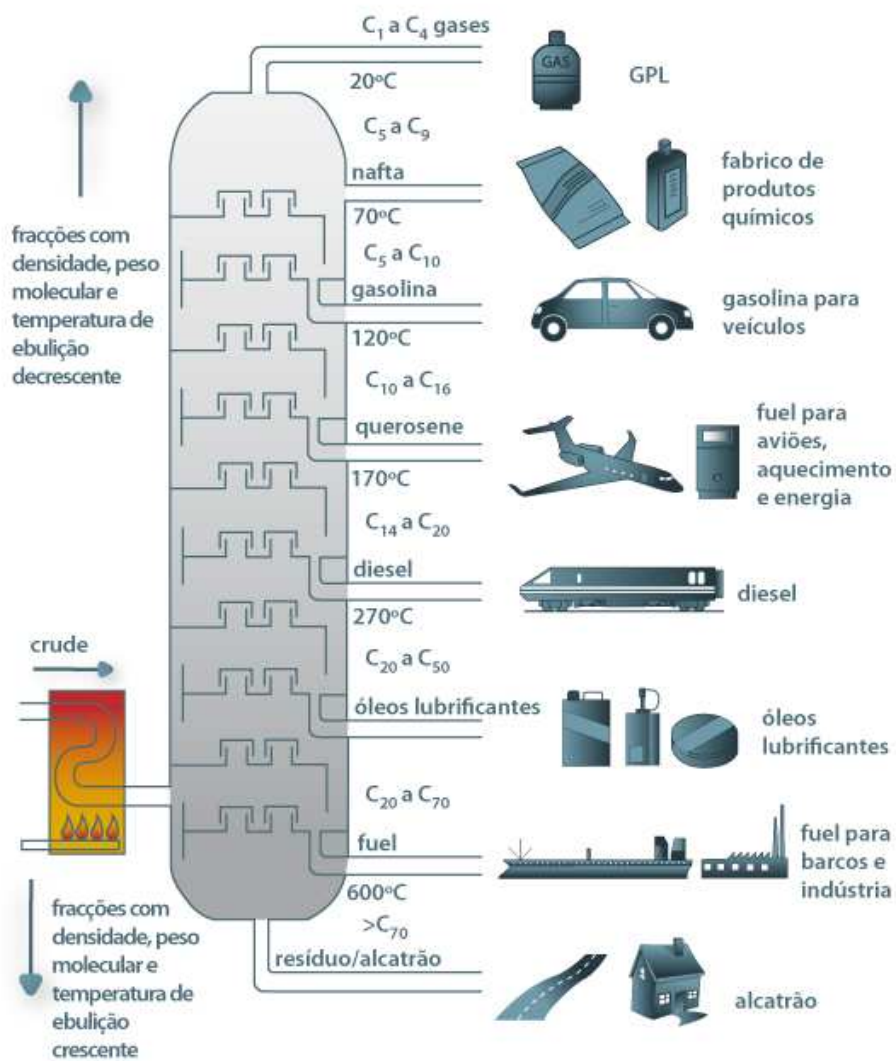


Figura 6- Esquema da destilação fraccionada do petróleo bruto [5].

3. CARACTERIZAÇÃO DA GASOLINA

3.1 PROCESSO DE PRODUÇÃO DA GASOLINA

A gasolina é uma combinação complexa de hidrocarbonetos, constituídos principalmente por compostos parafínicos, nafténicos, aromáticos e olefínicos, com número de carbonos predominantemente maiores que C₃ (de C₄ a C₁₂) e uma gama de temperaturas de ebulição de 30°C a 220°C [6].

Esta combinação é produzida a partir da mistura de várias correntes de refinaria de modo a obter um produto final com as características desejadas. Muitas gasolinas na Europa contêm também componentes de origem não petrolífera especialmente oxigenados (maioritariamente éteres e álcoois), Figura 7, e aditivos que podem ser usados para melhorar o desempenho do combustível [6].

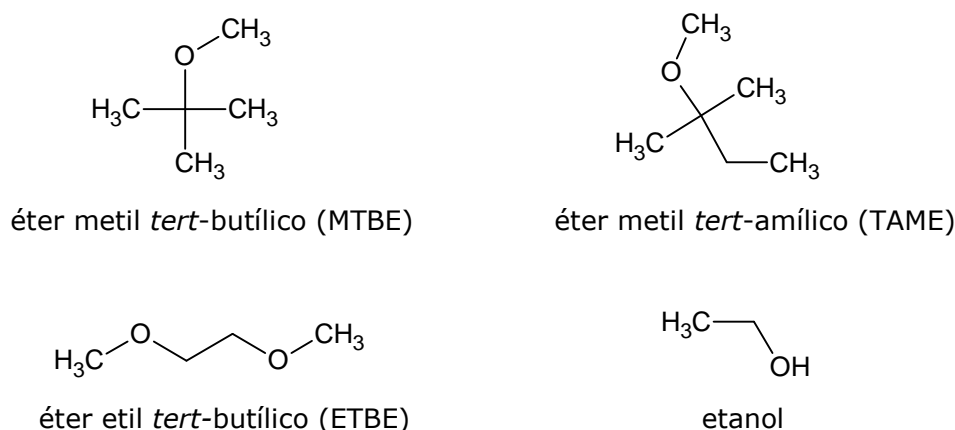


Figura 7- Estrutura química de alguns dos compostos oxigenados adicionados às gasolinas.

A composição dos hidrocarbonetos presentes na gasolina varia muito, dependendo do tipo e da natureza de cada petróleo bruto processado, dos processos de refinação disponíveis, das condições do processo e do balanço global exigido entre a gasolina e os outros produtos de refinação [6].

As correntes de refinaria mais usadas para a produção de gasolinas, são designadas por *Componentes de Gasolina* e derivam de vários processos produtivos [7]:

- *Cracking Catalítico* – consiste na quebra dos hidrocarbonetos maiores e com maior ponto de ebulição, em hidrocarbonetos na gama dos

existentes na gasolina e contendo cerca de 30% de compostos aromáticos e 20-30% de olefinas [7];

- *Hidrocracking* – quebra de moléculas e hidrogenação, produzindo fracções de gasolina mais saturadas e estáveis [7];
- *Isomerização* – conversão de hidrocarbonetos de cadeia linear em hidrocarbonetos de cadeia ramificada que como consequência produz um aumento do número de octano da gasolina [7];
- *Reforming* – converte os hidrocarbonetos saturados, com baixo número de octano, em hidrocarbonetos com elevado número de octano contendo cerca de 60% de compostos aromáticos [7];
- *Alquilação* – reacção de uma corrente gasosa de olefinas com isobutano de modo a produzir *i*-parafinas com elevado número de octano [7].

Devido às diferentes composições dos *Componentes de Gasolina*, para se obter uma gasolina final, é necessário proceder à sua mistura - *Blending*. O processo físico de *Blending* de componentes é simples mas a escolha dos componentes a usar, assim como a percentagem de cada um na mistura, é mais complexo [8]. A operação é feita por bombagem simultânea de todos os componentes em que cada propriedade, de cada componente, é uma variável e o seu efeito na mistura final é considerável. Assim é fundamental conhecer as características de cada componente antes de proceder ao seu *Blending*.

Nos últimos anos surgiu uma preocupação crescente dos governos em criar regulamentos para proteger o meio ambiente. Muitos desses regulamentos foram feitos com o objectivo de minimizar as emissões dos automóveis, nomeadamente em CO (monóxido de carbono), CO₂ (dióxido de carbono), NO_x (compostos de azoto), SO_x (compostos de enxofre), VOCs (compostos orgânicos voláteis) e partículas, para a atmosfera [7].

Inicialmente a maioria dos regulamentos foram dirigidos para os automóveis e tiveram resultados nas tecnologias, aumentando a eficiência dos motores, o que reduziu significativamente as emissões [7]. Posteriormente as regulamentações foram dirigidas para a mudança da composição dos combustíveis (fim dos aditivos à base de compostos de chumbo, menor volatilidade, menor teor em olefinas e aromáticos) com o objectivo de reduzir ainda mais as emissões dos veículos e torná-las menos tóxicas [9].

Estas novas considerações ambientais alteraram as especificações comerciais das gasolinas e introduziram uma série de dificuldades na sua produção [9]. As refinarias tiveram que alterar os seus processos de fabrico de modo a produzir gasolinas menos voláteis, com menor teor em olefinas e em compostos aromáticos, nomeadamente benzeno. Assim passaram a incorporar componentes de origem não petrolífera, compostos oxigenados e outros aditivos, com o objectivo de compensar as perdas no desempenho do combustível provocadas por estas variações na composição [7].

3.2 ESPECIFICAÇÃO COMERCIAL DA GASOLINA

Actualmente a composição da gasolina varia de refinaria para refinaria e não é apenas um único critério que caracteriza a sua qualidade [9]. A gasolina, para além de satisfazer as exigências dos motores automóveis, está adicionalmente exposta a uma variedade de condicionantes mecânicas, físicas e químicas. Assim as suas propriedades têm que ser um compromisso entre o bom desempenho do motor, perante condições operatórias adversas, e o cumprimento dos parâmetros ambientais [10].

Em 2004, a Directiva Europeia 2003/17/CE que regulamenta as especificações de Gasolinas e Gasóleos foi transposta para a lei Portuguesa pelo Decreto de Lei 235/2004. Esta especificação tem como objectivo assegurar que toda a gasolina comercializada apresenta um padrão de qualidade mínimo para as características fundamentais. A Tabela 3 [11] apresenta os valores das especificações para essas características nas Gasolinas Euro Super e Super Plus.

3.3 CARACTERÍSTICAS FUNDAMENTAIS DA GASOLINA

Todas as características presentes na especificação comercial das gasolinas são importantes e têm que ser cumpridas, umas por motivos ambientais, outras para permitirem o bom funcionamento dos motores automóveis. No caso das características associadas ao motor há quatro que têm predominante importância: a volatilidade, a estabilidade, a corrosão, e as características de combustão [7].

Tabela 3 – Especificações para as principais características das Gasolinas Euro Super e Super Plus [11].

Característica	Unidades	Euro Super		Super Plus	
		Mínimo	Máximo	Mínimo	Máximo
Aspecto		Claro e límpido		Claro e límpido	
Cor		Violeta		Violeta	
Massa Volúmica a 15°C	Kg/m ³	720	775	720	775
Número de Octano - RON		95	-	98	-
Número de Octano - MON		85	-	87	-
Pressão de Vapor:					
De 1 de Maio a 30 de Setembro	kPa	45,0	60,0	45,0	60,0
Meses de Outubro a Abril	kPa	60,0	90,0	60,0	90,0
De 1 de Novembro a 31 de Março	kPa	60,0	90,0	60,0	90,0
Destilação					
Evaporado a 70°C					
De 1 de Maio a 30 de Setembro	%v/v	20,0	48,0	20,0	48,0
De 1 de Outubro a 30 de Abril	%v/v	22,0	50,0	22,0	50,0
Evaporado a 100°C	%v/v	46,0	71,0	46,0	71,0
Evaporado a 150°C	%v/v	75,0	-	75,0	-
Ponto Final	°C	-	210	-	210
Resíduo	%v/v		2	-	2
Análise de hidrocarbonetos					
Olefinas	%v/v	-	18,0	-	18,0
Aromáticos	%v/v	-	42,0	-	42,0
Benzeno	%v/v	-	1,00	-	1,00
Teor em oxigénio	%m/m	-	2,7	-	2,7
Teor em enxofre	mg/kg	-	150	-	150
Teor de chumbo	g/l	-	0,005	-	0,005
Estabilidade à Oxidação	min	360	-	360	-
Gomas existentes (lavadas com solvente)	mg/100 ml	-	5	-	5
Corrosão à lâmina de Cobre (3h a 50°C)			1		1
Compostos Oxigenados					
Metanol	%v/v	-	3,0	-	3,0
Etanol	%v/v	-	5,0	-	5,0
Álcool Isopropílico	%v/v	-	10,0	-	10,0
Álcool Terbutílico	%v/v	-	7,0	-	7,0
Álcool Isobutílico	%v/v	-	10,0	-	10,0
Éteres com 5 ou mais átomos de carbono por molécula	%v/v	-	15,0	-	15,0
Outros Compostos Oxigenados	%v/v	-	10,0	-	10,0

A volatilidade afecta de várias formas o comportamento do motor. Se for muito baixa, o motor pode ter dificuldade em arrancar a frio e leva muito tempo a aquecer. Se for muito elevada, pode provocar problemas de falhas do motor ou funcionamento irregular em tempo quente. A volatilidade também tem efeito na economia e energia, podendo influenciar ainda a formação de depósitos no motor. A relação entre a volatilidade e o comportamento do motor está associada a grupos de hidrocarbonetos que se encontram em zonas distintas do intervalo de destilação da gasolina. Para a refinação isto representa uma limitação e portanto um cuidadoso equilíbrio dentro do intervalo de destilação, de modo que a gasolina permaneça entre os limites de vaporização especificados [12]. A volatilidade é controlada pela *pressão de vapor* e pelo *intervalo de destilação* [7].

A estabilidade da gasolina é conseguida com a incorporação de aditivos que a tornam mais estável, promovem a limpeza e evitam a formação de depósitos nos motores. Caso contrário a gasolina poderia originar sujidade nos motores (formação de gomas) provocada pela sua deterioração por oxidação [12]. A estabilidade da gasolina é controlada por duas das características mencionadas na especificação: a *estabilidade à oxidação* e as *gomas existentes* [7].

O enxofre presente na gasolina pode provocar corrosão, e depois da combustão pode formar gases corrosivos que atacam o motor, o escape e o meio ambiente. A *corrosão à lâmina de cobre* avalia a actividade do enxofre e o *teor em enxofre* reporta o enxofre total presente na gasolina. Estas duas características são usadas para avaliar a qualidade do combustível relativamente à corrosão [7].

Os combustíveis queimam-se nos motores dos automóveis para deslocar os êmbolos, tal como descrito à frente no ponto 4.1. Uma medida com interesse é comparar o volume do cilindro após o êmbolo ter sido totalmente deslocado pelos gases em expansão, com o volume do cilindro quando o êmbolo está no seu ponto superior. Isto representa a taxa de compressão de cada motor. Um automóvel com taxa de compressão de 7,5 necessita, como modelo típico, de uma gasolina com número de octano de 88 para evitar o fenómeno de detonação, definido à frente no ponto 4.2. Aumentando a taxa de compressão para 9 passa a necessitar de um número de octano de cerca de 98. Este

aumento de número de octano de 10 tem a acompanhá-lo um aumento de cerca de 10% de rendimento, mas esta proporção de aumento de 1% de rendimento para cada 1% de aumento do número de octano não se mantém indefinidamente. Taxas de compressão cada vez maiores traduzem-se em diferenças de rendimento cada vez menores [12]. Assim quanto maior for a relação de compressão de um motor, maior será a economia de combustível e a potência desenvolvida. Esta relação, ou taxa de compressão, é contudo limitada pelo aparecimento da detonação. O poder anti-detonante de um combustível é expresso em termos do *número de octano*. Esta qualidade é, como a volatilidade, uma propriedade fundamental das gasolinas. Quanto maior for o número de octano de uma gasolina, melhor poderá resistir ao aparecimento da detonação e melhor será o rendimento do motor [12].

4. APLICAÇÃO DA GASOLINA NO MOTOR

4.1 FUNCIONAMENTO DO MOTOR A GASOLINA

Os motores dos veículos, cujo combustível usado é a gasolina, são motores de explosão ou combustão interna, e seguem um ciclo denominado por *Ciclo a Quatro Tempos* ou *Ciclo de Otto*. Os quatro tempos do ciclo designam-se por: *Admissão*, *Compressão*, *Explosão* e *Escape*, Figura 8 [13].

Para ter início o funcionamento do motor, a válvula de admissão abre-se enquanto o pistão se move para baixo. A gasolina é bombeada a partir do depósito e misturada com o ar no cilindro, esta fase é designada por *Admissão*. O pistão sobe e comprime a mistura ar-gasolina, na fase de *Compressão*, o que torna a explosão mais potente. Quando o pistão atinge o topo do seu curso, a vela de ignição produz uma faísca que inflama a gasolina. A gasolina no cilindro entra em combustão e expande-se empurrando o êmbolo ao longo do cilindro e gerando assim a potência necessária para criar o movimento. Logo, a primeira exigência de funcionamento para uma gasolina é que a sua combustão seja rápida e completa, no interior das câmaras de combustão, de modo a libertar o máximo de energia – *Explosão* [13].

De seguida o pistão atinge o topo do seu curso, a válvula de escape abre-se e os gases queimados saem do cilindro pelo tubo existente para esse fim – *Escape* [13].

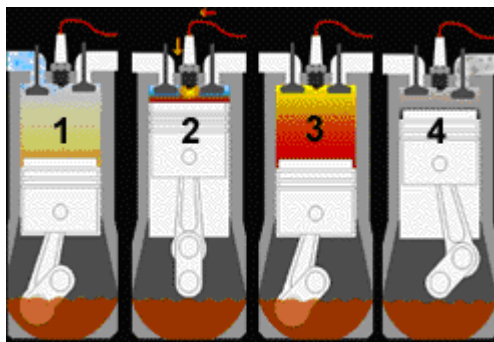


Figura 8 – Quatro tempos de um motor a gasolina: 1-Admissão, 2-Compressão, 3-Explosão e 4-Escape [14].

4.2 FENÓMENO DE DETONAÇÃO

Nos motores de combustão interna, são possíveis vários tipos de combustão. O processo normal é uma combustão, rápida e suave, da mistura combustível com o ar no motor, devido à propagação da chama proveniente da faísca entre os eléctrodos da vela, de acordo com o descrito no ponto anterior [13].

No entanto, quando bolsas com uma mistura de ar e combustível explodem espontaneamente na câmara antes da chama, ocorre um fenómeno indesejável chamado *detonação*. Como consequência há um aumento de pressão local seguido de vibrações, as quais são atenuadas apenas quando a pressão é igual em todos os pontos da câmara de combustão. Este fenómeno é acompanhado por um ruído característico e similar a uma série de estalares metálicos [15]. A Figura 9 mostra a câmara de combustão em condições normais e em condições de detonação, assim como os respectivos gráficos de pressão ao longo de um ciclo de funcionamento do motor.

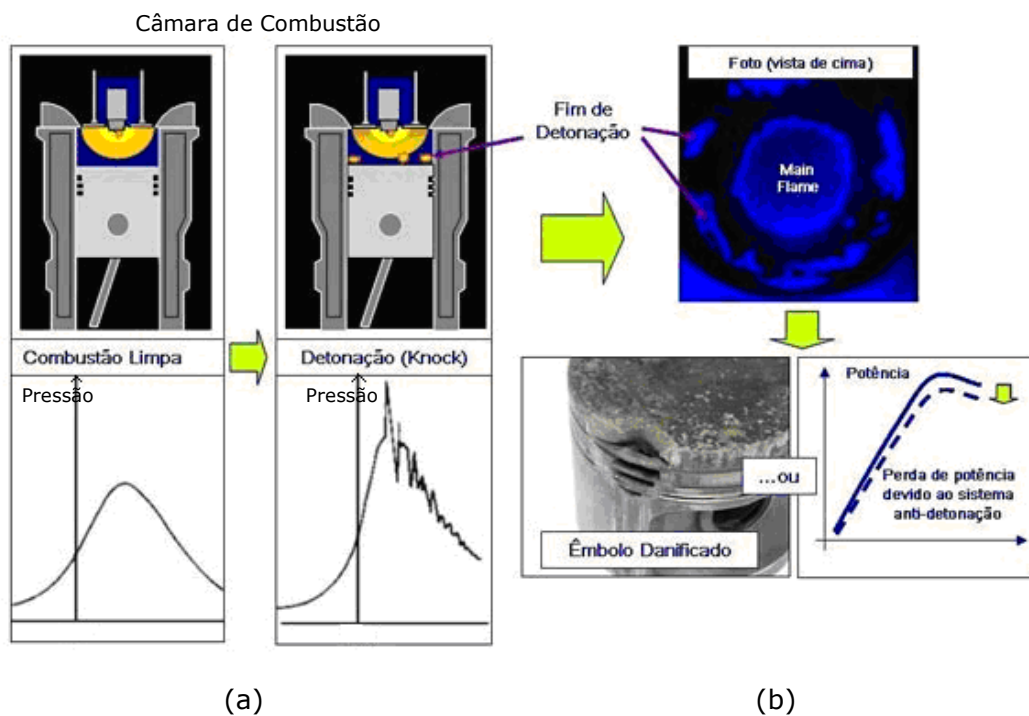


Figura 9 – (a) Diagrama de combustão em condições normais e de detonação.

(b) Consequências da detonação [15].

As detonações no motor levam a uma redução do desempenho, mas também podem danificar gravemente o motor. Podem causar sérios problemas

mecânicos, térmicos e incidentes destrutivos como a ruptura da cabeça do cilindro da junta, falha ou parcial fusão dos pistões, deterioração da cabeça dos cilindros e válvulas [13]. A [Figura 9](#) mostra um êmbolo danificado como consequência do fenômeno de detonação.

No caso dos motores mais recentes, o sistema de gestão do motor encarrega-se das detonações, dessincronizando o motor para evitar que isso ocorra. Obviamente isto significa que não ocorre detonação, no entanto o desempenho dos veículos reduz-se significativamente já que o motor está a proteger-se a si mesmo dos efeitos potencialmente nocivos da detonação, [Figura 9](#) [15].

4.3 DEFINIÇÃO DO NÚMERO DE OCTANO

Os pontos anteriores mostram como é importante que os combustíveis tenham grande resistência à auto-detonação. Qualitativamente são conhecidos os hidrocarbonetos com estruturas preferenciais nesta matéria: parafinas muito ramificadas, olefinas e aromáticos (benzeno, tolueno, xilenos) são muito resistentes à auto-detonação; inversamente as cadeias longas e sem ramificações de parafinas e olefinas, com mais do que quatro átomos de carbono, são mais susceptíveis aos fenômenos de auto-detonação [13].

De modo a caracterizar o comportamento dos combustíveis ou dos seus componentes relativamente à resistência à detonação durante a compressão e antes da ignição, sem envolver a composição química, foi introduzido e usado universalmente há mais de 50 anos o conceito de *Número de Octano* [13].

De acordo com este conceito, o combustível em teste é comparado com dois hidrocarbonetos puros escolhidos como combustíveis de referência primários. O primeiro é o 2,2,4-trimetilpentano ou isoctano, [Figura 2](#), o qual é muito resistente à auto-ignição e ao qual foi dado arbitrariamente o Número de Octano 100. O outro componente é o *n*-heptano, [Figura 2](#), o qual tem pouca resistência à detonação e ao qual foi atribuído o Número de Octano zero. Assim, um combustível com Número de Octano X , é uma mistura com X % v/v de isoctano e $(100-X)$ %v/v de *n*-heptano. As misturas do binário isoctano/*n*-heptano são consideradas combustíveis de referência [13].

A determinação do Número de Octano é feita recorrendo ao uso de um motor de referência designado por CFR (Cooperative Fuel Research) e existem dois processos padrão para a sua determinação, designados por RON (Research Octane Number) e MON (Motor Octane Number). A diferença entre os dois consiste essencialmente na velocidade de rotação, temperatura de admissão e avanço da faísca. O RON é mais usado do que o MON, e o valor obtido pelo MON é normalmente 10 a 12 pontos inferior ao RON. Esta diferença é designada por *sensibilidade* e é uma indicação da sensibilidade do combustível à modificação das condições experimentais, mais extremas (em particular ao aumento da temperatura) em que se processa o MON [13]. Convencionalmente as gasolinas têm um RON entre 90 e 100 e um MON entre 80 e 90 [13].

4.4 NÚMERO DE OCTANO vs DESEMPENHO DO MOTOR

Uma vez conhecido o Número de Octano do combustível, os motores devem ser construídos de modo a prevenir a detonação em quaisquer que sejam as condições de operação. As variáveis em causa são a razão de compressão e o avanço da ignição as quais determinam o desempenho do motor (eficiência térmica e potência). A potência pode sempre ser mantida por avanços tecnológicos como deslocamento de cilindros e razão de transmissão mas a eficiência térmica está sempre condicionada pelo Número de Octano. Assim quanto maior o Número de Octano, melhor será o seu desempenho e menor o consumo de combustível [13].

Verifica-se que a detonação depende de fenómenos físicos e químicos complexos e muito relacionados com o *design* do motor e as suas condições de operação. Assim não é possível caracterizar completamente o desempenho anti-detonante de uma gasolina com uma única medida, pois este está intimamente relacionado com o motor no qual é usada assim como com as condições operacionais do mesmo [10].

4.5 NÚMERO DE OCTANO DAS DIFERENTES FAMÍLIAS DE HIDROCARBONETOS

O RON dos hidrocarbonetos depende da sua estrutura química e varia com o ponto de ebulição de cada família, Figura 10. Para as *n*-parafinas, o RON é muito elevado para os componentes mais leves, e decresce com o aumento da cadeia alcançando o zero para o *n*-heptano. O RON, assim como o MON, aumenta sempre com o número e complexidade das ramificações laterais. O MON para parafinas é geralmente 2 a 3 pontos inferiores ao RON [13].

O RON para as olefinas, assim como para as parafinas, depende do comprimento e das ramificações da cadeia. O RON das olefinas é geralmente mais elevado do que o das parafinas para o mesmo esqueleto carbonado. A localização das ligações duplas no 'centro' da molécula tende a aumentar o RON, pelo menos para os primeiros membros da série. O MON das olefinas é sempre inferior ao correspondente RON numa média de 10 a 15 pontos, o que é explicado pela elevada energia de activação no processo cinético de auto-ignição. Na prática, o MON dos combustíveis comercializados correlaciona-se sempre fortemente com o conteúdo em olefinas [13].

Os compostos nafténicos têm sempre valores de RON e MON, superiores aos dos seus homólogos não-ciclícos. Por exemplo o RON do *n*-hexano é 24,8 enquanto o do ciclo-hexano é 83 [13].

O RON dos compostos Aromáticos é sempre superior a 100 (acima de 115-120). Estes valores estão numa ordem de magnitude em que há pouca precisão na determinação do RON. O MON dos Aromáticos também é superior a 100 mas sempre cerca de 10 pontos inferiores ao respectivo RON [13].

Para além dos hidrocarbonetos alguns compostos orgânicos oxigenados como álcoois e éteres são usados na formulação de gasolinas. São eles o metanol, etanol, propanol e butanol, assim como metil e etil éteres obtidos a partir de olefinas de C₄ e C₅: éter metil *terc*-butílico (MTBE), éter etil *terc*-butílico (ETBE) e o éter metil *terc*-amílico (TAME). Todos estes componentes são caracterizados por um elevado RON. O MON também é superior a 100 mas a diferença RON-MON é elevada, na ordem de 15 a 20 pontos. Os seus valores são geralmente estimados, uma vez que o procedimento do motor CFR não está adaptado a estas características. Estas substâncias têm um comportamento bastante diferente dos hidrocarbonetos derivados do petróleo.

Normalmente é mais importante conhecer o comportamento destes componentes na mistura do que no seu estado puro e o seu incremento nas gasolinas resulta sempre num ganho substancial no número de octano [13].

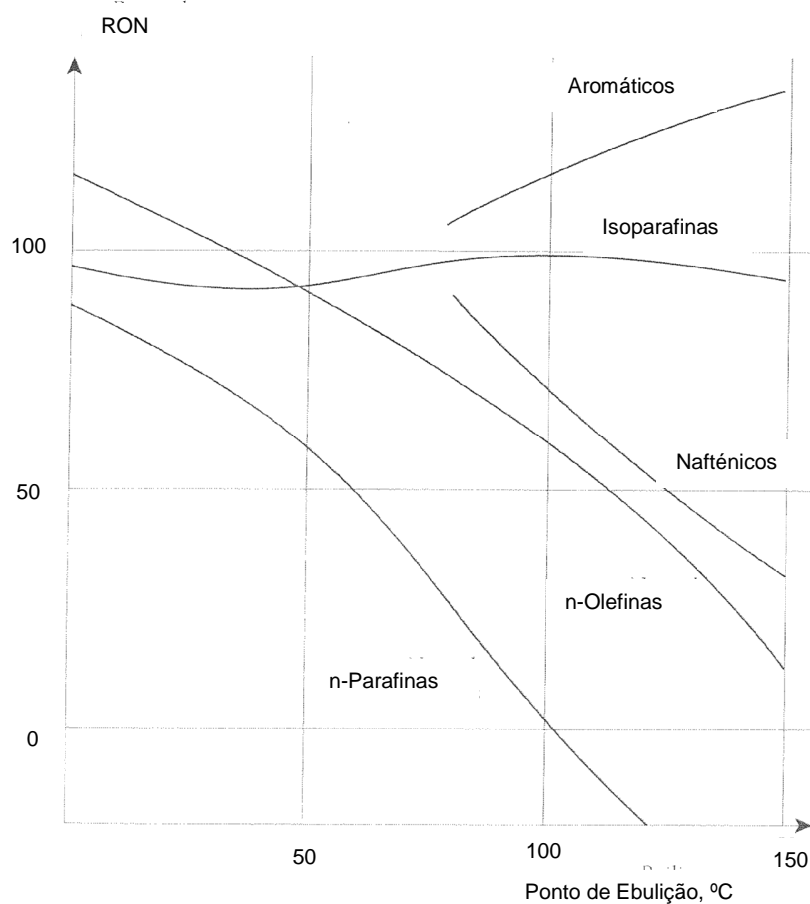


Figura 10 – Relação entre o RON e o ponto de ebulição das diferentes famílias de hidrocarbonetos [13].

5. MÉTODOS PARA A DETERMINAÇÃO DO NÚMERO DE OCTANO

5.1 MÉTODOS DE REFERÊNCIA- *ASTM D 2699* e *ASTM D 2700*

A razão anti-detonante da gasolina é medida num motor laboratorial de um único cilindro, de acordo com os métodos desenvolvidos para o efeito: *ASTM D 2699*- Caracterização do combustível pelo método *Research* (RON) e *ASTM D 2700*- Caracterização do combustível pelo método *Motor* (MON) [10].

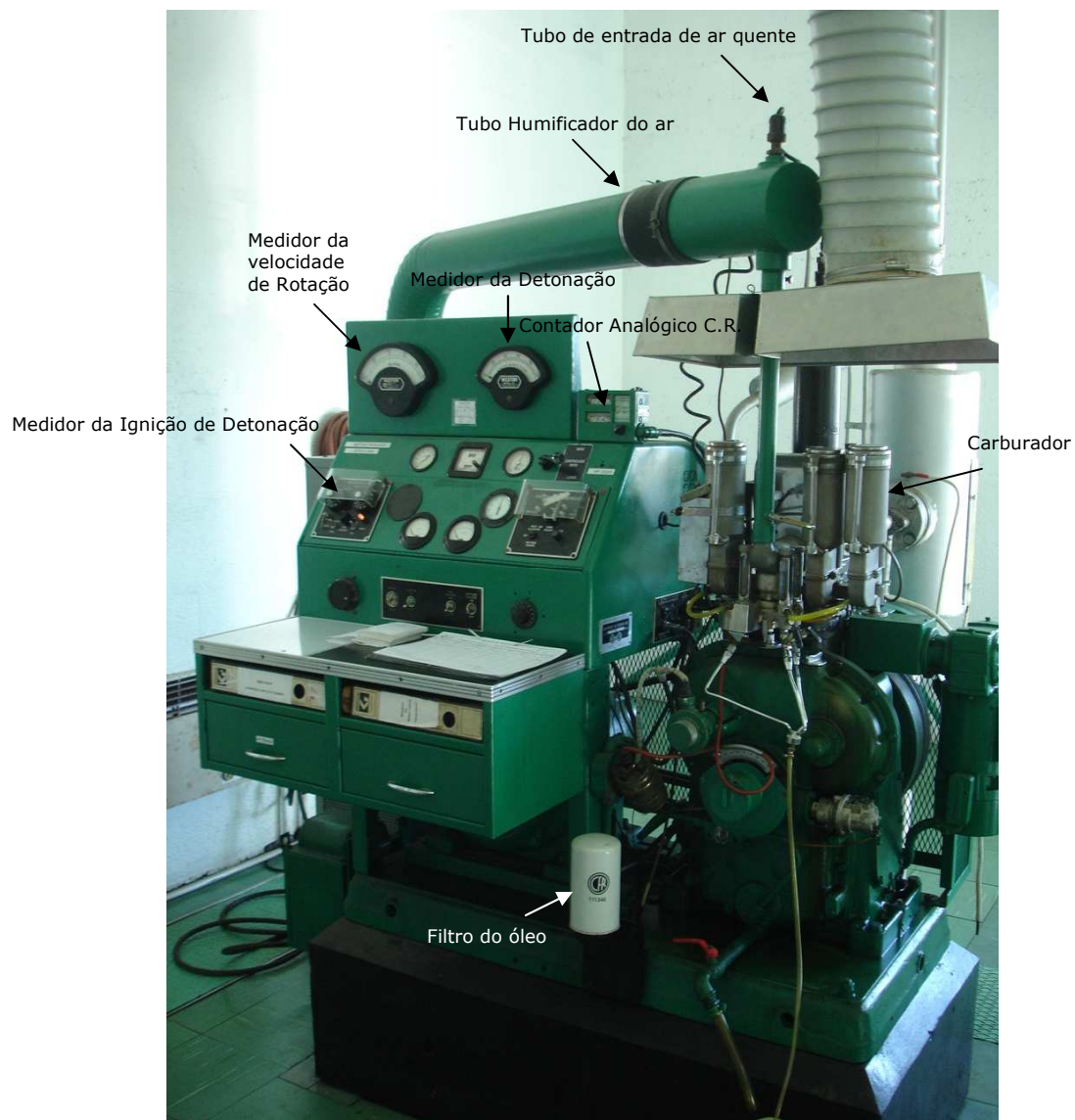


Figura 11 – Motor CFR do Laboratório da Refinaria do Porto.

O motor CFR (Cooperative Research Fuel), Figuras 11 e 12, funciona de acordo com as condições de operação definidas para cada método, usa um

único cilindro, um ciclo a quatro tempos e razão de compressão variável. A escala do número de octano é definida pela composição volumétrica das misturas Combustível de Referência Primárias (PRF). Estas misturas são preparadas a partir de isoctano e *n*-heptano, em proporções volumétricas de modo a obter misturas com números de octano de 0 a 100. A intensidade da detonação (KI) da amostra combustível é comparada com a de uma ou mais misturas PRF [16, 17].



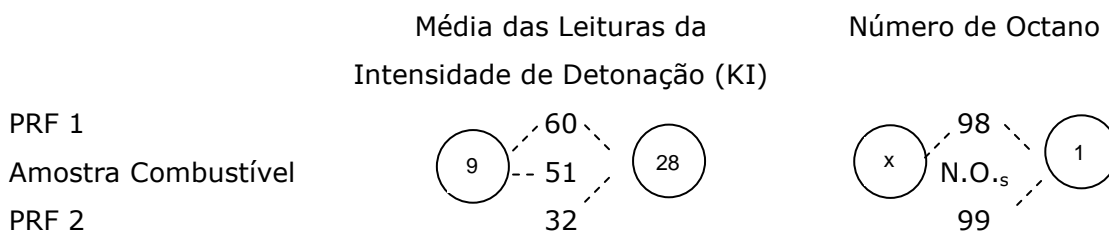
Figura 12 – Pormenores do Motor CFR: 1- Medidor da Velocidade de Rotação, 2- Medidor da Detonação, 3- Medidor da Ignição da Detonação, 4-Contador Analógico.

A escala do Número de Octano cobre a gama de 0 a 120 mas estes métodos têm uma gama de trabalho apenas de 40 a 120 [16,17].

A diferença entre os dois métodos consiste essencialmente na velocidade de rotação, temperatura de admissão do ar e da mistura. Durante a determinação do RON, o motor CFR opera a 600 ± 6 rpm, com temperatura de admissão do ar de $52 \pm 1^\circ\text{C}$ e sem pré-aquecimento do combustível. Em contraste o MON opera a 900 ± 9 rpm, com temperatura de admissão do ar de $38 \pm 2,8^\circ\text{C}$ e temperatura da mistura combustível de $149 \pm 1^\circ\text{C}$ [16,17].

O procedimento usado para a determinação do RON e MON é o da *Razão de Compressão*. Neste procedimento os valores de KI são estabelecidos a partir de PRF analisados com Números de Octano próximos da amostra a analisar e os vários parâmetros do motor são ajustados de acordo com as indicações do método em uso. A amostra a analisar é introduzida no carburador, o motor é colocado em funcionamento e o valor obtido no contador analógico é convertido em Número de Octano usando as '*Tabelas Guias de Intensidade de Detonação*' que constam nos métodos, para esse efeito [16,17].

Devem ser feitas duas determinações que não podem variar entre si mais do que 0,2, para ambos os métodos, e o Número de Octano é obtido por interpolação dessa média com as misturas padrão, como mostra o Esquema 1 [16,17].



$$N.O._s = N.O._{PRF1} + \left(\frac{K.I._{PRF1} - K.I._s}{K.I._{PRF1} - K.I._{PRF2}} \right) \cdot (N.O._{PRF2} - N.O._{PRF1}) = 98 + \left(\frac{60 - 51}{60 - 32} \right) \cdot (99 - 98) = 98.32$$

Esquema 1- Exemplo de cálculo da determinação do Número de Octano.

Em que:

Os valores dos círculos representam diferenças.

N.O._s= Número de Octano da amostra combustível;

N.O._{PRF1} = Número de Octano do PRF inferior;

N.O._{PRF2}= Número de Octano do PRF superior;

K.I._s= Intensidade de detonação da amostra combustível;

K.I._{PRF1}= Intensidade de detonação do PRF inferior;

K.I._{PRF2}= Intensidade de detonação do PRF superior.

A determinação do RON e MON usando o motor CFR é complexa e relativamente dispendiosa, uma vez que o motor requer manutenção constante e calibrações frequentes [16,17]. Nas gasolinas finais o número de

octano tem que ser determinado por estes métodos para cumprir as especificações e o produto poder ser colocado no mercado. No entanto para as correntes processuais e componentes de gasolina podem ser utilizados métodos alternativos mais rápidos e menos dispendiosos.

5.2 MÉTODOS ALTERNATIVOS

A complexidade da determinação do RON e MON pelo motor CFR, suscitou desde sempre interesse por parte de inúmeros investigadores, para o desenvolvimento de métodos alternativos à sua execução.

Uma vez que muitas das propriedades físicas e químicas das gasolinas estão directamente relacionadas com a sua composição, uma alternativa para a determinação do RON é o uso da composição da gasolina.

Existem várias técnicas analíticas para a determinação da composição das gasolinas, a partir da qual se pode calcular o RON. A desvantagem da maioria delas, dada a precisão necessária à obtenção do RON, é o facto da composição obtida ser pouco detalhada. Quanto mais detalhada for a composição obtida para a gasolina maior precisão é conseguida na previsão do RON [9].

Técnicas analíticas como a Adsorção com Indicador de Fluorescência, Infra-Vermelho e Ressonância Magnética Nuclear podem apenas fornecer informações limitadas sobre a composição correspondente a grupos estruturais como compostos aromáticos, olefinas e compostos saturados. Estas limitações analíticas na análise detalhada da composição fazem com que estes métodos não sejam tão precisos na previsão do número de octano, com a excepção dos casos em que a gasolina é constituída predominantemente por um único grupo estrutural [9].

Pelo contrário, os métodos que usam Cromatografia Gasosa (GC) com modernas colunas capilares, promovem uma excelente separação da maioria dos hidrocarbonetos constituintes da gasolina, e proporcionam a obtenção de dados de composição detalhados muito para além dos grupos estruturais, podendo por isso ser usados para a previsão mais precisa do número de octano [9].

Em 1955, *Schoen e Mrstik* [18] desenvolveram uma correlação gráfica para prever o número de octano para misturas binárias. Esta correlação tinha por

base o número de octano e a concentração em olefinas de cada um dos dois componentes que se pretendiam misturar. Este método foi redefinido, em 1959 por *Stewart* [19], e aplicado a misturas multi-componentes, com resultados mais consistentes.

Healy, et al. [20], em 1959, correlacionaram misturas de componentes de gasolina com diferentes níveis do número de octano e diferentes tipos de hidrocarbonetos em cada um dos componentes de gasolina. No entanto se este modelo for usado para prever o comportamento da mistura de novos componentes, podem ser obtidos valores pouco razoáveis, especialmente se o tipo de hidrocarbonetos ou o número de octano dos novos componentes estiver fora da gama dos componentes previamente testados.

Auckland e Charnock [21] desenvolveram, em 1969, um índice para o RON em misturas lineares obtido por combinação linear das propriedades molares parciais dos componentes que as constituem. Este índice de mistura é um valor hipotético obtido por extrapolação e a obtenção deste valor é análoga à obtenção da propriedade molar de uma solução. Desde que as propriedades molares parciais dependam da composição, é possível determinar o índice de mistura. No entanto esta aproximação é limitada, este método só pode ser usado para encontrar valores de componentes de mistura com uma composição particular e não pode ser usado para prever os seus valores de mistura em outras misturas [21].

Três anos mais tarde, em 1972, *Anderson et al.* [22], agruparam os compostos obtidos por cromatografia, na caracterização de gasolinas e componentes de gasolina, em 31 grupos, Tabela 4. Alguns desses grupos, consistiam apenas num componente enquanto outros eram misturas de componentes, agrupados pelo facto de terem estruturas químicas idênticas e por pertencerem a uma estreita gama de pontos de ebulição. Definiram um modelo para o cálculo do RON a partir da definição do RON numa mistura de *n*-heptano com o isoctano, eq.1 [22].

$$\text{RON} = a_1x_1 + a_2x_2 \quad (\text{eq.1})$$

Em que x_1 e x_2 são as fracções em volume do *n*-heptano e do isoctano na mistura e a_1 e a_2 são os valores do RON do *n*-heptano (RON=0) e do isoctano

(RON=100). E generalizaram esta equação para os restantes componentes da mistura, eq.2 [22].

$$RON = \sum_{g=1}^{g=31} b_g \cdot W_g \quad (\text{eq.2})$$

Em que:

W_g é a fracção em massa do grupo g na mistura.

b_g é o número de octano do grupo g na mistura.

No entanto, em misturas com outros componentes para além do n -heptano e do isoctano o RON efectivo dos compostos é função da sua concentração e da composição da mistura. O que significa que para se obter um RON preciso da mistura deve-se ter em conta os termos de interacção e os termos não-lineares. Porém devido à possível complexidade desses termos de interacção foram produzidos apenas modelos lineares simples. O que diferia entre cada um dos quatro modelos era a forma como os valores de b_g eram obtidos [22]:

➤ *Modelo 1 – RON dos componentes puros publicados*

O projecto API 45 [23] publicou tabelas com os números de octano para uma vasta gama de hidrocarbonetos. Estes foram usados para calcular os valores de b_g da eq.2. Nos grupos constituídos apenas por um componente o RON é igual ao apresentado pelo projecto API. Nos restantes casos os valores de RON foram estimados tendo em conta o RON dos componentes puros e a sua proporção na mistura [22].

➤ *Modelo 2 – RON das misturas publicados*

O projecto API 45 [23] também publicou o RON medido em misturas para uma elevada gama de hidrocarbonetos. Estes valores de RON foram calculados em misturas com 48%v/v de isoctano, 32%v/v de n -heptano e 20%v/v do hidrocarboneto em questão. Foi calculado o RON para cada grupo do mesmo modo mas usando agora estes novos valores de b_g [22].

➤ *Modelo 3 e 4 – RON efectivo estimado por regressão linear*

O número de octano da mistura, b_g , foi estimado por regressão linear múltipla a partir dos dados provenientes de 460 amostras, representativas de uma vasta gama de gasolinas finais e componentes de gasolina, Tabela 4 [22].

Os quatro modelos foram testados em 88 amostras cobrindo toda a gama de componentes de gasolina assim como de gasolinas finais.

Tabela 4 – Grupos usados para caracterizar as gasolinas (de acordo com a ordem de eluição) e coeficientes para estimar o RON das gasolinas [22].

Grupo	Componentes de cada grupo	Coefficiente de Regressão, b_g
1	componentes antes do <i>n</i> -butano	103,9
2	<i>n</i> -butano	88,1
3	<i>n</i> -butano a <i>i</i> -pentano (excluindo)	144,3
4	<i>i</i> -pentano	84,0
5	<i>i</i> -pentano a <i>n</i> -pentano (excluindo)	198,2
6	<i>n</i> -pentano	67,9
7	<i>n</i> -pentano a 2-metilpentano (excluindo)	95,2
8	2-metilpentano a 3-metilpentano (incluindo)	86,6
9	3-metilpentano a <i>n</i> -hexano (excluindo)	95,9
10	<i>n</i> -hexano	20,9
11	<i>n</i> -hexano a benzeno (excluindo)	94,9
12	benzeno	105,2
13	benzeno a 2-metil-hexano (excluindo)	113,6
14	2-metil-hexano a 3-metil-hexano (incluindo)	80,0
15	3-metil-hexano a <i>n</i> -heptano (excluindo)	97,8
16	<i>n</i> -heptano	-47,8
17	<i>n</i> -heptano a tolueno (excluindo)	62,3
18	tolueno	113,9
19	tolueno a 2-metilpentano (excluindo)	115,1
20	2-metil-heptano a 3-metil-heptano (incluindo)	81,7
21	3-metil-heptano a <i>n</i> -octano (excluindo)	109,7
22	<i>n</i> -octano	10,5
23	<i>n</i> -octano a etilbenzeno (excluindo)	96,1
24	etilbenzeno	122,6
25	etilbenzeno a <i>p</i> -xileno (excluindo)	45,4
26	<i>p</i> -xileno e <i>m</i> -xileno	102,0
27	<i>m</i> -xileno a <i>o</i> -xileno (excluindo)	73,3
28	<i>o</i> -xileno	123,6
29	fim <i>o</i> -xileno a fim do <i>n</i> -nonano	35,0
30	<i>n</i> -nonano a <i>n</i> -decano (excluindo)	112,0
31	<i>n</i> -decano e seguintes	85,6

Verifica-se que os Modelos 1 e 2 mostram erros sistemáticos, indicando que o RON efectivo é geralmente superior ao RON reportado pelo projecto API. Como era de esperar o RON das misturas apresentado no projecto API é

melhor do que o apresentado para os compostos puros pelo mesmo projecto. Em média o Modelo 1 apresenta resultados que são 10% abaixo do valor verdadeiro do RON, enquanto o Modelo 2 apresenta resultados apenas 2 a 3 % inferiores a este. Os Modelos 3 e 4 cujos coeficientes foram determinados por regressão linear são ambos melhores do que os Modelos 1 e 2 [22].

Verifica-se que equações lineares podem ser desenvolvidas e usadas no cálculo do RON. Este pode ser estimado a partir do RON dos compostos, do RON das misturas ou por regressão linear múltipla. O último método dá-nos valores mais realistas que embora não sejam suficientemente precisos para o controlo de qualidade das gasolinas finais, são muito úteis para cálculos processuais, controlo do processo de produção de componentes de mistura, entre outros [22].

Uma outra equação foi proposta por *Morris, et al.* [24] em 1975, que descreve o comportamento não linear, no que se refere às variáveis, em misturas de gasolinas.

$$\text{Número de Octano} = x_1a_1 + x_2a_2 + b_{12}x_1x_2 \quad (\text{eq.3})$$

Em que a_i é o número de octano do componente i , x_i a fracção em volume do componente i e b_{12} é o coeficiente de interacção dos componentes 1 e 2 [24].

No entanto existem algumas desvantagens em usar este modelo para descrever o comportamento da mistura de gasolina. Em primeiro lugar o parâmetro b_{12} é uma constante empírica que depende das características dos componentes 1 e 2, não depende só do tipo de componentes mas também do número de octano e da diferença entre o valor de octano de cada componente. Os valores de b_{12} podem variar entre o muito negativo e o muito positivo e podem alterar-se se as características dos componentes se alterarem. Os parâmetros binários obtidos pela equação só podem ser usados para os componentes a partir dos quais estes foram encontrados. É muito difícil generalizar estes parâmetros de interacção como função dessas propriedades. A segunda desvantagem é que quando novos componentes são adicionados a uma gasolina, novos parâmetros de interacção são necessários para descrever o comportamento do novo componente com os componentes presentes. Assim, é necessário um novo estudo de mistura com o novo componente de modo a determinar o termo de interacção adicional. Com n componentes, são

necessários $n(n-1)/2$ parâmetros de interação. Se quatro novos componentes forem adicionados a um mistura com 8 componentes o número de parâmetros de interação aumenta drasticamente de 28 para 66 [24].

Rusin, et al [25], em 1981 apresentaram um método de transformação complexo para estimar o RON para misturas de gasolinas a partir do RON dos seus componentes. Este método consiste em três passos:

- Transformação das propriedades dos componentes.
- Mistura linear dessas propriedades transformadas.
- Transformação inversa dos resultados.

Este método é similar ao do índice de mistura, no entanto pode causar inconsistência na transformação dos dados entre estes três passos [25].

Duran et al. [13], em 1987 desenvolveram uma técnica, por cromatografia gasosa, que lhes permitiu identificar e analisar cerca de 200 componentes de gasolina cujos números de octano eram conhecidos. Neste caso a primeira aproximação foi desenvolver um modelo linear, eq. 4 [13]:

$$RON = \sum (RON_{ppi} + K_i) \cdot C_i \quad (\text{eq. 4})$$

Em que RON é o número de octano da gasolina, RON_{ppi} o número de octano do componente i puro, C_i a concentração em % m/m e K_i o coeficiente que representa o desvio à idealidade. O valor de K_i é idêntico para componentes da mesma família de hidrocarbonetos com o mesmo grau de ramificação e com o mesmo número de átomos de carbono [13].

Zahed, et al. [26] propuseram em 1993 um modelo com 5 variáveis independentes para prever o número de octano. No entanto verifica-se que o modelo funcionava bem para as misturas usadas na regressão, mas tinha uma prestação pobre para outras misturas. As 5 variáveis derivadas deste modelo eram internamente inconsistentes. Os números de octano previstos a partir desta equação não eram muito próximos dos números de octano originais dos componentes da mistura. No caso do n -heptano o valor previsto a partir deste modelo era de 108,77 em oposição ao valor predefinido de zero. Similarmente o valor previsto para o isoctano por este modelo era de 108,95 versus o definido teoricamente de 100. O conjunto de variáveis obtidas por este modelo

pode apenas ser usado para um conjunto específico de componentes, na mesma gama de composições a partir das quais as variáveis foram definidas. Assim as vantagens do uso deste método são limitadas [26].

Vera e Luiz [27], desenvolveram um modelo para a determinação do RON e MON por regressão linear, a partir da análise cromatográfica (PIANO), usando o *Statgraphics* como ferramenta para o ajuste dos modelos propostos. Da análise cromatográfica foram detectados cerca de 300 componentes, o que tornava inviável o tratamento de todas as variáveis. Assim procederam à selecção das variáveis mais promissoras. A escolha destas variáveis foi feita com base na natureza química dos diferentes componentes e tendo em conta considerações teóricas para o seu agrupamento. O ponto de partida foi um modelo em que o número de octano da mistura era considerado o somatório dos números de octano dos diversos componentes. Devido ao grande número de componentes presentes na análise cromatográfica e tendo em conta as interacções verificou-se que é enorme o número de modelos que podem ser sugeridos [27]. Neste trabalho foram usadas 100 amostras de gasolinas provenientes de pelo menos 5 refinarias da *Petrobras* oriundas de processos de produção diferentes. Os dados foram avaliados por meio de histogramas de frequência das variáveis explicativas. Foram levados em conta os gráficos-caixa, para os valores de RON referentes a diversos hidrocarbonetos, usando-se os valores recomendados pelo *Manual de Dados Técnicos da Petrobras, Rio de Janeiro*. A divisão destes componentes em grupos foi feita de acordo com a sua natureza química: Parafinas (P), Olefinas (O), Nafténicos (N) e Aromáticos (A).

O uso de gráficos-caixa levou à escolha das variáveis explicativas do modelo, e a sua manipulação fez com que algumas delas fossem posteriormente desmembradas ou agrupadas até à obtenção do modelo considerado mais adequado (sete conjuntos diferentes de variáveis explicativas). Neste caso foram também utilizados os métodos de *stepwise* e de regressão múltipla para as tentativas de obtenção de regressão da variável RON em função dos componentes da análise cromatográfica. Foi sempre analisada a qualidade do ajuste, os gráficos de resíduos e a coerência dos valores dos coeficientes obtidos, obtendo o seguinte modelo [27]:

$$\text{RON} = 98,43 + 0,10 \cdot \text{AROM} - 0,18 \cdot \text{NAFRAR1} - 1,35 \cdot \text{NC7ENC8M} \quad (\text{eq.5})$$

Em que são usadas como variáveis explicativas para o melhor modelo: AROM, NAFFAR1 e NC7ENC8M [27].

Os resultados obtidos por este método foram comparados com os obtidos pelo método ASTM D2699 por meio do teste da média das diferenças. Os testes mostraram que os resultados obtidos pelo método cromatográfico podem ser considerados iguais aos obtidos pelo motor. Foram obtidas regressões consideradas adequadas, ou seja, foram criados modelos cujos gráficos dos resíduos não apresentavam qualquer tipo de estrutura, embora alguns dos valores absolutos fossem maiores do que a reprodutibilidade do método de obtenção do RON [27].

Recentemente, e tendo como base os contínuos avanços dos computadores, foram desenvolvidos alguns novos e promissores métodos não lineares, baseados em patentes reconhecidas, para a previsão do RON [9]. O novo método proposto por *N. Nikolaou et al*, [9] é usado para a previsão do RON em gasolinas de isomerização, eq. 6:

$$\text{RON}_G = \sum_{i=1}^N K_i \text{RON}_i y_i \quad (\text{eq.6})$$

Em que y_i é a fracção em volume do hidrocarboneto presente na gasolina, e obtido por GC, o RON_i é o RON do hidrocarboneto puro (publicado), Tabela 5, e K_i é o factor de peso que é calculado por:

$$K_i = \frac{\text{RON}_i}{\text{BRON}_i} \cdot \frac{\sum_{i=1}^N y_i \text{BRON}_i}{\sum_{i=1}^N y_i \text{RON}_i} \quad (\text{eq.7})$$

Em que BRON_i é o RON do hidrocarboneto na mistura (publicado), Tabela 5. O produto $K_i \text{RON}_i$ dá-nos um RON calculado efectivo ou da mistura, Tabela 5, característico para uma determinada gasolina, o qual também depende, do número, tipo e concentração dos hidrocarbonetos presentes na gasolina especificada [9].

Foram consideradas 20 amostras de produtos puros de isomerização de refinarias com diferente número de componentes e concentração de cada um. Foi determinado o RON pelo CFR e as mesmas amostras foram analisadas por

cromatografia. Foi detectado um número máximo de 26 hidrocarbonetos, e mais de 98% m/m da composição da gasolina foi determinada quantitativamente [9].

Tabela 5 – Números de Octano dos vários hidrocarbonetos presentes em gasolinas de isomerização [9].

Hidrocarbonetos	y _i	K _i	RON	RON	RON
			puro	mistura	calculado
<i>n</i> -butano	0,0043	0,876	93,8	113,0	82,1
2,2-dimetilpropano	0,0003	0,902	85,5	100,0	77,1
2-metilbutano	0,4507	0,974	92,3	100,0	89,9
<i>n</i> -pentano	0,1335	1,05	61,7	62,0	64,8
2,2-dimetilbutano	0,2342	1,088	91,8	89,0	99,9
ciclo-pentano	0,0116	0,758	101,3	141,0	76,8
2,3-dimetilbutano	0,0270	1,137	103,5	96,0	117,7
2-metilpentano	0,0490	0,944	73,4	82,0	69,3
3-metilpentano	0,0200	0,914	74,5	86,0	68,1
<i>n</i> -hexano	0,0260	1,377	24,8	19,0	34,1
2,2-dimetilpentano	0,0004	1,100	92,8	89,0	102,1
metil ciclo-pentano	0,0113	0,900	91,3	107,0	82,2
2,4-dimetilpentano	0,0003	1,153	83,1	76,0	95,8
2,2,3-trimetilbutano	0,0001	1,056	112,1	112,0	118,4
3,3-dimetilpentano	0,0002	1,015	80,8	84,0	82,0
ciclo-hexano	0,0249	0,796	83	110,0	66,1
2-metil-hexano	0,0006	1,118	42,4	40,0	47,4
2,3-dimetilpentano	0,0003	1,102	91,9	88,0	101,2
1,1-dimetil ciclo-pentano	0,0002	1,014	92,3	96,0	93,6
3-metil-hexano	0,0005	0,979	52	56,0	50,9
1,3 dimetil- <i>cis</i> -ciclopentano	0,0002	0,852	79,2	98,0	67,5
1,3 dimetil- <i>trans</i> -ciclopentano	0,0002	0,945	80,6	90,0	76,1
3-etilpentano	0,0002	1,071	65	69,6	69,6
<i>n</i> -heptano	0,0002		0	0	0
metil ciclo-hexano	0,0039	0,759	74,8	104	56,7
etil ciclo-hexano	0,0001	1,158	67,2	61,2	77,8

Verifica-se que os valores encontrados para o RON estão dentro do intervalo de $\pm 0,5$ RON do RON(ASTM) medido. Verifica-se que o RON(ASTM) do motor CFR é na maioria dos casos superior ao obtido a partir dos dados do GC e esta diferença pode dever-se a perda de informação do GC, uma vez que a composição determinada quantitativamente é inferior a 100% m/m. A desvantagem é que o RON da mistura calculado por este método tem uma correlação negativa com o RON das misturas apresentadas no projecto API 45 e ambos são comparados com os valores puros de RON. No entanto este é o efeito dos factores de peso K_i , que dependem não só da composição específica das gasolinas mas também da mistura dos RON's [9].

Das várias metodologias apresentadas são evidentes duas grandes desvantagens. A primeira é que embora os métodos baseados em análises de GC nos possam dar dados de composição detalhados, o agrupamento dos hidrocarbonetos analisados em grupos com estruturas similares, para a previsão do número de octano, automaticamente remove a maior vantagem oferecida por esses métodos. Este agrupamento é sempre baseado no incerto pressuposto de que os grupos estruturais se comportam do mesmo modo durante a combustão de diferentes gasolinas. A outra desvantagem dos métodos propostos é o facto dos valores de RON dos componentes puros e das misturas terem sido obtidos sob condições específicas e serem aplicados de igual modo para todos os tipos de misturas de gasolinas. Por exemplo, os números de octano determinados no projecto API 45, no caso das misturas, foram estimados a partir da análise de misturas de 48% v/v isoctano, 32% v/v *n*-heptano e 20% v/v do hidrocarboneto específico. Deve ser reconhecido que esse número de octano nos dá alguma informação qualitativa em termos do desempenho do RON, relacionado com esse hidrocarboneto específico nessa mistura. No entanto a sua extensão para qualquer número, tipo e concentração de hidrocarbonetos presentes na gasolina pode ser questionável.

5.3 APLICAÇÃO DOS MÉTODOS ALTERNATIVOS AO REFORMADO

Dos modelos apresentados nos vários trabalhos publicados, verifica-se que em alguns deles não são mencionados todos os dados necessários à sua aplicação em outras amostras. Outros por sua vez, embora apresentem os dados

necessários e bons resultados, nas amostras a partir das quais foram desenvolvidos, quando aplicados em novas amostras, já não apresentam resultados satisfatórios [20].

No trabalho publicado por *Anderson et al.* [22] os vários compostos detectados na análise cromatográfica foram agrupados em 31 grupos, de acordo com as suas estruturas químicas e ordem de eluição. Verifica-se que apresentam uma separação muito diferente da obtida com o equipamento disponível para o desenvolvimento da presente dissertação. Para além disso verifica-se que as matrizes de gasolina usadas são diferentes da matriz de Reformado. As gasolinas são constituídas apenas por *i-parafinas*, *n-parafinas* e compostos *aromáticos*, como consequência alguns dos compostos que constituem o Reformado não são apresentados no trabalho publicado. É o caso das *olefinas* e dos compostos *nafténicos*, que representam cerca de 2%*m/m* da composição das amostras de Reformado. Deste modo os modelos desenvolvidos não são aplicáveis na previsão do RON das amostras de reformado, objecto de estudo desta dissertação.

Vera e Luiz [22], desenvolveram o seu trabalho em amostras de gasolinas e classificaram inicialmente os compostos em grupos, de acordo com a sua natureza química: *Parafinas (P)*, *Olefinas (O)*, *Nafténicos (N)* e *Aromáticos (A)*. No entanto no processo de escolha das variáveis explicativas e sua manipulação, estas foram desmembradas ou agrupadas de um modo não muito explícito. No final as variáveis explicativas para o melhor modelo são: *AROM*, *NAFPAR1* e *NC7ENC8*. Tendo por base a descrição no trabalho publicado não é possível rastrear quais os componentes que fazem parte de cada uma das três variáveis explicativas, não sendo por isso possível aplicar o modelo proposto a outras amostras.

No caso do trabalho de *N. Nikolaou et al* [9], desenvolvido em amostras de gasolinas de isomerização, verifica-se que os componentes que constituem essas gasolinas são diferentes dos que constituem as amostras de Reformado. Estas gasolinas não têm compostos Aromáticos que representam em média 67%*m/m* da composição das amostras de Reformado. Por este motivo não é também possível aplicar o modelo desenvolvido neste trabalho às amostras de Reformado em estudo.

Os restantes trabalhos publicados, apresentam essencialmente os resultados finais e o modo de proceder sem que apresentem dados suficientes que permitam a aplicação dos modelos em outras amostras, nomeadamente nas amostras de Reformado. Assim, o ponto seguinte descreve o método proposto no presente trabalho para determinar o RON em amostras de Reformado, a partir da composição determinada por cromatografia.

6. METODOLOGIA PROPOSTA NESTE TRABALHO PARA A DETERMINAÇÃO DO RON

Embora o GC ofereça informação detalhada sobre a composição da gasolina, o problema mais comum para o cálculo preciso do número de octano é a contribuição não-linear das características dos componentes individuais da amostra, a qual é normalmente muito complexa. Assim, o número de octano determinado para cada componente não pode simplesmente ser combinado linearmente, de acordo com a composição da gasolina, para o cálculo do número de octano da mistura. A combinação não-linear dos componentes puros é o mais complexo para o cálculo preciso do número de octano.

O que se propõe neste trabalho é o desenvolvimento de um modelo de regressão linear múltipla para calcular os valores de RON em amostras de **Reformado**, a partir dos resultados da análise cromatográfica das mesmas. Este modelo não usa os valores de RON publicados (para hidrocarbonetos puros ou em misturas) para deste modo evitar o efeito da contribuição não-linear das características dos hidrocarbonetos individuais na mistura. Uma vez que em todas as amostras de Reformado nas quais é determinado o número de octano (RON) também é determinada a composição por cromatografia, esta nova metodologia apresenta desde já a vantagem de reduzir o número de análises necessárias a uma única - a análise cromatográfica. A esta acrescenta-se o facto da análise cromatográfica ser mais rápida, mais económica e ambientalmente preferível devido ao reduzido volume de amostra a manipular.

Esta nova metodologia poderá ser extendida a outras correntes processuais ou componentes de gasolina, se se mostrar vantajosa.

6.1 AMOSTRAS

As amostras estudadas neste trabalho, num total de 375, são de *Reformado* proveniente da Fábrica de Combustíveis da Refinaria do Porto da Galp Energia. Foram recolhidas no período compreendido entre 2003 e 2006, três vezes por semana. Todos os dados referentes a estas amostras são apresentados no Ficheiro 4: anexos, do CD-ROM anexo à dissertação. O critério de selecção das

amostras, durante este período de tempo, foi a análise simultânea da composição por cromatografia e do número de octano pelo motor CFR (método ASTM D 2699). Todo o trabalho laboratorial foi realizado em rotina, pelos analistas do laboratório, e fora do contexto desta pós-graduação.

6.2 DADOS DO MOTOR CFR

Todas as amostras consideradas neste trabalho foram analisadas de acordo com o método ASTM D 2699 – Determinação do Número de Octano (RON). Para tal foi usado um motor CFR (Cooperative Research Fuel), Figura 11, a 600 ± 6 rpm, temperatura de admissão do ar de $52 \pm 1^\circ\text{C}$, sem pré-aquecimento do combustível e usando o procedimento da *Razão de Compressão*. Para cada amostra foram feitas duas determinações que não variaram entre si mais do que 0,2 e o RON foi obtido por interpolação dessa média com as misturas padrão, como mostra o exemplo do Esquema 1 [16,17].

6.3 DADOS CROMATOGRÁFICOS

A composição das amostras de Reformado foi determinada por cromatografia gasosa multidimensional. O laboratório dispõe de um cromatógrafo, Hewlett-Packard 5890 com Tecnologia AC PIONA, que separa a amostra nas suas 5 famílias de compostos (*n*-parafinas, *i*-parafinas, olefinas, compostos nafténicos e compostos aromáticos), por número de carbonos, Figura 13.

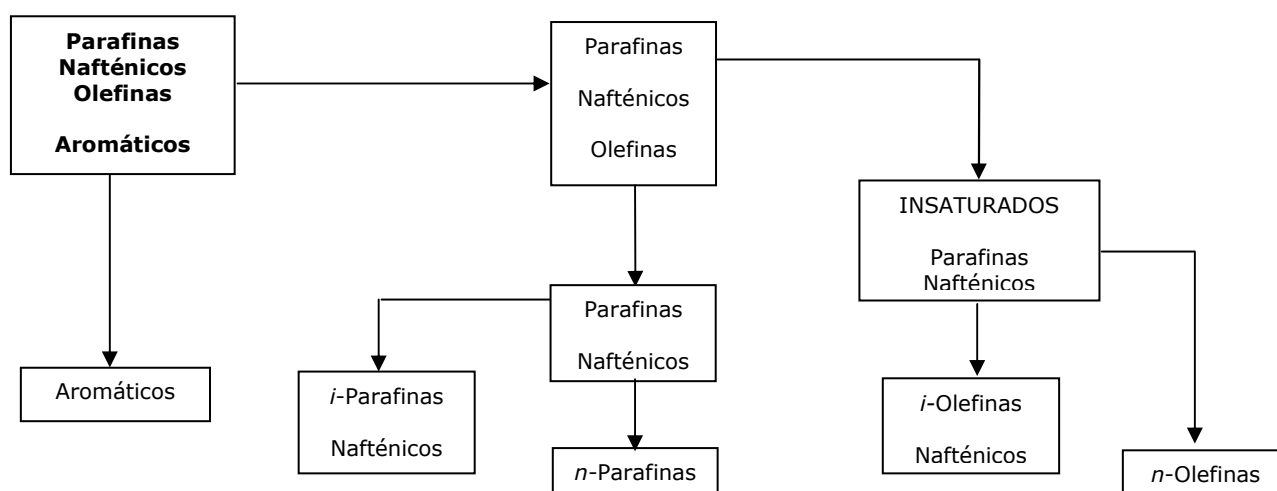


Figura 13 – Diagrama de separação do Cromatógrafo PIONA [28].

Este sistema é constituído por um conjunto de válvulas e colunas que permitem a separação desejada, sendo que cada coluna, Tabela 6, tem um propósito específico.

Tabela 6 – Descrição de cada uma das colunas do cromatógrafo PIONA [28].

Coluna	Tipo
OV275	Coluna de enchimento com 3,0mx2,0mm (d.i.), com 30% OV275 em <i>Chromosorb PAW 60-80 mesh</i>
HP-1	Coluna capilar, 15mx0,53mm (d.i.) com 5mm HP-1 (<i>Crosslinked MetilSiloxano</i>)
Trap Tenax	Coluna de enchimento, 0,15mx2,5mm (d.i.)
Trap de Olefinas	Coluna de enchimento, 0,30mx2,5mm (d.i.)
5A <i>Molecular Sieve</i>	Coluna de enchimento, 0,10mx2,5mm (d.i.)
13X <i>Molecular Sieve</i>	Coluna de enchimento, 1,80mx1,7mm (d.i.)
Pt	Catalizador Hidrogenador de Platina

A *Coluna OV-275* é usada como uma pré-coluna de separação, separa a amostra em três fracções as quais são analisadas posteriormente noutras colunas do sistema. Esta coluna retarda especialmente a eluição dos compostos aromáticos: o benzeno (ponto de ebulição de 80°C) elui depois do dodecano (ponto de ebulição 200°C). Isto significa que a fracção da amostra que elui antes do benzeno contém parafinas e compostos nafténicos com pontos de ebulição até 200°C. Estes componentes são separados sequencialmente no cromatógrafo pela *Trap de Olefinas*, 5A e 13X. O eluente proveniente da coluna OV-275 com início no benzeno é separado em duas fracções. Estas fracções são transferidas para a coluna HP-1 por via da *Trap Tenax* [28].

A *Trap Tenax* absorve, focaliza e transfere as fracções de aromáticos da coluna OV-275 para a HP-1. Esta coluna absorve essa fracção à temperatura de 70°C e desorve para a coluna HP-1 à temperatura de 280°C [28].

A *Coluna HP-1* é usada para separar os aromáticos com baixos pontos de ebulição das parafinas e compostos nafténicos com elevado ponto de ebulição. Devido ao número limitado de compostos aromáticos com ponto de ebulição abaixo dos 200°C a separação pode ser feita por diferença dos números de

carbono dos compostos aromáticos. Os componentes com ponto de ebulição abaixo dos 200°C são enviados de volta, da coluna para o detector [28].

A *Trap de Olefinas* é usada para separar as olefinas dos compostos saturados. A fracção de hidrocarbonetos com ponto de ebulição até 200°C (C₄ a C₁₁) elui da coluna OV-275 para a *Trap de Olefinas*. As olefinas são retidas enquanto os compostos saturados passam pela coluna; de seguida, e por aquecimento, a *Trap* liberta as olefinas. A sua selectividade depende da concentração e distribuição dos componentes olefínicos. Para amostras que contenham mais de 15% de olefinas a selectividade é superior ao número de carbono sete. O resultado é a separação entre saturados C₁₁ e olefinas C₄. A *Trap* tem uma temperatura de adsorção entre 90-160°C e de dessorção de 280°C [28].

A *Coluna do Catalisador Hidrogenado de Platina* hidrogena todos os componentes olefínicos à sua forma saturada antes de entrarem nas colunas 5A e 13X. Uma vez que estas colunas só actuam com compostos saturados, caso algum componente olefínico elua desta coluna irá resultar em picos fantasmas. A hidrogenação funciona a temperaturas entre os 180-220°C, geralmente é escolhida a temperatura de 220°C e reduzida a um mínimo de 180°C quando a actividade da coluna está muito elevada [28].

A *Coluna 5A Molecular Sieve* é usada para separar as *n*-parafinas das *i*-parafinas na gama do C₄ ao C₁₁. Funciona como uma *trap* reversível para as *n*-parafinas nesta gama de número de carbono. À temperatura de adsorção as *i*-parafinas e os compostos nafténicos passam pela coluna sem serem impedidos enquanto as *n*-parafinas são retidas e mais tarde são dessorvidas a uma temperatura mais elevada. A temperatura de adsorção varia entre 110-160°C e a de dessorção é de 450°C [28].

A *Coluna 13X Molecular Sieve* é usada para separar as parafinas dos compostos nafténicos por número de carbono, o ponto de ebulição dos compostos a serem analisados é de 270°C [28].

Depois de separados, os compostos são enviados para o detector de ionização de chama (FID) que os detecta e posteriormente quantifica, a partir de factores de resposta baseados na percentagem em massa de carbono. Considera-se que o metano tem um factor de resposta unitário e o factor de resposta de cada componente é calculado com base na massa atómica do carbono 12,011 e do hidrogénio 1,0080, assim:

$$\text{Factor Resposta} = \frac{0,7487x \frac{(C_{aw} \cdot C_n) + (H_{aw} \cdot H_n)}{C_n}}{C_{aw}} \quad (\text{eq.8})$$

Em que C_{aw} é a massa atómica do carbono, C_n é o número de átomos de carbono na molécula, H_{aw} é a massa atómica do hidrogénio, H_n é o número de átomos de hidrogénio na molécula e 0,7487 é a correcção do factor de resposta do metano para a unidade.

A Figura 14 mostra a separação obtida para uma das amostras de Reformado, usada para o modelo, analisada pelo cromatógrafo PIONA e, cujas concentrações estão descritas na Tabela 7.

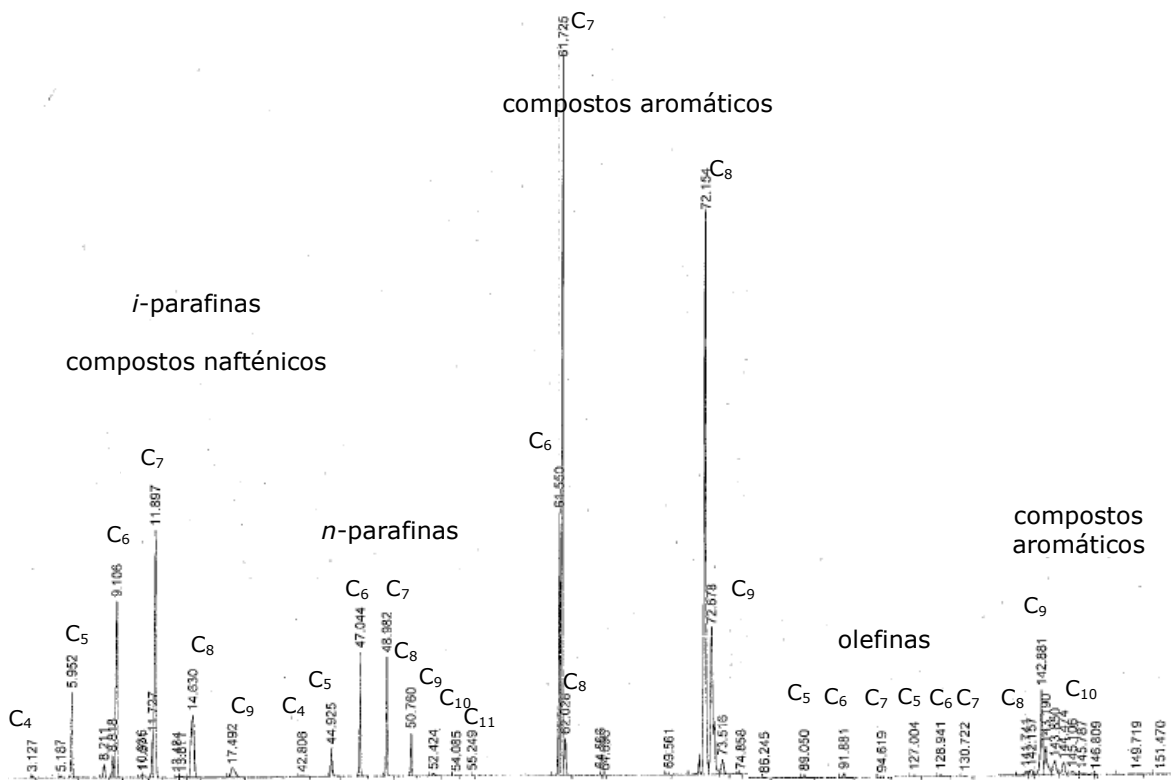


Figura 14 – Exemplo de um cromatograma de uma amostra de Reformado usada no modelo.

Tabela 7 – Exemplo da composição de uma amostra de Reformado usada no modelo.

Composto	%, m/m	Composto	%, m/m		
<i>i</i> -Parafinas	C4	0,00	C4	0,00	
	C5	2,52	C5	0,00	
	C6	8,33	C6	0,10	
	C7	8,57	C7	0,22	
	C8	2,61	C8	0,00	
	C9	0,11	C9	0,00	
	C10	0,08	C10	0,00	
	C11	0,03	Total	0,32	
Total	22,25	<i>n</i> -Olefinas	C4	0,00	
<i>n</i> -Parafinas	C4		0,00	C5	0,00
	C5		1,41	C6	0,03
	C6		3,44	C7	0,00
	C7	2,20	Total	0,03	
	C8	0,69	Nafténicos	C5	0,19
	C9	0,58		C6	0,55
	C10	0,00		C7	0,28
C11	0,00	C8		0,11	
Total	8,32	C9		0,00	
<i>i</i> -Olefinas	C4	0,00	C10	0,00	
	C5	0,00	Total	1,13	
	C6	0,21	Aromáticos	C6	5,69
	C7	0,35		C7	17,02
	C8	0,00		C8	23,08
	C9	0,00		C9	17,63
C10	0,00	C10		2,87	
Total	0,56	Total	66,29		

6.4 ANÁLISE ESTATÍSTICA DOS DADOS

6.4.1 Programa Informático para Estatística – SPSS

O SPSS (Statistical Package for the Social Sciences), é por tradição um programa usado para a análise estatística em ciências sociais [29]. No entanto este programa também é usado nas ciências exactas, como no caso da presente dissertação onde toda a análise estatística foi executada recorrendo ao SPSS.

Este programa faz a análise de dados em 4 etapas [29]:

- Introduzem-se as variáveis da amostra no editor de dados
- Seleccionam-se os procedimentos de análise de um dos menus
- Seleccionam-se as variáveis a analisar
- Examinam-se e interpretam-se os resultados

Na primeira etapa é necessário ter particular atenção ao tipo de variáveis em estudo (escala de medida). Na segunda etapa seleccionam-se os procedimentos estatísticos a utilizar. Depois de seleccionadas as variáveis e as opções dos diversos campos obtêm-se os resultados. Na interpretação dos resultados é necessário ter em conta que o programa não testa *à priori* os pressupostos que é necessário que sejam verificados para utilizar determinados testes [29].

6.4.2 O modelo de regressão linear [29]

A regressão linear é uma técnica estatística usada para modelar relações entre variáveis e prever a partir de um conjunto de variáveis independentes o valor da variável dependente. No modelo de regressão linear múltipla, a relação funcional entre uma variável dependente Y e as várias variáveis independentes X_i ($i=1, \dots, p$) é do tipo:

$$Y_j = \beta_0 + \beta_1 \cdot X_{1j} + \beta_2 \cdot X_{2j} + \dots + \beta_p \cdot X_{pj} + \varepsilon_j \quad (j=1, \dots, n) \quad (eq.9)$$

Em que:

β_i –são os coeficientes de regressão.

ε_j - erros aleatórios ou resíduos do modelo.

As hipóteses subjacentes a este modelo são:

- O termo ε_j reflecte apenas os erros da medição e a variação natural em Y_j . Os valores de X_i são encarados como constantes predeterminadas, sem erro.
- Os erros ε_j são aleatórios, independentes, têm valor esperado nulo, variância constante (σ^2) e são normalmente distribuídos, isto é, $\varepsilon_j \sim N(0, \sigma I)$.
- As variáveis independentes são ortogonais (não estão correlacionadas, ou quanto muito apresentam apenas correlações fracas).

6.4.3 Estimação dos coeficientes de regressão – método dos mínimos quadrados [29]

Os parâmetros $\beta_0, \beta_1, \dots, \beta_p$ do modelo podem ser calculados a partir de um conjunto de observações $(X_{1j}, X_{2j}, \dots, X_{pj}, Y_j)$ recorrendo ao método dos mínimos quadrados. Neste método, as estimativas dos coeficientes de regressão são obtidas de modo a que os erros ou resíduos do modelo de regressão linear, obtidos por $e_j = y_j - \hat{y}_j$ sejam mínimos. A partir do somatório do quadrado de e_j é possível avaliar as diferenças existentes entre os valores estimados pela recta (\hat{y}_j) e os valores observados (y_j). Assim a estimação dos parâmetros $\beta_0, \beta_1, \dots, \beta_p$ do modelo, resume-se à determinação do mínimo da função designada por Soma dos Quadrados dos Erros ou Resíduos do modelo (SQE), relativamente a cada um dos coeficientes de regressão, eq.10.

$$SQE = \sum_{j=1}^n e_j^2 \quad (\text{eq. 10})$$

O modelo em forma matricial é:

$$y = X\beta + \varepsilon \quad (\text{eq. 11})$$

em que:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad \text{é o vector das } n \text{ observações da variável dependente, (eq.12)}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{bmatrix} \quad \text{é o vector dos } p+1 \text{ coeficientes do modelo, (eq.13)}$$

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{p1} \\ 1 & x_{12} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{pn} \end{bmatrix} \quad \text{é a matriz das } n \text{ observações de } X_i \text{ (} i=1, \dots, p \text{), (eq.14)}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \text{ é o vector dos erros com } \varepsilon \sim N(0, \sigma I), \text{ (eq.15)}$$

A minimização da soma dos quadrados dos erros é dada por $(y - X\beta)'(y - X\beta)$ e obtém-se para o vector $\hat{\beta}$ que satisfaça a condição:

$$\frac{\partial}{\partial \hat{\beta}} [(y - X\hat{\beta})'(y - X\hat{\beta})] = 0 \quad (\text{eq.16})$$

A solução desta equação diferencial, resolvendo em ordem a $\hat{\beta}$ é:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (\text{eq.17})$$

6.4.4 Inferências sobre o modelo de regressão linear

6.4.4.1 A análise de variância do modelo de regressão linear [29]

Depois de desenvolvida a regressão linear é necessário avaliar, a partir das estimativas amostrais, se alguma das variáveis independentes pode ou não influenciar a variável dependente, isto é, se o modelo ajustado é ou não significativo:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad \text{vs.}$$

$$H_1: \exists i: \beta_i \neq 0 \quad (i=1, \dots, p)$$

(H_1 lê-se como "existe pelo menos um i tal que β_i é diferente de zero")

Para o teste destas hipóteses, a variabilidade total em Y , que é expressa pela soma dos quadrados dos desvios de Y em relação a \bar{Y} (SQT), divide-se em duas componentes. Uma delas é explicada pelo modelo ajustado (SQR) e a outra não é explicada pelo modelo mas sim pelos erros (SQE), eq. 18.

$$\sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2 + \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 \quad (\text{eq. 18})$$

$$SQT = SQR + SQE$$

Se a variância total em Y , explicada pelo modelo de regressão for significativamente maior do que a variância total não explicada pelo modelo,

então o modelo ajustado é significativo. A variância explicada pela regressão é estimada dividindo a SQR pelos respectivos graus de liberdade p , e a variância dos erros pode ser estimada dividindo a SQE pelos respectivos graus de liberdade $(n-p-1)$:

$$F = \frac{SQR/p}{SQR/(n-p-1)} = \frac{QMR}{QME} \quad (\text{eq.19})$$

Em que QMR e QME são respectivamente os quadrados médios da regressão e dos erros. O SPSS cria um p -value para esta estatística de teste, em que se p -value $\leq \alpha$, rejeitamos a H_0 a favor da H_1 . Concluindo deste modo que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente, logo o modelo ajustado aos dados é significativo. É no entanto necessário analisar se todas ou apenas algumas variáveis independentes influenciam a variação da variável dependente.

6.4.4.2 Testes aos coeficientes do modelo de regressão [29]

Para saber quais dos β_i são diferentes de zero é necessário testar os β_i , de acordo com as seguintes hipóteses:

$$\begin{aligned} H_0: \beta_i &= k & \text{vs.} \\ H_1: \beta_i &\neq k & (i=1, \dots, p) \end{aligned}$$

Em que k pode tomar qualquer valor, no entanto na maioria dos *software* de análise estatística $k=0$. A estatística de teste é:

$$T_{\beta_i} = \frac{b_i - k}{\sqrt{s_{y,x}^2 \times C_{ii}}} \quad (\text{eq.20})$$

Em que b_i é a estimativa de β_i , $s_{y,x}^2 = \text{QME}$ é a estimativa da variância do modelo e C_{ii} é o elemento i da diagonal da matriz $(X'X)^{-1}$ correspondente a β_i . Para H_0 , T_{β} tem uma distribuição t -student com $(n-p-1)$ graus de liberdade. O SPSS produz um p -value e rejeita-se H_0 se p -value $\leq \alpha$.

6.4.4.3 O coeficiente de determinação

Existem vários parâmetros que ajudam a qualificar o modelo de regressão como: o coeficiente de correlação (r), coeficiente de determinação (R^2) e o coeficiente de determinação ajustado (R_a^2).

O coeficiente de correlação mede a associação linear entre duas variáveis como por exemplo os valores conhecidos e os respectivos valores previstos por um modelo de calibração (y_j, \hat{y}_j). E neste caso serve para avaliar a proximidade entre os valores observados e os valores estimados pelo modelo. Varia entre -1 e +1, indicando nestes extremos a existência de uma forte correlação entre as duas variáveis (correlação negativa e correlação positiva respectivamente). Quando não há uma relação linear entre as duas variáveis então o valor de r distancia-se de -1 e +1 e é próximo de 0 (zero) [29].

O coeficiente de determinação (geralmente representado por R^2) é uma das medidas da qualidade do ajuste mais usadas, eq. 21.

$$R^2 = \text{SQR} / \text{SQT} \quad (\text{eq.21})$$

O R^2 mede a proporção da variabilidade total que é explicada pela regressão ($0 \leq R^2 \leq 1$). Quando $R^2 = 0$ o modelo não se ajusta aos dados, pelo contrário quando $R^2 = 1$ o ajuste é perfeito. O valor de R^2 para o qual se considera um ajuste adequado é muito subjectivo; é geralmente aceite um bom ajuste, no caso das ciências exactas, se $R^2 > 0,9$. No entanto o uso do R^2 como indicador da qualidade do ajuste deve ser usado com cuidado e não deve ser usado para comparar modelos cujo número de variáveis independentes seja muito diferente entre eles. Na maioria dos casos, o aumento do número de variáveis independentes provoca um aumento no R^2 mesmo que estas tenham pouca influência sobre a variável dependente [29].

Já o coeficiente de determinação ajustado R_a^2 , eq.22, pode ser usado sem este tipo de restrições, como indicador de qualidade do ajuste [29]. A adição de uma nova variável no modelo só leva ao aumento do R_a^2 se esta conduzir a um melhor ajuste do modelo aos dados, isto é se a variância dos erros (QME) diminuir relativamente à variância total (QMT) [29].

$$R_a^2 = 1 - \frac{QME}{QMT} = 1 - \frac{n-1}{n-p-1}(1-R^2) = R^2 - \frac{p(1-R^2)}{n-p-1} \quad (\text{eq. 22})$$

6.4.5 Validação dos pressupostos do modelo de regressão linear

O modelo de regressão linear só pode ser usado, depois de validado um conjunto de pressupostos do próprio modelo. Assim após a estimação dos coeficientes da regressão é necessário proceder à validação dos pressupostos respeitantes aos erros ou resíduos e à (quasi)ortogonalidade entre as variáveis independentes.

6.4.5.1 A análise de resíduos [29]

Os erros ou resíduos do modelo de regressão linear servem para estimar os coeficientes de regressão mas também para validar os pressupostos da sua aplicação. Assim a inferência sobre o modelo só é válida quando:

- $\varepsilon_j \sim N(0, \sigma)$ - os erros possuem distribuição normal de média nula e variância constante;
- $\text{Cov}(\varepsilon_k, \varepsilon_l) = 0$ ($k \neq l$; $k, l = 1, \dots, n$) - os erros são independentes.

A normalidade da distribuição dos erros pode ser feita por um procedimento gráfico. Idealmente os erros ou resíduos quando representados graficamente em função dos valores de \hat{y}_j têm uma distribuição aleatória em torno de $e_i = 0$. O pressuposto da distribuição normal dos erros também pode ser verificado através do gráfico de probabilidade normal, onde no eixo das abcissas se representa a probabilidade observada acumulada dos erros, e no eixo das ordenadas se representa a probabilidade acumulada que se observaria se os erros possuísem distribuição normal. Quando os erros possuem distribuição normal os valores representados neste gráfico distribuem-se em torno da diagonal principal. Também é possível testar a normalidade recorrendo a testes de ajustamento, como o caso do teste de Kolmogorov-Smirnov (K-S). O pressuposto da independência dos resíduos, pode ser verificado pelo teste de Durbin-Watson, em que se testa a presença de auto-correlação entre os erros ou resíduos do modelo de regressão linear. Assim se $d \approx 2$, concluí-se que

não existe auto-correlação entre os resíduos; para d muito menor do que 2 existe auto-correlação positiva, e para valores de d muito maiores do que 2 existe auto-correlação negativa. O SPSS não produz o p -value associado a esta estatística de teste, logo de uma forma mais ou menos empírica, considera-se que não existe auto-correlação para $d \approx 2,0 (\pm 0,2)$ [29].

6.4.5.2 Multicolinearidade [29]

O modelo de regressão linear múltipla pressupõe que as variáveis explicativas são linearmente independentes, isto é, não estão correlacionadas (são ortogonais), por isso não se verifica multicolinearidade. A intensidade da multicolinearidade é analisada essencialmente através de dois pontos:

- Correlação entre as variáveis independentes [29]

A partir da matriz das correlações das variáveis independentes, é possível verificar se estas estão ou não correlacionadas entre si. No entanto não existe um valor limite a partir do qual seja possível prever problemas de estimação do modelo por causa da colinearidade entre as variáveis independentes. Considera-se que correlações elevadas $|R| > 0,75$ normalmente conduzem a problemas de multicolinearidade [29].

- Tolerância e VIF

A Tolerância mede o grau em que uma variável X_i é explicada por todas as outras variáveis independentes, eq. 23.

$$T = 1 - R_i^2 \text{ (eq.23)}$$

A Tolerância varia entre 0 e 1, e quanto mais próxima estiver de zero, maior será a multicolinearidade. Normalmente considera-se como o limite a partir do qual há multicolinearidade 0,1 [31].

O inverso da Tolerância designa-se por VIF (Variance Inflation Factor) e quanto mais próximo estiver de zero menor será a multicolinearidade. Considera-se como limite acima do qual existe multicolinearidade 10 [31].

6.4.6 Selecção de regressores

Num problema de regressão linear múltipla podem-se conhecer à partida quais as variáveis independentes (regressores) a incluir no modelo de regressão.

Porém, principalmente na fase inicial da análise de regressão pode ser desconhecida qual ou quais as variáveis que conduzem ao melhor modelo e a decisão de escolha pode ser complicada pela presença de multicolinearidade e dos seus efeitos sobre a magnitude dos coeficientes de regressão. Os procedimentos de selecção de variáveis apresentam a vantagem de indicar com base num critério exacto quais as variáveis que apresentam relações mais fortes com a variável dependente. A selecção dos regressores pode ser feita por vários métodos:

6.4.6.1 Método Exaustivo [30]

Conceptualmente muito simples consiste em construir os modelos de regressão que combinem de todas as maneiras possíveis os potenciais regressores. Ordenar os modelos de acordo com o critério de qualidade escolhido e avaliar os modelos considerados melhores.

Este método apresenta no entanto uma limitação enorme que se prende com a dificuldade associada à construção de todos os modelos de regressão possíveis. Reconhecida esta limitação, foram propostos métodos alternativos que, de alguma forma, constituem uma aproximação ao método exaustivo.

6.4.6.2 Método Progressivo (ou *Forward*) [30]

Inicialmente são ajustados tantos modelos de regressão linear simples quantos os potenciais regressores. De entre os regressores, que isoladamente explicam uma proporção significativa da variação total da variável dependente, é incluído no modelo aquele que explica maior proporção. De seguida são construídos modelos de regressão linear dupla que associem, como variáveis independentes, o regressor seleccionado no passo anterior e cada um dos potenciais regressores ainda não incluídos. Neste processo está envolvida a re-estimação de parâmetros já calculados no passo anterior. O próximo regressor a adicionar ao modelo é aquele que explica uma proporção adicional significativa da variação total (adicional relativamente aquela que era explicada no passo anterior). Prossegue-se a tentativa de construção de modelos de ordem superior. O método termina quando nenhum dos potenciais

regressores ainda não incluídos no modelo, ao entrar nele, explicar uma proporção adicional significativa da variação total ou quando todos os potenciais regressores forem incluídos no modelo.

6.4.6.3 Método Regressivo (ou *Backward*) [30]

Este método tem uma base conceptual semelhante ao método progressivo, é um procedimento de alguma forma simétrico. O modelo é iniciado com todos os potenciais regressores e posteriormente, são retirados ao modelo, um a um os regressores cuja presença não contribua para explicar uma proporção significativa da variação total da variável dependente.

6.4.6.4 Método de Regressão Passo a Passo (ou *Stepwise*) [30]

Este método é um híbrido dos dois métodos anteriores. O primeiro passo inicia-se só com um regressor, tal como no método Progressivo, mas a significância da adição de uma nova variável independente ao modelo é testada como no método Regressivo. A vantagem deste método, é que permite a remoção de uma variável cuja importância no modelo se torna reduzida pela adição de novas variáveis (o método Passo a Passo é particularmente apropriado quando existem correlações significativas entre as variáveis independentes).

6.5 ABORDAGEM AOS DADOS

Inicialmente foram consideradas as 375 amostras e foi estudada a influência da discriminação da composição da amostra, na previsão do RON pelo modelo. Foram desenvolvidos três modelos com o objectivo de verificar se o aumento da discriminação da composição da amostra, usada para a elaboração dos modelos, provoca ou não um aumento da exactidão da previsão do RON por parte dos modelos.

No primeiro modelo a composição das amostras foi agrupada, de acordo com a sua família química, em 4 variáveis independentes: parafinas, olefinas, compostos nafténicos e compostos aromáticos. No segundo modelo a composição foi agrupada em 7 variáveis independentes: *i*-parafinas, *n*-

parafinas, *i*-olefinas, *n*-olefinas, *c*-olefinas, compostos nafténicos e compostos aromáticos. Por último no terceiro modelo a discriminação da composição da amostra foi máxima, considerando-se 47 variáveis independentes.

A fim de validar os modelos e verificar a homogeneidade dos dados, estes foram divididos em três grupos. No primeiro grupo foram considerados os dados referentes a 2003, no segundo os referentes a 2003-2004 e no terceiro os de 2003-2005. Com cada um destes grupos foi desenvolvido um modelo, considerando as 47 variáveis independentes, modelo esse posteriormente aplicado aos dados não incluídos na sua elaboração com o objectivo de validar cada um dos modelos.

Em todos os modelos as variáveis independentes foram caracterizadas e foi analisada a correlação que existia entre as mesmas. Foram retiradas as amostras cujos resultados eram atípicos e foi feito o diagnóstico aos pressupostos da regressão linear para cada um dos modelos estimados.

Depois de elaborado cada um dos modelos, estes foram usados para prever o RON e os resultados obtidos foram comparados com os obtidos pelo motor CFR.

7. RESULTADOS OBTIDOS COM A METODOLOGIA PROPOSTA

7.1 MODELO COM 4 VARIÁVEIS INDEPENDENTES

7.1.1 Análise das Variáveis

Os resultados cromatográficos obtidos foram agrupados, de acordo com a sua família química, em 4 variáveis: *Parafinas*, *Olefinas*, *Nafténicos* e *Aromáticos*. Considerando os resultados das 375 amostras disponíveis foi calculada a média e o desvio padrão para cada uma das variáveis, Tabela 8.

Tabela 8- Estatística descritiva da variável RON e das 4 variáveis independentes: *Parafinas*, *Olefinas*, *Nafténicos* e *Aromáticos*.

	Média	Desvio Padrão	N
RON	97,576	1,164	375
Olefinas	0,981	0,248	375
Parafinas	29,898	2,851	375
Aromáticos	67,269	3,073	375
Nafténicos	1,180	0,416	375

De seguida foi analisada a correlação existente entre a variável dependente *RON* e cada uma das variáveis independentes, assim como as correlações entre as variáveis independentes, Tabela 9, Figura 15 e Figura 16.

Tabela 9- Correlação entre as 4 variáveis independentes (*Parafinas*, *Olefinas*, *Nafténicos* e *Aromáticos*) e com a variável dependente *RON*.

	RON	Olefinas	Parafinas	Aromáticos	Nafténicos
RON	1,000	-0,244	-0,676	0,743	-0,601
Olefinas	-0,244	1,000	-0,069	-0,117	0,534
Parafinas	-0,676	-0,069	1,000	-0,969	0,307
Aromáticos	0,743	-0,117	-0,969	1,000	-0,497
Nafténicos	-0,601	0,534	0,307	-0,497	1,000

Verifica-se que a variável independente *Aromáticos* é a que mais se correlaciona com a variável dependente *RON* (0,743), e é a única que apresenta uma correlação positiva. A correlação positiva justifica-se pelo facto dos compostos aromáticos terem um RON muito elevado, sempre superior a

100 [13], assim quanto maior for a percentagem de compostos aromáticos presentes na amostra, mais elevado será o RON da amostra.

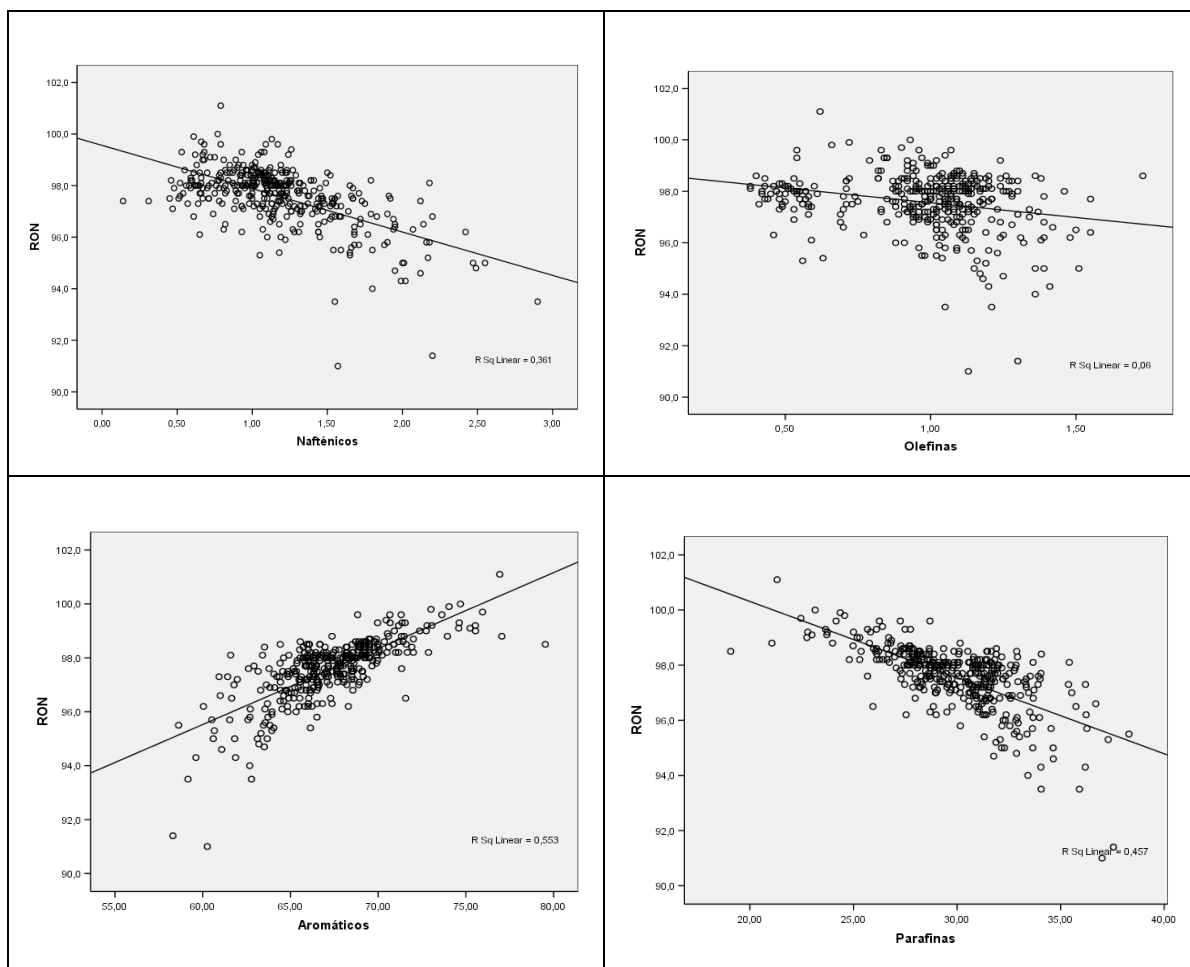


Figura 15- Representação gráfica da correlação entre cada uma das 4 variáveis independentes com a variável dependente RON.

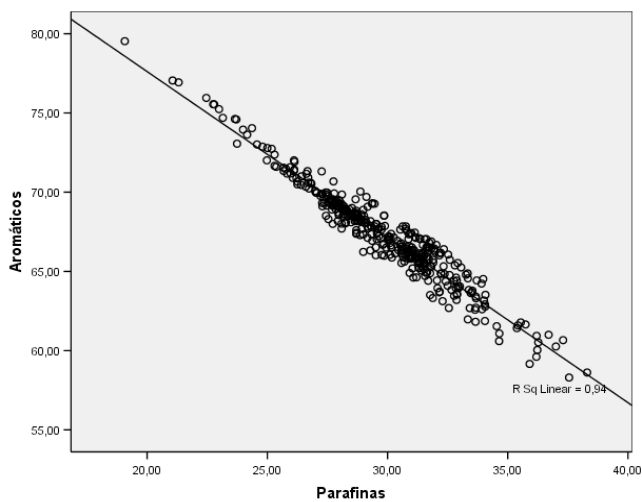


Figura 16- Representação gráfica da correlação entre as variáveis independentes *Parafinas* e *Aromáticos*.

A forte correlação que existe entre as duas variáveis deve-se a que mais de 60% da amostra é constituída por compostos aromáticos, logo é normal que esta influencie e muito o valor de RON da amostra.

As restantes variáveis independentes apresentam uma correlação mais fraca com a variável dependente RON, *Parafinas* (0,676), *Nafténicos* (0,601) e por fim a variável *Olefinas* (0,244), e em todos os casos a correlação é negativa. Verifica-se que quanto menor é a percentagem dessa variável independente na mistura menor é a correlação que esta tem com o RON, o que se percebe pelo facto da influência no valor final do RON na mistura ser menor. Por outro lado as correlações negativas são justificadas pelo facto do RON das parafinas, compostos nafténicos e olefinas variar com o tamanho da cadeira e com a complexidade das ramificações [13]. Assim um aumento da percentagem destas variáveis na mistura não corresponde a um aumento do valor de RON na mistura, depende sempre do composto cuja percentagem aumenta e do seu valor de RON.

Quanto às variáveis independentes, apresentam uma correlação fraca entre si à excepção das variáveis *Aromáticos* e *Parafinas*. A correlação apresentada por estas duas variáveis é negativa e muito forte (0,969), é superior à apresentada por qualquer das variáveis independentes relativamente ao RON. Este valor de correlação resulta simplesmente das amostras serem constituídas essencialmente por Parafinas e Aromáticos, a soma das duas é em média ~98% da composição das amostras. Assim quando a percentagem de uma delas aumenta na mistura a outra diminui e vice-versa, pelo que fatalmente estarão fortemente correlacionadas.

7.1.2 Estimação do modelo

Uma vez que é desconhecida qual ou quais as variáveis independentes a incluir no modelo de regressão, a selecção de regressores foi feita por três métodos diferentes: *Backward*, *Forward* e *Stepwise*. No entanto verifica-se que o ajuste do modelo obtido pelos métodos *Backward* e *Forward* não é melhor do que o desenvolvido pelo método *Stepwise*. Assim, e uma vez que este último é um híbrido dos outros dois, foi considerado apenas este método

de selecção de regressor para a estimação deste e dos próximos modelos de regressão linear.

O método de selecção de regressores *Stepwise*, quando aplicado às 375 amostras e considerando as quatro variáveis independentes, desenvolveu quatro modelos (M_1 , M_2 , M_3 e M_4), cujos coeficientes e regressores são apresentadas na Tabela 12.

No primeiro modelo (M_1) foi incluída apenas a variável independente *Aromáticos*, uma vez que é a que apresenta maior correlação em valor absoluto com a variável RON. A Tabela 10 mostra ainda que 55,3% da variação do RON no modelo M_1 é explicada pela variável *Aromáticos*. Este resultado é confirmado com a análise da ANOVA, Tabela 11, em que a soma total dos quadrados (507,019), na sua maior parte (280,224), o que corresponde a 55,3%, é explicada pela variável *Aromáticos*.

Tabela 10- Características dos modelos com 4 variáveis.

Modelo	r	R ²	R _a ²
M ₁	0,743	0,553	0,551
M ₂	0,790	0,624	0,622
M ₃	0,793	0,628	0,625
M ₄	0,792	0,628	0,626

Tabela 11- Resultados da análise ANOVA aos modelos com 4 variáveis.

Modelo	Soma dos Quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	p-value
M ₁	Regressão	1	280,224	460,871	0,000
	Residual	373	0,608		
	Total	374			
M ₂	Regressão	2	158,110	308,267	0,000
	Residual	372	0,513		
	Total	374			
M ₃	Regressão	3	106,146	208,824	0,000
	Residual	371	0,508		
	Total	374			
M ₄	Regressão	2	159,194	313,947	0,000
	Residual	372	0,507		
	Total	374			

Uma vez que antes da entrada da variável *Aromáticos* no modelo (M_1) o R^2 era nulo, a variação do coeficiente de determinação com a entrada da variável *Aromáticos* é igual ao valor desse coeficiente ($R^2=0,553$).

Do mesmo modo, a alteração do valor do teste F usado na ANOVA, Tabela 11, aquando da introdução da variável *Aromáticos* é igual ao valor do teste F (460,871).

A variável seguinte a entrar no modelo (M_2) é a variável independente que apresentar a maior correlação com a variável dependente RON depois de ajustados os efeitos da variável *Aromáticos* sobre a variável RON [29]. A adição da nova variável conduz a um maior acréscimo da estatística F -parcial da ANOVA da regressão de RON nessa variável, depois de considerado o efeito da variável *Aromáticos* [29]. Assim ao contrário do previsto na Tabela 9, a próxima variável independente a entrar no modelo é a variável *Nafténicos*, e só no terceiro modelo (M_3) é incluída a variável *Parafinas*.

De acordo com o teste F , Tabela 11, verifica-se que os quatro modelos ajustados em termos globais são significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Assim rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

A Tabela 12 apresenta os coeficientes de cada modelo, o seu desvio padrão, os testes t , e as tolerâncias de cada variável, para os quatro modelos desenvolvidos pelo método *Stepwise*.

O teste t permite testar a hipótese nula de inexistência de uma relação linear entre a variável dependente RON com cada uma das variáveis independentes, bem como com as constantes consideradas nos quatro modelos [31]. Em todos os modelos, à excepção do M_3 , os testes t têm associado um nível de significância inferior a 0,05, o que nos permite concluir que todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes de cada variável independente são diferentes de zero [31]. No caso do modelo M_3 , verifica-se que a variável *Aromáticos* não tem capacidade explicativa significativa do RON e por isso essa variável foi eliminada quando da elaboração do modelo seguinte, M_4 .

Tabela 12- Coeficientes dos modelos com 4 variáveis.

Modelo	Coeficientes		t	p-value	Estatística de Colinearidade		
	B	Desvio Padrão			Tolerância	VIF	
M ₁	(Constante)	78,627	0,884	88,985	0,000		
	Aromáticos	0,282	0,013	21,468	0,000	1,000	1,000
M ₂	(Constante)	83,534	1,001	83,466	0,000		
	Aromáticos	0,224	0,014	16,113	0,000	0,753	1,329
	Nafténicos	-0,859	0,103	-8,378	0,000	0,753	1,329
M ₃	(Constante)	102,733	9,245	11,112	0,000		
	Aromáticos	0,030	0,094	0,314	0,753	0,016	61,410
	Nafténicos	-1,166	0,179	-6,515	0,000	0,245	4,086
	Parafinas	-0,193	0,092	-2,089	0,037	0,020	51,038
M ₄	(Constante)	105,638	0,388	272,137	0,000		
	Nafténicos	-1,214	0,093	-13,064	0,000	0,906	1,104
	Parafinas	-0,222	0,014	-16,337	0,000	0,906	1,104

A variável independente *Olefinas* apresenta uma correlação muito fraca com a variável RON, não tem uma capacidade explicativa significativa e por isso não foi incluída em nenhum dos modelos. O método de selecção de regressores *Stepwise* criou estes quatro modelos, até conseguir que as variáveis não introduzidas no modelo não tenham capacidade explicativa da variável dependente e todas as que estão no modelo o tenham [29]. O modelo final estimado corresponde ao modelo M₄ e é dado por:

$$RON = 105,638 - 1,214 * Nafténicos - 0,222 * Parafinas$$

Depois de estimado o modelo de regressão, é necessário analisar as hipóteses que estão na sua base.

7.1.3 Validação dos pressupostos do modelo estimado

7.1.3.1 Homocedasticidade

Quando não se verifica a hipótese de homocedasticidade, os parâmetros do modelo estimados embora sejam centrados, não são, eficientes [31]. Para avaliar a homocedasticidade pode-se observar as relações que existem, por um lado entre os resíduos normalizados e os valores estimados de Y, e por outro lado entre os resíduos normalizados e os valores estimados de Y também normalizados [31], Figura 17.

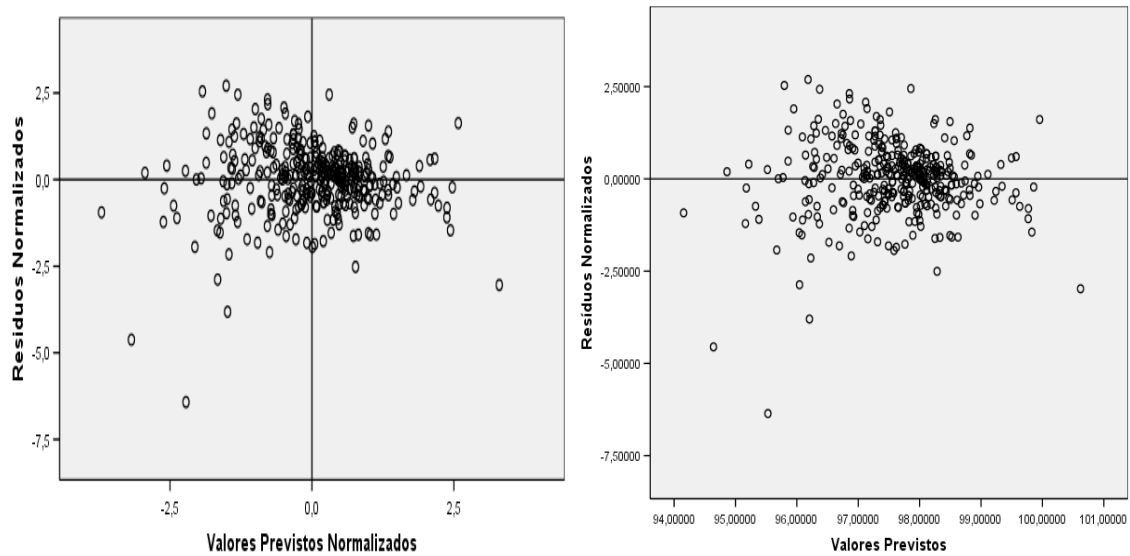


Figura 17- Avaliação da Homocedasticidade para o modelo M_4 , com 4 variáveis.

Da análise dos dois gráficos é visível que os resíduos têm uma amplitude aproximadamente constante em torno do eixo horizontal zero, ou seja, não existe uma tendência crescente ou decrescente, logo não se rejeita a hipótese da homocedasticidade.

7.1.3.2 Covariância Nula

A covariância é analisada pelo teste de Durbin-Watson em que para o modelo M_4 $d=1,102$, logo existe auto-correlação entre os resíduos.

7.1.3.3 Normalidade dos Resíduos

A Normalidade é testada usando o teste de Kolmogorov-Smirnov, Tabela 13. Os desvios à normalidade podem também ser observados graficamente, Figura 18.

Tabela 13- Teste à Normalidade dos resíduos, modelo M_4 , com 4 variáveis.

	Kolmogorov-Smirnov		
	Estadística	Graus Liberdade	Significância
Resíduos Padrão	0,069	375	0,000

De acordo com o teste Kolmogorov-Smirnov, verifica-se que os resíduos não seguem uma distribuição normal para um nível de significância de 0,05.

Graficamente também é visível a não normalidade, uma vez que nem todas as observações se dispõem ao longo das rectas oblíqua e horizontal.

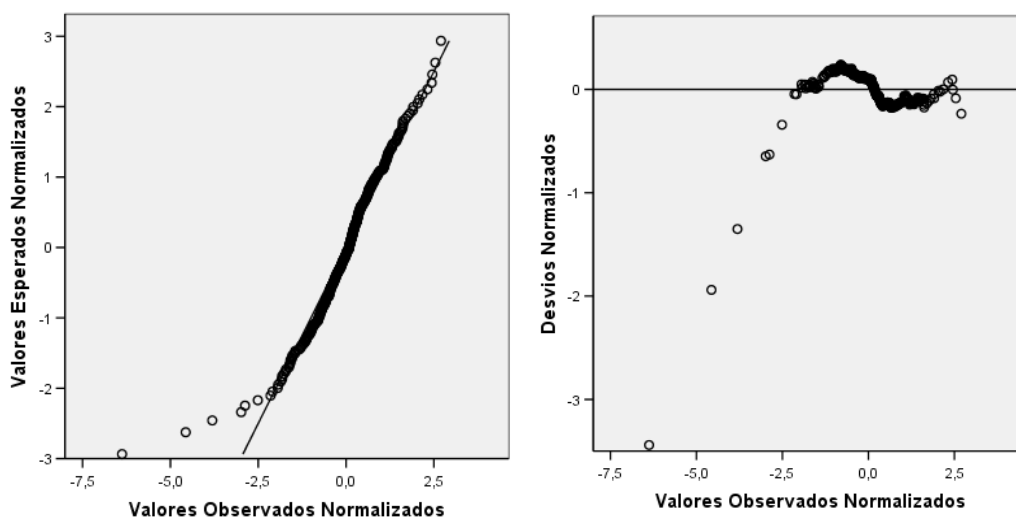


Figura 18- Análise gráfica da normalidade dos resíduos para o modelo M_4 , com 4 variáveis.

7.1.3.4 Multicolinearidade

O modelo de regressão linear múltipla pressupõe que não se verifique multicolinearidade, ou seja, que as variáveis explicativas sejam linearmente independentes [31]. A intensidade da multicolinearidade é analisada através da correlação entre as variáveis independentes e dos valores da Tolerância e do VIF. Relativamente à correlação entre as variáveis independentes já foi visto no ponto 7.1.1 que é muito forte entre as variáveis *Aromáticos* e *Parafinas* mas apenas pelo facto do somatório das duas representar em média ~98% da composição das amostras.

Os valores obtidos para a Tolerância e para o VIF, [Tabela 12](#), confirmam a multicolinearidade entre as variáveis independentes *Aromáticos* e *Parafinas*, no modelo M_3 . No entanto esta situação é ultrapassada uma vez que a variável *Aromáticos* é excluída no modelo M_4 .

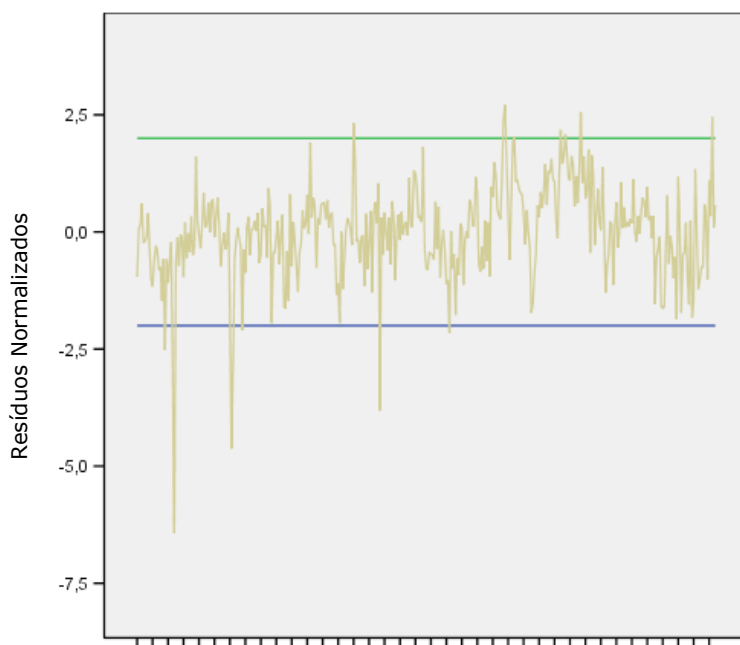
7.1.4 Análise dos Valores Atípicos

Da análise gráfica dos resíduos é possível identificar eventuais valores atípicos ou aberrantes. Estes valores são observações extremas, que apresentam

resíduos bastante superiores, em valor absoluto, aos resíduos das restantes observações, por exemplo um valor superior a dois desvios-padrão relativamente à média [29]. Assim é necessário proceder à análise dos resíduos para identificar e eliminar os valores atípicos, [Figura 19](#).

Neste caso são considerados atípicos todos os resíduos normalizados com valores superiores a 2 em valor absoluto. Verifica-se a existência de alguns valores atípicos. As amostras correspondentes a estes valores devem ser retiradas e o modelo deve ser estimado novamente sem a sua presença.

Estes valores atípicos podem ser consequência de um erro analítico, de um erro de transcrição de resultados, ou mesmo de um dia de produção anómalo que provocou alterações na composição da amostra desse dia. No entanto a esta distância não é possível rastrear qual o motivo destes valores atípicos.



[Figura 19](#)- Análise dos Resíduos, modelo M_4 , com 4 variáveis.

Verifica-se que este modelo M_4 não cumpre os requisitos necessários para um Modelo de Regressão Linear Multipla. Os resíduos são auto-correlacionados e não seguem uma distribuição normal, e confirma-se a presença de valores atípicos. Assim serão excluídos os valores atípicos, encontrados a partir da análise de resíduos, e desenvolvido um novo modelo considerando as mesmas 4 variáveis independentes (*Parafinas, Olefinas, Nafténicos e Aromáticos*)

7.2 MODELO COM 4 VARIÁVEIS INDEPENDENTES EXCLUÍDO OS VALORES ATÍPICOS

7.2.1 Análise das Variáveis

Retirando as amostras que apresentavam valores atípicos, o número de amostras passa a ser de 359 e cada variável apresenta uma nova média e desvio padrão, Tabela 14.

Tabela 14- Estatística descritiva do RON e das 4 variáveis independentes: *Parafinas, Olefinas, Nafténicos e Aromáticos*, sem as amostras atípicas.

	Média	Desvio Padrão	N
RON	97,630	1,0203	359
Nafténicos	1,1675	0,41138	359
Aromáticos	67,3736	2,92615	359
Parafinas	29,8039	2,72542	359
Olefinas	0,9757	0,24814	359

As correlações entre a variável dependente *RON* e cada uma das variáveis independentes, assim como as correlações entre as variáveis independentes, são idênticas às obtidas antes da exclusão das amostras atípicas e apresentadas na Tabela 9.

7.2.2 Estimação do Modelo

O método de selecção de regressores *Stepwise*, quando aplicado às 359 amostras, considerando as 4 variáveis independentes, desenvolveu 5 modelos (M_1 , M_2 , M_3 , M_4 e M_5), os seus coeficientes e regressores estão apresentados na Tabela 15. A primeira variável independente a entrar no modelo M_1 foi novamente a *Aromáticos*, uma vez que continua a apresentar o maior coeficiente de correlação em valor absoluto [29]. A Tabela 16 mostra ainda que 62,1% da variação do RON no modelo M_1 é explicada pela variável *Aromáticos*.

Tal como no ponto 7.1, e pelos mesmos motivos, mesmo depois da exclusão das amostras atípicas, a segunda variável a entrar no modelo M_2 foi a variável *Nafténicos*, a terceira variável a ser incluída no modelo M_3 foi a variável *Parafinas* e no modelo M_4 foi novamente excluída a variável *Aromáticos*. No

entanto neste caso o método *Stepwise* desenvolveu mais um modelo M_5 no qual incluí a variável *Olefinas*, o que não aconteceu no ponto 7.1. Assim a variável *Olefinas*, depois da exclusão das amostras atípicas, passa a ter uma capacidade explicativa significativa e por isso é incluída no último modelo (M_5) [29]. Verifica-se ainda que os coeficientes estimados agora para os modelos de M_1 a M_4 são idênticos aos obtidos no ponto 7.1 antes da exclusão das amostras atípicas.

Tabela 15- Coeficientes dos modelos, com 4 variáveis, sem as amostras atípicas.

Modelo		Coeficientes		t	p-value	Estatística da Colinearidade	
		B	Desvio Padrão			Tolerance	VIF
M ₁	(Constante)	79,120	0,766	103,246	0,000	1,000	1,000
	Aromáticos	0,275	0,011	24,176	0,000		
M ₂	(Constante)	83,651	0,829	100,870	0,000	0,769	1,300
	Aromáticos	0,221	0,012	19,147	0,000		
	Nafténicos	-0,792	0,082	-9,643	0,000		
M ₃	(Constante)	98,283	7,412	13,260	0,000	0,018	55,887
	Aromáticos	0,073	0,075	0,969	0,333		
	Nafténicos	-1,028	0,144	-7,136	0,000		
	Parafinas	-0,147	0,074	-1,986	0,048		
M ₄	(Constante)	105,458	0,324	325,203	0,000	0,923	1,084
	Nafténicos	-1,147	0,075	-15,355	0,000		
	Parafinas	-0,218	0,011	-19,305	0,000		
M ₅	(Constante)	105,932	0,377	280,812	0,000	0,593	1,685
	Nafténicos	-1,014	0,093	-10,956	0,000		
	Parafinas	-0,227	0,012	-19,149	0,000		
	Olefinas	-0,358	0,148	-2,411	0,016		

Tabela 16- Características dos modelos com 4 variáveis, sem as amostras atípicas.

Modelo	r	R ²	R _a ²
M ₁	0,788	0,621	0,620
M ₂	0,836	0,699	0,698
M ₃	0,838	0,703	0,700
M ₄	0,838	0,702	0,700
M ₅	0,841	0,707	0,704

Verifica-se que o melhor ajuste é o do modelo M_5 com $r=0,841$; $R^2=0,707$ e $R_a^2=0,704$; a exclusão das amostras atípicas foi uma boa decisão uma vez que aumentou o coeficiente de determinação dos modelos. O R_a^2 passou de 0,626 para 0,704, Tabela 16.

De acordo com o teste F , Tabela 17, todos os cinco modelos ajustados são globalmente significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) em cada modelo, e verifica-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

Tabela 17- Resultados da análise ANOVA aos modelos com 4 variáveis, sem amostras atípicas.

Modelo		Soma dos quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	<i>p-value</i>
M_1	Regressão	231,360	1	231,360	584,493	0,000
	Residual	141,311	357	0,396		
	Total	372,671	358			
M_2	Regressão	260,626	2	130,313	414,042	0,000
	Residual	112,045	356	0,315		
	Total	372,671	358			
M_3	Regressão	261,858	3	87,286	279,627	0,000
	Residual	110,814	355	0,312		
	Total	372,671	358			
M_4	Regressão	261,564	2	130,782	419,043	0,000
	Residual	111,107	356	0,312		
	Total	372,671	358			
M_5	Regressão	263,355	3	87,785	285,077	0,000
	Residual	109,316	355	0,308		
	Total	372,671	358			

A Tabela 15 apresenta os coeficientes dos vários modelos desenvolvidos e o teste t aos coeficientes dos regressores e tal como antes da exclusão das amostras atípicas, verifica-se no modelo M_3 que a variável *Aromáticos* deixa de ser significativa e é excluída logo no modelo seguinte. O modelo que apresenta o melhor ajuste é modelo M_5 :

$$RON = 105,932 - 1,014 * Nafténicos - 0,227 * Parafinas - 0,358 * Olefinas$$

7.2.3 Validação dos pressupostos do modelo estimado

7.2.3.1 Homocedasticidade

Tal como no modelo anterior, os resíduos apresentam uma amplitude aproximadamente constante em torno do eixo horizontal zero, e não se verificam tendências crescentes ou decrescentes. Assim não se rejeita a hipótese da homocedasticidade.

7.2.3.2 Covariância Nula

Para o modelo M_5 , o teste de Durbin-Watson tem um valor $d=1,192$, de acordo com os critérios deste teste não se pode dizer que não existe auto-correlação entre os resíduos. No entanto de seguida é possível verificar que apesar da auto-correlação os resíduos seguem uma distribuição normal.

7.2.3.3 Normalidade dos Resíduos

Depois de retiradas as amostras atípicas e de acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, para um nível de significância de 0,05, Tabela 18. Graficamente verifica-se que os valores esperados e observados normalizados se dispõem ao longo da recta oblíqua; os desvio normalizados também se dispõem ao longo da recta horizontal mas com uma grande amplitude e revelando uma certa tendência oscilatória, Figura 20.

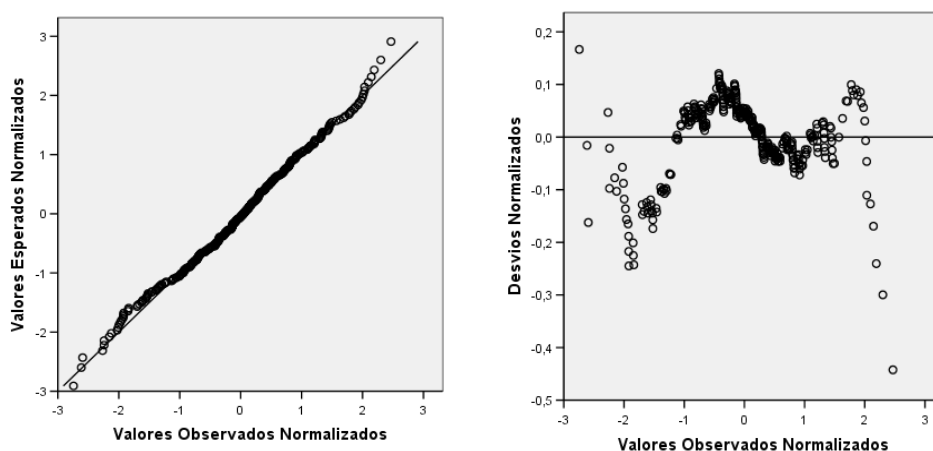


Figura 20- Análise gráfica da normalidade dos resíduos para o modelo M_5 com 4 variáveis, sem amostras atípicas.

Tabela 18- Teste à normalidade dos resíduos, modelo M₅ com 4 variáveis, sem amostras atípicas.

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,038	359	0,200

7.2.3.4 Multicolinearidade

Neste modelo M₃, tal como no anterior modelo M₃ de 7.1, e pelos mesmos motivos verifica-se a existência de multicolinearidade entre as variáveis *Aromáticos* e *Parafinas*, Tabela 15. Esta multicolinearidade é confirmada pelo baixo valor da Tolerância e elevado valor do VIF apresentados no modelo M₃, Tabela 15.

Com a exclusão dos valores atípicos verifica-se que o coeficiente de correlação do modelo M₅ é melhor, a homocedasticidade mantém-se e os resíduos passam a ter uma distribuição normal, de acordo com o teste de Kolmogorov-Smirnov. Continua a existir multicolinearidade entre as variáveis *Aromáticos* e *Parafinas*, no entanto este facto não afecta o modelo seleccionado M₅ uma vez que a variável *Aromáticos* não foi incluída neste modelo.

7.2.4 Previsão do RON usando o modelo

Considerando o modelo desenvolvido é possível prever o RON para as amostras usadas na sua elaboração, a partir da sua composição obtida por cromatografia.

Os valores obtidos a partir do modelo matemático para o RON podem ser comparados com os valores obtidos para o RON pelo motor CFR, Figura 21. Idealmente a relação entre as duas variáveis deve ser uma recta com declive um e ordenada na origem igual a zero, no entanto a relação obtida foi um pouco diferente:

$$\text{RON}_{\text{modelo}} = 28,66 + 0,71 * \text{RON}_{\text{CFR}} \text{ com } R^2 = 0,71.$$

Outra forma de comparar os dois métodos de determinação do RON, e avaliar a exactidão dos resultados obtidos pelo modelo, face aos obtidos pelo CFR, é o

teste *t* de Student das médias [32]. Neste caso é necessário calcular as médias dos valores obtidos por cada um dos métodos, \bar{x}_1 e \bar{x}_2 , e os respectivos desvios padrão s_1 e s_2 , Tabela 19.

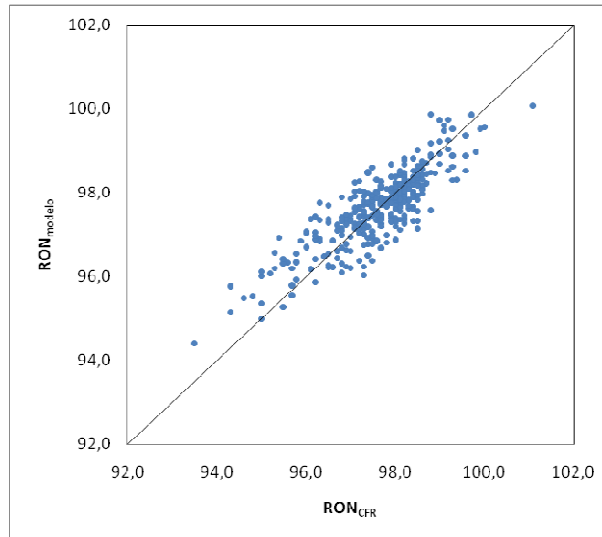


Figura 21- Representação gráfica do valor do RON_{CFR} e do RON_{modelo}, com 4 variáveis, sem as amostras atípicas.

Tabela 19- Resultados da comparação do RON_{CFR} e do RON_{modelo}, com 4 variáveis, sem amostras atípicas.

	RON _{CFR}	RON _{modelo}
Média	97,63	97,63
Desvio Padrão	1,02	0,86
Variância	1,04	0,73
n	359	359
<hr/>		
F _{calc}		1,42
F _{tab}		0,83
F _{calc} > F _{tab}		
<hr/>		
t _{calc}		0,05
t _{tab}		1,96
t _{calc} < t _{tab}		

Considera-se como hipótese nula que os dois métodos originam os mesmos resultados, $H_0: \mu_1 = \mu_2$, e é necessário avaliar se $(\bar{x}_1 - \bar{x}_2)$ é significativamente diferente de zero. Para decidir se a diferença entre as duas médias é significativa, e testar a hipótese nula, surge o teste *t de Student*, eq. 24 [32].

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (\text{eq.24})$$

em que:

$$s = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} \quad (\text{eq.25})$$

O valor de *t* calculado (t_{calc}) é comparado com o valor de *t* tabelado (t_{tab}) para $(n_1 + n_2 - 1)$ graus de liberdade, Tabela 19. Como $t_{\text{calc}} < t_{\text{tab}}$, para um nível de significância de 0,05, não se rejeita a hipótese nula, considera-se que as médias dos resultados obtidos pelos dois métodos não são significativamente diferentes.

A fim de verificar se as variâncias obtidas pelos dois métodos são significativamente diferentes ou não, aplica-se o teste *F*, em que:

$$F = \frac{s_1^2}{s_2^2} \quad F \geq 1 \quad (\text{eq. 26})$$

Considera-se como hipótese nula que os dois métodos têm variâncias semelhantes, $H_0: \sigma_1 = \sigma_2$, e compara-se o valor de *F calculado* (F_{calc}) com o valor de *F tabelado* (F_{tab}) com $(n_1 - 1)$ e $(n_2 - 1)$ graus de liberdade, Tabela 19.

Como neste caso $F_{\text{calc}} > F_{\text{tab}}$, para um nível de significância de 0,05, rejeita-se a hipótese nula, as variâncias obtidas pelos dois métodos não são semelhantes; sendo que os resultados obtidos a partir do motor CFR apresentam uma variância superior aos obtidos pelo modelo.

É também importante avaliar as diferenças obtidas entre os resultados provenientes pelos dois métodos, Figura 22. Verifica-se que estas diferenças não apresentam uma tendência crescente ou decrescente no entanto em 22% dos casos o RON obtido pelo modelo difere do RON obtido pelo motor CFR mais do que $|0,7|$ (reprodutibilidade do método de referência ASTM D 2699).

Embora estes dados não sirvam para comparar os dois métodos são fundamentais para o uso do modelo como uma alternativa ao motor CFR, uma vez que este é o desvio máximo permitido pelo método de referência.

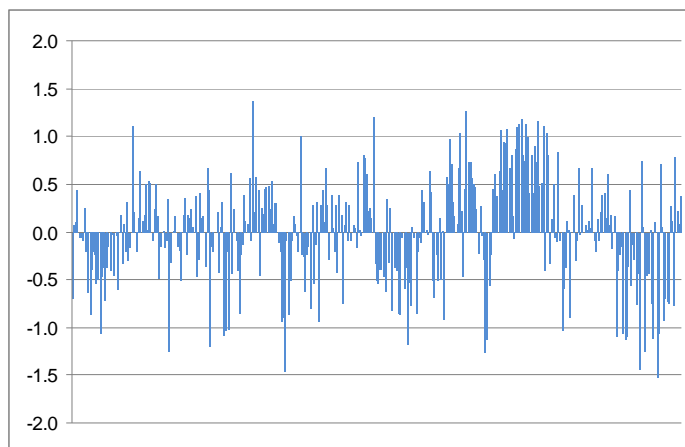


Figura 22- Diferença entre os resultados obtidos para o RON pelo modelo, com 4 variáveis sem as amostras atípicas, e pelo motor CFR.

7.3 MODELO COM 7 VARIÁVEIS INDEPENDENTES

7.3.1 Análise das Variáveis

Neste novo modelo foram consideradas 7 variáveis: *i-Parafinas*, *n-Parafinas*, *i-Olefinas*, *n-Olefinas*, *c-Olefinas*, *Nafténicos* e *Aromáticos*. Ao conjunto total das 375 amostras foram retirados as amostras atípicas, após análise de resíduos, e foram consideradas 355 amostras para desenvolver o modelo, Tabela 20.

Tabela 20- Estatística descritiva do RON e das 7 variáveis independentes: *i-Parafinas*, *n-Parafinas*, *i-Olefinas*, *n-Olefinas*, *c-Olefinas*, *Nafténicos* e *Aromáticos*.

	Média	Desvio Padrão	N
RON	97,653	1,0032	355
<i>i-Parafinas</i>	21,7637	2,11107	355
<i>n-Parafinas</i>	8,0492	0,86507	355
<i>i-Olefinas</i>	0,6799	0,17177	355
<i>n-Olefinas</i>	0,2499	0,08361	355
<i>c-Olefinas</i>	0,0457	0,03269	355
Nafténicos	1,1643	0,41075	355
Aromáticos	67,3710	2,94347	355

Da análise da correlação entre a variável dependente *RON* e cada uma das variáveis independentes verifica-se que a variável independente *Aromáticos* continua a ser a que apresenta uma correlação mais forte (0,786) com o *RON*, Tabela 21,

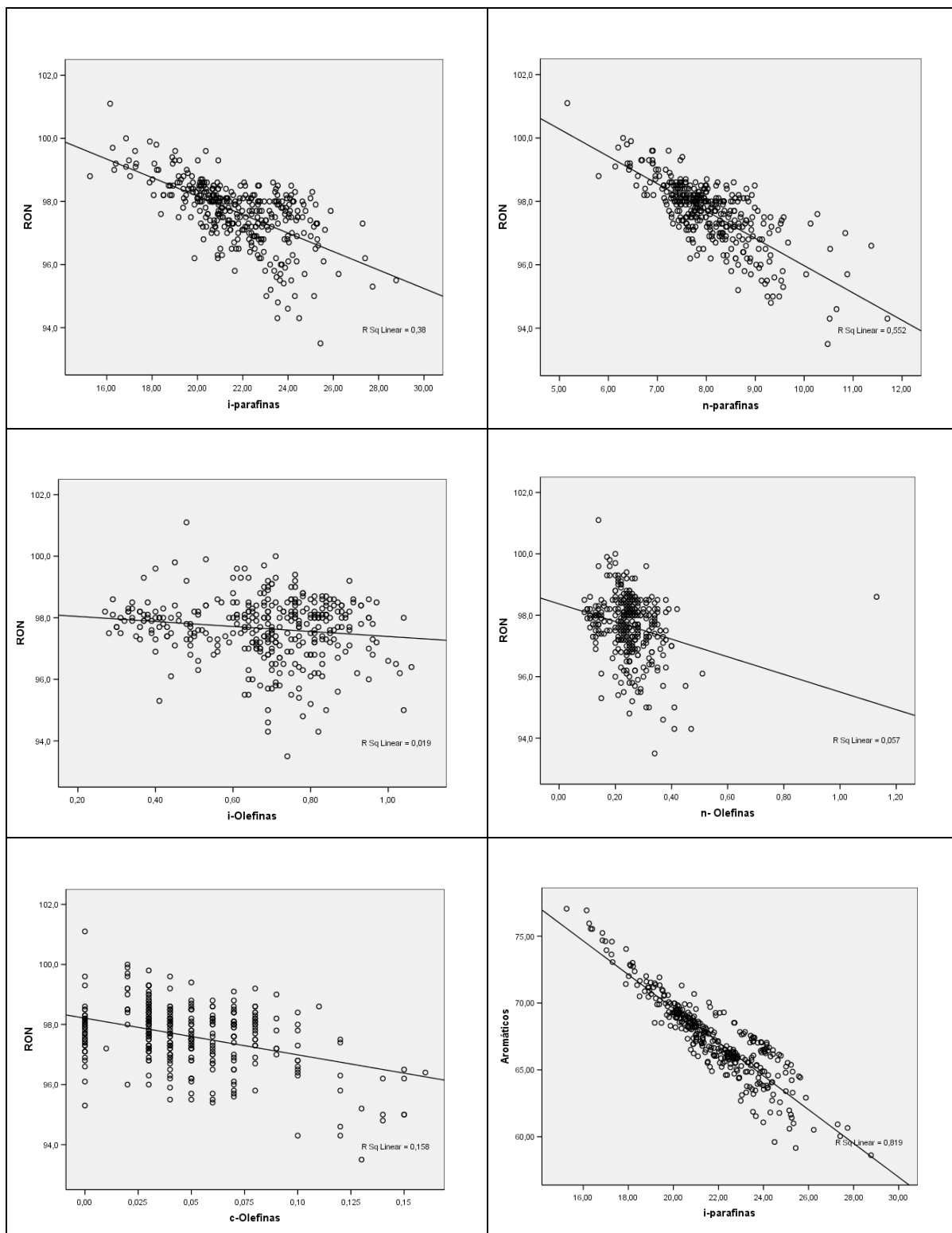


Figura 20- Representação gráfica da correlação entre as variáveis independentes e a variável dependente RON, e das variáveis independentes Aromáticos e *i*-Parafinas.

As correlações das restantes variáveis independentes com o RON não diferem das já apresentadas no ponto 7.1.1. Quanto às variáveis independentes correlacionam-se pouco entre si à excepção da variável *Aromáticos* com as variáveis *i*-Parafinas e *n*-Parafinas. A correlação apresentada por estas duas variáveis relativamente ao RON é muito forte, tal como já explicado em pontos anteriores, Tabela 21, Figura 20.

	RON	<i>i</i> -parafinas	<i>n</i> -parafinas	<i>i</i> -olefinas	<i>n</i> -olefinas	<i>c</i> -olefinas	nafténicos	aromáticos
RON	1,000	-0,617	-0,743	-0,139	-0,239	-0,398	-0,616	0,786
<i>i</i> -parafinas	-0,617	1,000	0,634	-0,349	0,018	-0,175	0,157	-0,905
<i>n</i> -parafinas	-0,743	0,634	1,000	0,165	0,434	0,290	0,502	-0,858
<i>i</i> -olefinas	-0,139	-0,349	0,165	1,000	0,451	0,673	0,388	0,044
<i>n</i> -olefinas	-0,239	0,018	0,434	0,451	1,000	0,454	0,465	-0,279
<i>c</i> -olefinas	-0,398	-0,175	0,290	0,673	0,454	1,000	0,738	-0,160
nafténicos	-0,616	0,157	0,502	0,388	0,465	0,738	1,000	-0,478
aromáticos	0,786	-0,905	-0,858	0,044	-0,279	-0,160	-0,478	1,000

Tabela 21- Correlação entre as 7 variáveis independentes (*i*-Parafinas, *n*-Parafinas, *i*-Olefinas, *n*-Olefinas, *c*-Olefinas, Nafténicos e Aromáticos) e com a variável dependente RON.

7.3.2 Estimação do Modelo

O método de selecção de regressores *Stepwise*, quando aplicado às 7 variáveis independentes desenvolveu cinco modelos (M_1 , M_2 , M_3 , M_4 , M_5), cujos coeficientes e regressores estão apresentados na Tabela 24.

A primeira variável independente a entrar no modelo (M_1) foi novamente a *Aromáticos*. Esta variável continua a ser aquela que apresenta maior coeficiente de correlação em valor absoluto com a variável RON. A Tabela 22 mostra ainda que 61,8% da variação do RON no modelo M_1 é explicada pela variável *Aromáticos*.

A variável adicionada ao modelo (M_2) é a *c*-olefinas, ao modelo (M_3) é a *n*-olefinas, ao modelo (M_4) os nafténicos e por fim as *n*-parafinas ao último modelo (M_5). A variável *i*-parafinas não é adicionada a nenhum dos modelos,

sendo por isso implícito que não apresenta uma capacidade explicativa significativa das variações da variável dependente RON. A ordem de entrada das variáveis no modelo é bastante diferente da prevista quando analisadas inicialmente as correlações entre a variável RON e cada uma das variáveis independentes, Tabela 21. No entanto como é sabido o método *Stepwise* recalcula o efeito de cada uma das restantes variáveis independentes na variável dependente RON após a adição de cada variável independente ao modelo [29]. E só termina a elaboração do modelo quando todas as variáveis não introduzidas no modelo não tiverem capacidade explicativa da variável RON e todas as que estejam no modelo tenham [29].

Tabela 22- Características dos modelos com 7 variáveis.

Modelo	r	R ²	R _a ²
M ₁	0,786	0,618	0,617
M ₂	0,833	0,694	0,692
M ₃	0,840	0,706	0,704
M ₄	0,848	0,718	0,715
M ₅	0,853	0,728	0,724

Tabela 23- Resultados da análise ANOVA aos modelos com 7 variáveis.

Modelo	Soma dos quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	p-value
M ₁ Regressão	220,158	1	220,158	570,991	0,000
Residual	136,107	353	0,386		
Total	356,265	354			
M ₂ Regressão	247,255	2	123,627	399,199	0,000
Residual	109,010	352	0,310		
Total	356,265	354			
M ₃ Regressão	251,549	3	83,850	281,059	0,000
Residual	104,716	351	0,298		
Total	356,265	354			
M ₄ Regressão	255,924	4	63,981	223,173	0,000
Residual	100,341	350	0,287		
Total	356,265	354			
M ₅ Regressão	259,463	5	51,893	187,088	0,000
Residual	96,802	349	0,277		
Total	356,265	354			

De acordo com o teste *F*, Tabela 23, verifica-se que todos os modelos ajustados em termos globais são significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Assim rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

De acordo com o teste *t*, Tabela 24, todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes das variáveis independente são diferentes de zero, em todos os modelos [31].

Tabela 24- Coeficientes dos modelos com 7 variáveis.

Modelo	Coeficientes		t	p-value	Estatísticas da Colinearidade	
	B	Desvio Padrão			Tolerância	VIF
M ₁ (Constante)	79,603	0,756	105,281	0,000		
Aromáticos	0,268	0,011	23,895	0,000	1,000	1,000
M ₂ (Constante)	81,018	0,694	116,688	0,000		
Aromáticos	0,253	0,010	24,828	0,000	0,975	1,026
c-Olefinas	-8,573	0,917	-9,354	0,000	0,975	1,026
M ₃ (Constante)	80,094	0,724	110,679	0,000		
Aromáticos	0,262	0,010	25,478	0,000	0,921	1,086
c-Olefinas	-10,208	0,997	-10,234	0,000	0,793	1,262
n- Olefinas	1,521	0,401	3,794	0,000	0,750	1,334
M ₄ (Constante)	82,029	0,865	94,813	0,000		
Aromáticos	0,238	0,012	20,249	0,000	0,676	1,480
c-Olefinas	-6,304	1,398	-4,508	0,000	0,388	2,580
n- Olefinas	1,689	0,395	4,272	0,000	0,741	1,350
Nafténicos	-0,479	0,123	-3,906	0,000	0,319	3,134
M ₅ (Constante)	87,959	1,866	47,149	0,000		
Aromáticos	0,178	0,020	8,743	0,000	0,218	4,596
c-Olefinas	-5,470	1,395	-3,921	0,000	0,377	2,654
n- Olefinas	2,147	0,409	5,243	0,000	0,668	1,496
Nafténicos	-0,514	0,121	-4,243	0,000	0,317	3,154
n-parafinas	-0,249	0,070	-3,572	0,000	0,216	4,636

O melhor ajuste é o apresentado pelo modelo M₅ com $r=0,853$; $R^2=0,728$ e $R_a^2=0,724$, Tabela 22. O uso de 7 variáveis em vez de 4 aumentou o R_a^2 do modelo, no entanto verifica-se que o melhor modelo obtido não usa as 7 variáveis, modelo M₅:

$$RON = 87,959 + 0,178 * \text{Aromáticos} - 5,470 * \text{c-Olefinas} + 2,147 * \text{n-Olefinas} - 0,514 * \text{Nafténicos} - 0,249 * \text{n-Parafinas}$$

Depois de estimado o modelo de regressão é necessário analisar as hipóteses que estão na sua base.

7.3.3 Validação dos pressupostos do modelo estimado

7.3.3.1 Homocedasticidade

Também neste modelo os resíduos apresentam uma amplitude aproximadamente constante em volta do eixo horizontal zero, e não se verificam tendências crescentes ou decrescentes, Figura 21, pelo que não se rejeita a hipótese da homocedasticidade.

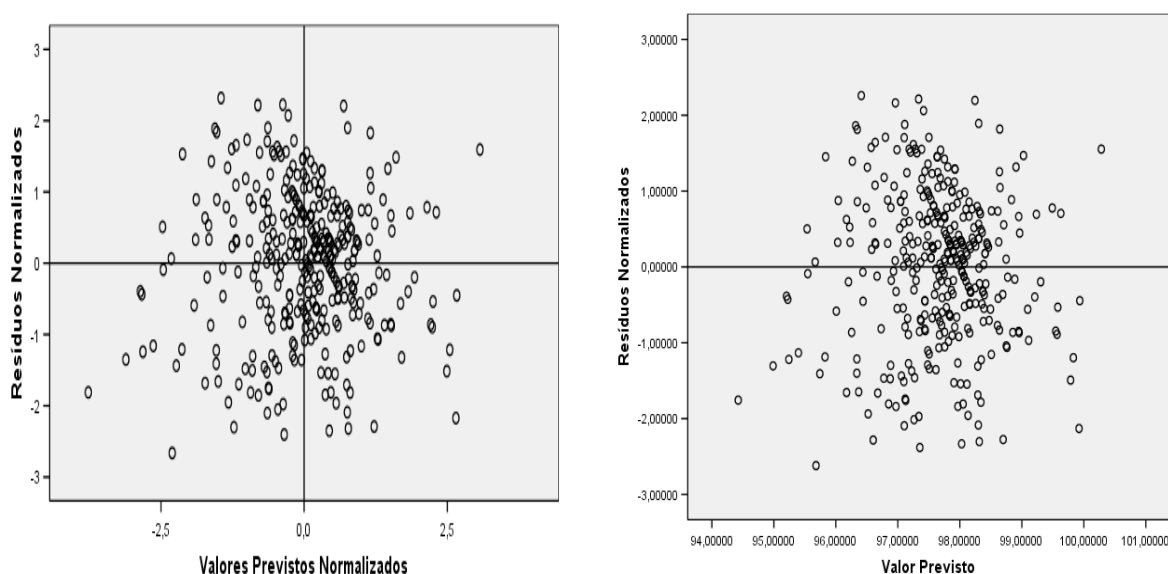


Figura 21- Avaliação da Homocedasticidade para o modelo M₅ com 7 variáveis.

7.3.3.2 Covariância Nula

Para o modelo M₅, o teste de Durbin-Watson tem um valor de $d=1,360$, o que de acordo com os critérios deste teste não se pode dizer que não existe autocorrelação entre os resíduos.

7.3.3.3 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, para um nível de significância de 0,05, Tabela 25.

Graficamente os valores esperados normalizados e os valores observados normalizados dispõem-se ao longo da recta oblíqua, Figura 22.

Os desvios normalizados também se dispõem ao longo da recta horizontal, sendo que apresentam uma menor amplitude que no modelo anterior e revelam uma tendência oscilatória menor, Figura 22.

Tabela 25- Teste à Normalidade dos resíduos, para o modelo M₅ com 7 variáveis.

	Kolmogorov-Smirnov		
	Estadística	Graus Liberdade	Significância
Resíduos Padrão	0,051	355	0,200

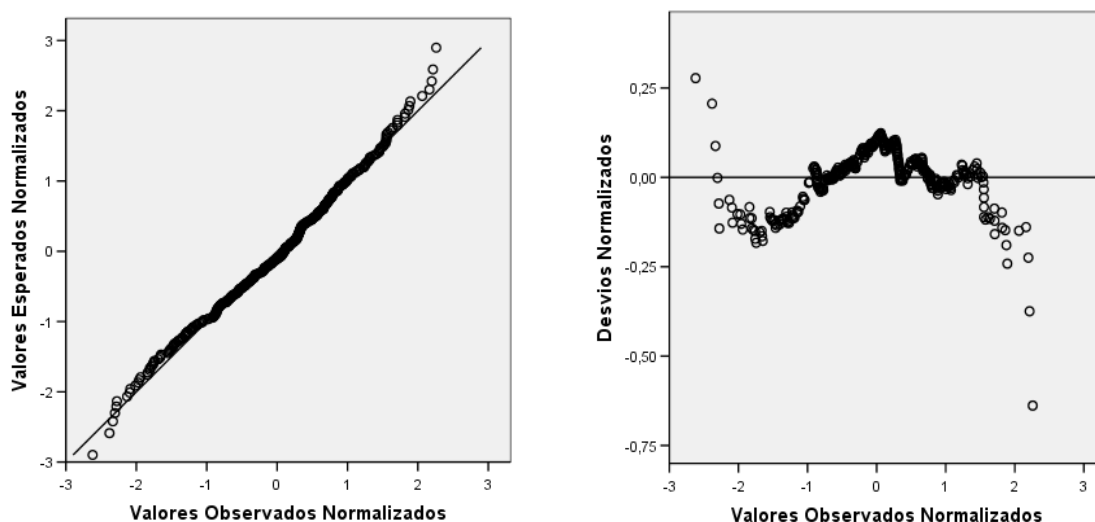


Figura 22- Análise gráfica da normalidade dos resíduos, para o modelo M₅, com 7 variáveis.

7.3.3.4 Multicolinearidade

Apesar de se verificar uma correlação forte entre as variáveis *Aromáticos* e *i-Parafinas* e *Aromáticos* e *n-Parafinas*, Tabela 21, esta não se confirma quando da elaboração dos modelos. A variável *i-Parafinas* não é incluída em nenhum dos modelos desenvolvidos logo o problema da multicolinearidade não se põe.

A variável *n-Parafinas* é introduzida apenas no modelo M₅ e não é verificada multicolinearidade uma vez que os valores da Tolerância são todos superiores a 0,1 e os valores de VIF são inferiores a 10, Tabela 24.

7.3.4 Previsão do RON usando o modelo

Considerando os coeficientes obtidos no modelo M₅, foi previsto o valor do RON para as amostras usadas na sua elaboração. Estes valores foram comparados com os valores do RON obtidos pelo motor CFR, Figura 23. Verifica-se que a recta obtida se aproxima pouco mais da relação ideal (declive=1 e ordenada na origem=0) e o coeficiente de determinação melhora:

$$\text{RON}_{\text{modelo}} = 26,55 + 0,73 * \text{RON}_{\text{CFR}} \text{ com } R^2 = 0,73.$$

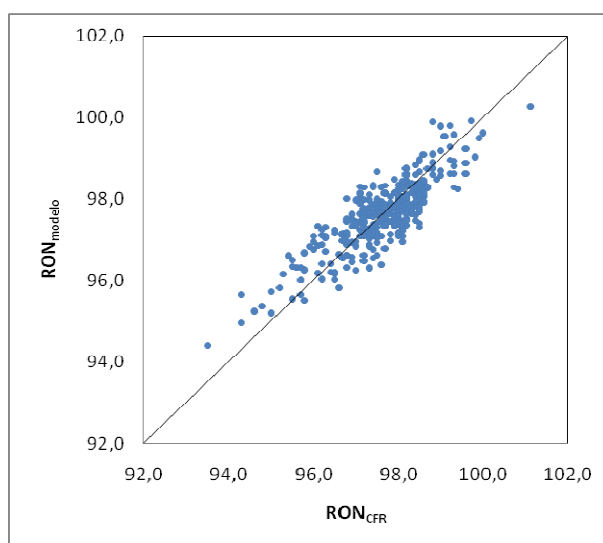


Figura 23- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 7 variáveis.

Na comparação das médias obtidas pelos dois métodos, de acordo com o teste *t* não se rejeita a hipótese nula, para um nível de significância de 0,05, uma vez que $t_{\text{calc}} < t_{\text{tab}}$, Tabela 26. Assim considera-se que as médias dos resultados obtidos pelos dois métodos não são significativamente diferentes.

Na comparação das variâncias, como $F_{\text{calc}} > F_{\text{tab}}$, para um nível de significância de 0,05, rejeita-se a hipótese nula, as variâncias obtidas pelos dois métodos

não são semelhantes. No entanto os resultados obtidos a partir do motor CFR apresentam uma variância superior aos obtidos pelo modelo, Tabela 26. Com este novo modelo as diferenças entre os resultados obtidos pelos dois métodos diminuem, incluindo a percentagem de diferenças superiores a $|0,7|$ que diminui para 20%, Figura 24.

Tabela 26- Resultados da comparação do RON_{CFR} e RON_{modelo} com 7 variáveis.

	RON_{CFR}	RON_{modelo}
Média	97,7	97,6
Desvio Padrão	1,00	0,86
Variância	1,01	0,73
n	355	355
F_{calc}		1,37
F_{tab}		0,89
$F_{calc} > F_{tab}$		
t_{calc}		0,25
t_{tab}		1,96
$t_{calc} < t_{tab}$		

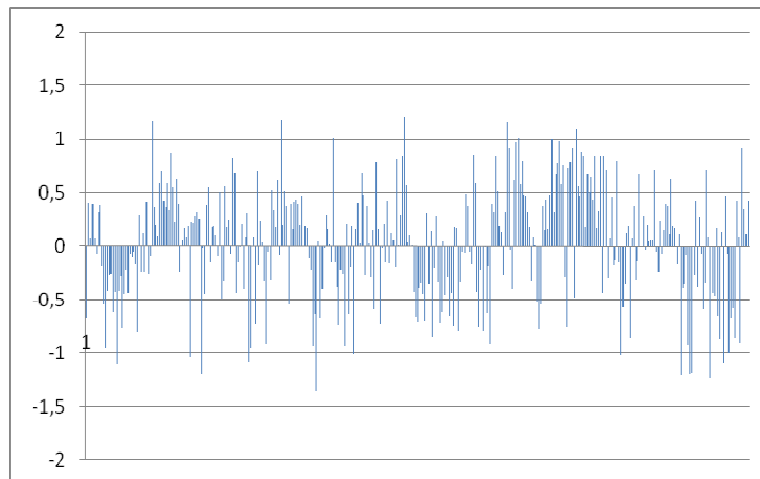


Figura 24- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} com 7 variáveis.

7.4 MODELO COM 47 VARIÁVEIS INDEPENDENTES

7.4.1 Análise das Variáveis

Neste novo modelo foram consideradas 47 variáveis, descritas na Tabela 27, sobre as quais foi calculada a média e o desvio padrão. No entanto as variáveis *n-Olefinas C₄* e *c-Olefinas C₁₀*, não foram consideradas uma vez que eram constantes com média e desvio padrão igual a zero. Assim o número de variáveis foi reduzido para 45. Do conjunto total das 375 amostras foram retiradas aquelas que apresentavam um comportamento atípico, depois da análise de resíduos, e consideradas 347 amostras para desenvolver o modelo pelo método *Stepwise*.

Tabela 27- Estatística descritiva da variável RON e das 47 variáveis independentes.

Variável	Média	Desvio Padrão	N	Variável	Média	Desvio Padrão	N
RON	97,614	1,0634	347	C nr 5 n olef	0,0204	0,01336	347
C nr 4 i par	0,1013	0,33766	347	C nr 6 n olef	0,1002	0,02660	347
C nr 5 i par	2,4249	0,68934	347	C nr 7 n olef	0,0892	0,05666	347
C nr 6 i par	8,2524	1,81958	347	C nr 8 n olef	0,0327	0,03003	347
C nr 7 i par	7,8521	0,88534	347	C nr 9 n olef	0,0085	0,01816	347
C nr 8 i par	2,6129	0,69266	347	C nr 10 n olef	0,0010	0,00585	347
C nr 9 i par	0,4509	0,31231	347	C nr 4 c olef	0,0003	0,00288	347
C nr 10i par	0,0719	0,08038	347	C nr 5 c olef	0,0029	0,01086	347
C nr 11 i par	0,0171	0,01846	347	C nr 6 c olef	0,0342	0,02240	347
C nr 4 n par	0,1386	0,33774	347	C nr 7 c olef	0,0090	0,01677	347
C nr 5 n par	1,4246	0,30561	347	C nr 8 c olef	0,0008	0,00688	347
C nr 6 n par	3,4793	0,56830	347	C nr 9 c olef	0,0001	0,00268	347
C nr 7 n par	2,1453	0,29535	347	C nr 10 c olef	0,0000	0,00000	347
C nr 8 n par	0,6313	0,22005	347	C nr 5 Naft	0,2029	0,10330	347
C nr 9 npar	0,2469	0,15880	347	C nr 6 Naft	0,5639	0,23788	347
C nr 10 npar	0,0076	0,02920	347	C nr 7 Naft	0,2743	0,09908	347
C nr 11 npar	0,0004	0,00282	347	C nr 8 Naft	0,1203	0,05644	347
C nr 4 i olef	0,0008	0,00970	347	C nr 9 Naft	0,0041	0,02267	347
C nr 5 i olef	0,0231	0,03687	347	C nr 10 Naft	0,0003	0,00591	347
C nr 6 i olef	0,2252	0,07468	347	C nr 6 Arom	6,3548	1,13838	347
C nr 7 i olef	0,3226	0,08254	347	C nr 7 Arom	18,0727	1,67272	347
C nr 8 i olef	0,1078	0,06537	347	C nr 8 Arom	23,6677	1,61909	347
C nr 9 i olef	0,0016	0,01399	347	C nr 9 Arom	16,4759	1,66360	347
C nr 10 i olef	0,0002	0,00322	347	C nr 10 Arom	2,7454	0,87844	347
C nr 4 n olef	0,0000	0,00000	347	C nr 10 Arom	2,7454	0,87844	347

De seguida foi analisada a correlação existente entre a variável dependente *RON* e cada uma das variáveis independentes assim como as correlações entre as variáveis independentes.

Verifica-se que a variável independente *Nafténicos C₆* é a que mais se correlaciona com a variável dependente *RON* (0,599), e as variáveis independentes têm uma correlação fraca entre si.

7.4.2 Estimação do Modelo

Os modelos foram desenvolvidos pelo método de selecção de regressores *Stepwise*, num total de 12. A primeira variável a ser introduzida no modelo (M_1) foi a *Nafténicos C₆*, uma vez que é a que apresenta maior correlação com a variável dependente *RON*. As restantes variáveis foram sendo introduzidas progressivamente no modelo de acordo com os princípios deste método de selecção de regressores.

A Tabela 28 mostra os ajustes obtidos para os vários modelos, sendo que o melhor ajuste é obtido para o modelo M_{12} com $r=0,933$; $R^2=0,870$ e $R_a^2=0,766$.

Tabela 28- Características dos modelos com 47 variáveis.

Modelo	r	R ²	R _a ²
M ₁	0,599	0,359	0,357
M ₂	0,716	0,513	0,510
M ₃	0,776	0,602	0,598
M ₄	0,860	0,739	0,736
M ₅	0,887	0,786	0,783
M ₆	0,906	0,822	0,818
M ₇	0,906	0,822	0,819
M ₈	0,916	0,839	0,837
M ₉	0,923	0,852	0,849
M ₁₀	0,929	0,862	0,859
M ₁₁	0,931	0,867	0,864
M ₁₂	0,933	0,870	0,866

O aproveitamento das potencialidades da cromatografia e conseqüente uso do número máximo de variáveis disponíveis (47) melhorou a qualidade do ajuste, o R_a^2 passou de 0,724 para 0,866, para o modelo com melhor ajuste, modelo M_{12} :

$$\begin{aligned}
 \text{RON} = & 93,190 + 0,174 * \text{AromáticosC}_8 - 1,217 * n - \text{ParafinasC}_7 - 0,551 * n - \\
 & \text{ParafinaC}_6 - 1,236 * i - \text{ParafinaC}_9 + 0,226 * \text{AromáticosC}_9 - 0,346 * n - \\
 & \text{ParafinasC}_5 - 50,624 * c - \text{OlefinasC}_9 + 2,222 * \\
 & i - \text{OlefinasC}_5 + 0,083 * \text{AromáticosC}_7 + 0,094 * \text{AromáticosC}_6
 \end{aligned}$$

De acordo com o teste *F*, Tabela 29, todos os modelos ajustados são globalmente significativos, isto é $p\text{-value} \leq \alpha$ ($\alpha=0,05$). Assim rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) [31].

Tabela 29- Resultados da ANOVA aos modelos com 47 variáveis.

Modelo	Soma dos quadrados	Graus Liberdade	Média da Soma dos Quadrados	F	<i>p-value</i>
M ₁ Regressão	140,305	1	140,305	192,874	0,000
Residual	250,969	345	0,727		
Total	391,274	346			
M ₂ Regressão	200,638	2	100,319	181,025	0,000
Residual	190,635	344	0,554		
Total	391,274	346			
M ₃ Regressão	235,446	3	78,482	172,751	0,000
Residual	155,828	343	0,454		
Total	391,274	346			
M ₄ Regressão	289,113	4	72,278	241,964	0,000
Residual	102,160	342	0,299		
Total	391,274	346			
M ₅ Regressão	307,676	5	61,535	251,006	0,000
Residual	83,598	341	0,245		
Total	391,274	346			
M ₆ Regressão	321,474	6	53,579	260,987	0,000
Residual	69,800	340	0,205		
Total	391,274	346			
M ₇ Regressão	321,459	5	64,292	314,026	0,000
Residual	69,814	341	0,205		
Total	391,274	346			
M ₈ Regressão	328,447	6	54,741	296,242	0,000
Residual	62,827	340	0,185		
Total	391,274	346			
M ₉ Regressão	333,485	7	47,641	279,469	0,000
Residual	57,789	339	0,170		
Total	391,274	346			
M ₁₀ Regressão	337,384	8	42,173	264,513	0,000
Residual	53,890	338	0,159		
Total	391,274	346			
M ₁₁ Regressão	339,376	9	37,708	244,862	0,000

M ₁₂	Residual	51,898	337	0,154	225,535	0,000
	Total	391,274	346			
	Regressão	340,540	10	34,054		
	Residual	50,733	336	0,151		
	Total	391,274	346			

Apesar de estarem disponíveis as 47 variáveis, este método de selecção de regressores seleccionou apenas 10 variáveis como explicativas do modelo que apresenta melhor ajuste M₁₂, Tabela 30:

Tabela 30- Coeficientes do modelo com melhor ajuste (M₁₂), com 47 variáveis.

Modelo	Coeficientes		t	p-value	Estatísticas de Colinearidade	
	B	Erro Padrão			Tolerância	VIF
M ₁₂ (Constante)	93,190	1,651	56,455	0,000		
C nr 8 Arom	0,174	0,021	8,198	0,000	0,368	2,715
C nr 7 n par	-1,217	0,150	-8,086	0,000	0,221	4,526
C nr 6 n par	-0,551	0,073	-7,557	0,000	0,255	3,928
C nr 9 i par	-1,236	0,108	-11,420	0,000	0,382	2,616
C nr 9 Arom	0,226	0,025	9,022	0,000	0,250	3,996
C nr 5 n par	-0,346	0,103	-3,378	0,001	0,444	2,251
C nr 9 c olef	-50,624	8,201	-6,173	0,000	0,901	1,110
C nr 5 i olef	2,222	0,616	3,610	0,000	0,847	1,180
C nr 7 Arom	0,083	0,026	3,236	0,001	0,236	4,243
C nr 6 Arom	0,094	0,034	2,777	0,006	0,296	3,374

7.4.3 Validação dos pressupostos do modelo estimado

7.4.3.1 Homocedasticidade

Os resíduos mantêm uma amplitude constante relativamente ao eixo horizontal zero, Figura 25, logo não se rejeita a hipótese da homocedasticidade.

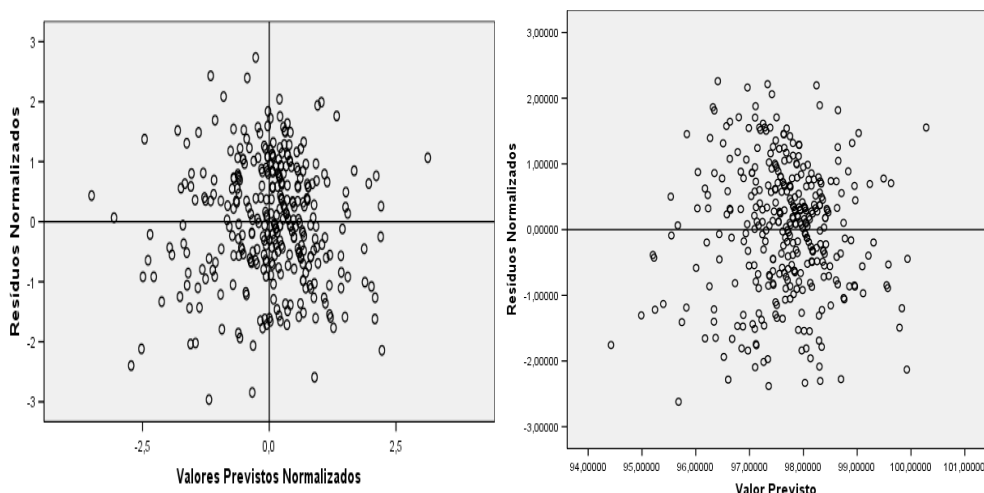


Figura 25- Avaliação da Homocedasticidade para o modelo M_{12} com 47 variáveis.

7.4.3.2 Covariância Nula

De acordo com o teste de Durbin-Watson o modelo M_{12} apresenta um valor de teste $d=1,438$, logo existe auto-correlação entre os resíduos.

7.4.3.3 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, Tabela 31.

Graficamente os valores esperados e observados normalizados dispõem-se ao longo da recta oblíqua; os desvios normalizados também se dispõem ao longo da recta horizontal, sendo que apresentam uma menor amplitude que nos modelos anteriores e uma menor tendência oscilatória, Figura 26.

Tabela 31- Teste à Normalidade dos resíduos, modelo M_{12} com 47 variáveis.

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,034	347	0,200

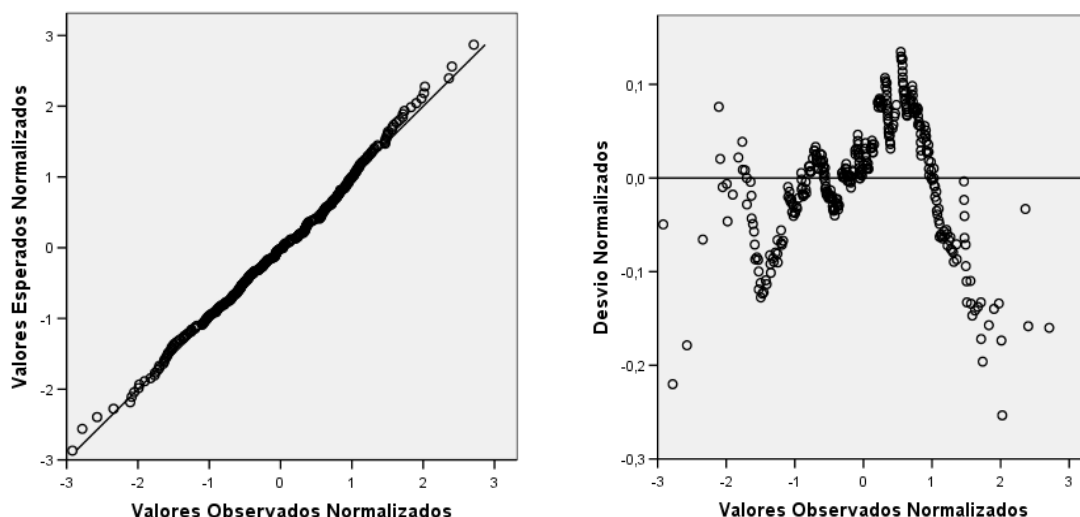


Figura 26- Análise gráfica da normalidade dos resíduos, para o modelo M_{12} com 47 variáveis.

7.4.3.4 Multicolinearidade

Não se verifica a existência de multicolinearidade entre as variáveis independentes. Esta análise é feita com base na matriz das correlações e nos valores da Tolerância e VIF, Tabela 30.

7.4.4 Previsão do RON usando o modelo

Com base no modelo desenvolvido é possível calcular o valor de RON, a partir da composição cromatográfica, para as amostras usadas na sua elaboração. Os valores obtidos são comparados com os obtidos pelo CFR através de uma regressão linear, Figura 27, e pelos *testes t e F*, Tabela 32.

Da comparação dos resultados por regressão linear, verifica-se que os resultados para o RON obtidos por este modelo são mais próximos dos obtidos pelo método do motor CFR, uma vez que a recta obtida se aproxima mais da relação ideal e apresenta um coeficiente de determinação superior:

$$RON_{\text{modelo}} = 12,67 + 0,87RON_{\text{CFR}} \text{ com } R^2 = 0,87.$$

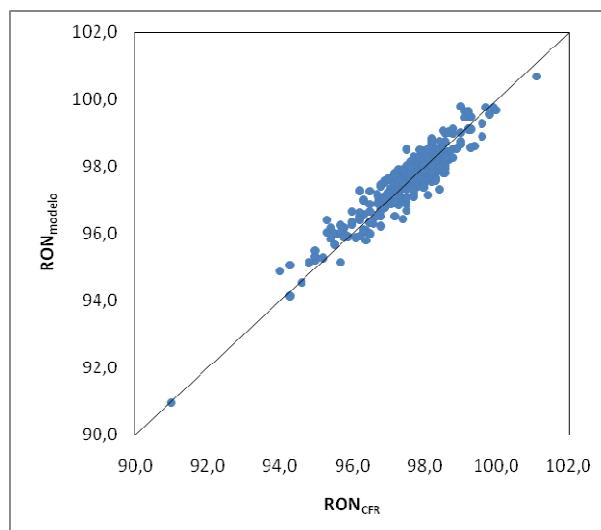


Figura 27- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis.

Dos *testes t e F* concluí-se que não existem desvios significativos entre os resultados obtidos por um ou outro método.

Relativamente às diferenças obtidas entre os dois métodos estas também diminuíram e a percentagem de resultados cuja diferença é superior a $|0,7|$ diminuiu drasticamente de 20% para 5% dos resultados, Figura 28.

Tabela 32- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis.

	RON_{CFR}	RON_{modelo}
Média	97,6	97,6
Desvio Padrão	1,06	0,99
Variância	1,13	0,98
n	347	347
F_{calc}		1,15
F_{tab}		1,27
$F_{calc} < F_{tab}$		
t_{calc}		0,24
t_{tab}		1,96
$t_{calc} < t_{tab}$		

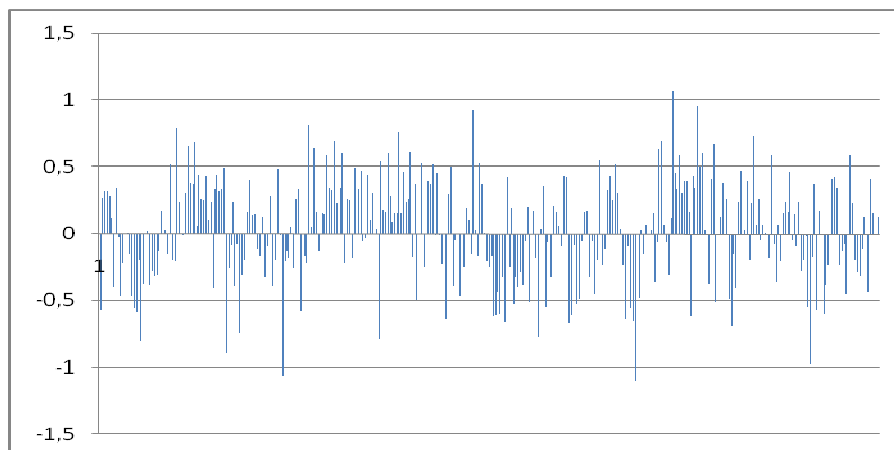


Figura 28- Diferença entre os resultados obtidos com o modelo e com o motor CFR, com 47 variáveis.

7.5 MODELO COM 47 VARIÁVEIS NORMALIZADAS

Uma vez que as concentrações médias de cada uma das 47 variáveis apresentam uma vasta amplitude, estas foram normalizadas antes de desenvolver um novo modelo.

Para isso foi calculada a média e o desvio padrão de cada uma das variáveis e estas foram transformadas em variáveis normalizadas Z_i , eq.27:

$$Z_i = \frac{X_i - \bar{X}}{\text{desv. pad}} \quad (\text{eq.27})$$

Foram excluídas as variáveis *n-Olefinas* C_4 e *n-Olefinas* C_{10} pelo facto de apresentarem desvio padrão igual a zero e não ser possível a sua normalização.

No entanto os resultados obtidos pelo modelo com as variáveis normalizadas foram iguais aos obtidos no ponto 7.4, sem as variáveis normalizadas. Uma vez que não se obtêm melhorias esta hipótese foi abandonada.

7.6 MODELO COM 47 VARIÁVEIS INDEPENDENTES – DADOS DE 2003

De acordo com os resultados obtidos nos modelos anteriores verifica-se que o melhor modelo é aquele para o qual foram consideradas as 47 variáveis, desenvolvido pelo método de selecção de regressores *Stepwise*.

A fim de validar os modelos e verificar se existem variações ao longo do tempo, os dados foram divididos em grupos e criados modelos a partir desses grupos. O primeiro modelo foi desenvolvido a partir dos dados do ano de 2003, e aplicado às restantes amostras (2004 a 2006), a fim de verificar a resposta do modelo a amostras que não foram usadas na sua elaboração.

7.6.1 Análise das Variáveis

Neste novo modelo foram consideradas 47 variáveis, sobre as quais foi calculada a média e o desvio padrão, Tabela 33. As variáveis *n-Olefinas C₄* e *c-Olefinas C₁₀*, foram rejeitadas uma vez que eram constantes com média e desvio padrão igual a zero, logo o número de variáveis passou para 45. Do conjunto total de 123 amostras, foram retiradas as amostras atípicas e consideradas 112 amostras para desenvolver o modelo.

Neste caso, considerando as amostras apenas de 2003, a variável *n-Parafinas C₉* é aquela que apresenta uma correlação mais forte com o RON e por isso a primeira a entrar no modelo.

Tabela 33- Estatística descritiva da variável RON e das 47 variáveis independentes (2003).

	Média	Desvio Padrão	N		Média	Desvio Padrão	N
RON	97,904	1,1387	112	C nr 5 n olef	0,0273	0,00968	112
C nr 4 i par	0,0584	0,19876	112	C nr 6 n olef	0,1038	0,02476	112
C nr 5 i par	2,2688	0,36628	112	C nr 7 n olef	0,0920	0,02760	112
C nr 6 i par	7,3221	1,03483	112	C nr 8 n olef	0,0326	0,02003	112
C nr 7 i par	7,4436	0,79744	112	C nr 9 n olef	0,0049	0,01230	112
C nr 8 i par	2,6046	0,69533	112	C nr 10 n olef	0,0007	0,00439	112
C nr 9 i par	0,4732	0,26530	112	C nr 4 c olef	0,0004	0,00378	112
C nr 10i par	0,0818	0,06900	112	C nr 5 c olef	0,0032	0,01015	112
C nr 11 i par	0,0231	0,01617	112	C nr 6 c olef	0,0388	0,01693	112
C nr 4 n par	0,1504	0,34447	112	C nr 7 c olef	0,0192	0,02023	112
C nr 5 n par	1,3689	0,22436	112	C nr 8 c olef	0,0023	0,01185	112
C nr 6 n par	3,2740	0,46438	112	C nr 9 c olef	0,0004	0,00472	112
C nr 7 n par	2,1475	0,30998	112	C nr 10 c olef	0,0000	0,00000	112
C nr 8 n par	0,6240	0,21690	112	C nr 5 Naft	0,1703	0,07399	112
C nr 9 npar	0,2589	0,09137	112	C nr 6 Naft	0,5188	0,14579	112
C nr 10 npar	0,0080	0,02650	112	C nr 7 Naft	0,2579	0,06738	112

C nr 11 npar	0,0014	0,00517	112	C nr 8 Naft	0,1249	0,04289	112
C nr 4 i olef	0,0011	0,00798	112	C nr 9 Naft	0,0046	0,02620	112
C nr 5 i olef	0,0390	0,04118	112	C nr 10 Naft	0,0010	0,01039	112
C nr 6 i olef	0,2711	0,05582	112	C nr 6 Arom	6,7641	1,13211	112
C nr 7 i olef	0,3721	0,06739	112	C nr 7 Arom	18,2303	1,77145	112
C nr 8 i olef	0,1395	0,04891	112	C nr 8 Arom	23,4730	1,32037	112
C nr 9 i olef	0,0035	0,02249	112	C nr 9 Arom	17,1788	1,48207	112
C nr 10 i olef	0,0005	0,00567	112	C nr 10 Arom	3,3575	0,86683	112
Cnr 4 n olef	0,0000	0,0000	112				

7.6.2 Estimação do Modelo

O método *Stepwise* desenvolveu um total de 17 modelos até que todas as variáveis não introduzidas no modelo não tivessem capacidade explicativa da variável dependente RON e todas as que estão no modelo o tivessem [29].

A Tabela 34 mostra os ajustes obtidos para os vários modelos; o melhor ajuste corresponde ao modelo M_{17} , com $r=0,974$; $R^2=0,949$ e $R_a^2=0,942$.

Embora se parta de um conjunto inicial de 45 variáveis o modelo obtido por este método usa apenas 13, Tabela 35. Algumas destas variáveis são comuns às consideradas como significativas para o modelo anterior (ponto 7.4) no entanto verifica-se uma forte presença das variáveis Nafténicos, não incluídas no último modelo M_{17} :

$$\begin{aligned}
 \text{RON} = & 98,292 - 4,249 * i\text{-ParafinasC}_8 - 0,313 * i\text{-} \\
 & \text{ParafinasC}_6 + 0,241 * \text{AromáticosC}_9 + 0,089 * \text{AromáticosC}_7 + 6,232 * i\text{-} \\
 & \text{ParafinasC}_{10} + 1,533 * \text{NafténicosC}_5 - 0,970 * n\text{-ParafinasC}_5 - 1,838 * n\text{-} \\
 & \text{ParafinasC}_7 - 24,662 * \text{NafténicosC}_{10} + 7,362 * \text{NafténicosC}_9 + 0,320 * i\text{-} \\
 & \text{ParafinasC}_7 + 9,864 * n\text{-OlefinasC}_5 - 0,326 * i\text{-ParafinasC}_4
 \end{aligned}$$

Tabela 34- Características dos modelos com 47 variáveis (2003).

Modelo	r	R ²	R _a ²
M ₁	0,778	0,606	0,602
M ₂	0,886	0,785	0,781
M ₃	0,928	0,860	0,857
M ₄	0,940	0,884	0,880
M ₅	0,948	0,899	0,894
M ₆	0,955	0,912	0,907
M ₇	0,958	0,917	0,912
M ₈	0,957	0,915	0,910

M ₉	0,959	0,919	0,913
M ₁₀	0,960	0,922	0,916
M ₁₁	0,963	0,928	0,922
M ₁₂	0,966	0,934	0,927
M ₁₃	0,968	0,938	0,931
M ₁₄	0,968	0,937	0,931
M ₁₅	0,970	0,941	0,935
M ₁₆	0,973	0,946	0,940
M ₁₇	0,974	0,949	0,942

O modelo M₁₇ é significativo na globalidade, logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para o modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31]. A Tabela 35 apresenta os coeficientes do modelo M₁₇, o seu desvio padrão, os testes *t*, e as tolerâncias de cada variável. Os testes *t* permitem testar a hipótese nula de inexistência de uma relação linear entre a variável dependente RON com cada uma das variáveis independentes, bem como com as constantes consideradas no modelo [31]. Verifica-se que todos os testes *t* têm associado um nível de significância inferior a 0,05, excepto o do coeficiente associado ao regressor *Aromáticos C₇*; este último é justificado pelo facto deste coeficiente apresentar um valor muito próximo de zero (0,089). O resultado destes testes permitem-nos concluir que todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes de cada variável independente são diferentes de zero, mesmo o associado ao regressor *Aromáticos C₇* [31].

Tabela 35- Coeficientes do modelo com melhor ajuste M₁₇, com 47 variáveis (2003).

Modelo	Coeficientes		t	p-value	Estatísticas de Colinearidade	
	B	Erro Padrão			Tolerância	VIF
M ₁₇ (Constante)	98,292	2,312	42,508	0,000		
C nr 9 i par	-4,249	0,246	-	0,000	0,159	6,293
C nr 6 i par	-0,313	0,055	-5,716	0,000	0,210	4,756
C nr 9 Arom	0,241	0,046	5,212	0,000	0,143	6,989
C nr 7 Arom	0,089	0,048	1,870	0,065	0,094	10,641
C nr 10i par	6,232	0,756	8,243	0,000	0,248	4,034
C nr 5 Naft	1,533	0,454	3,377	0,001	0,598	1,673
C nr 5 n par	-0,970	0,161	-6,030	0,000	0,518	1,930

C nr 7 n par	-1,838	0,351	-5,237	0,000	0,057	17,555
C nr 10 Naft	-	4,534	-5,440	0,000	0,304	3,292
C nr 9 Naft	7,362	1,923	3,829	0,000	0,266	3,761
C nr 7 i par	0,320	0,100	3,201	0,002	0,106	9,399
C nr 5 n olef	9,864	3,004	3,284	0,001	0,798	1,253
C nr 4 i par	-0,326	0,141	-2,317	0,023	0,863	1,158

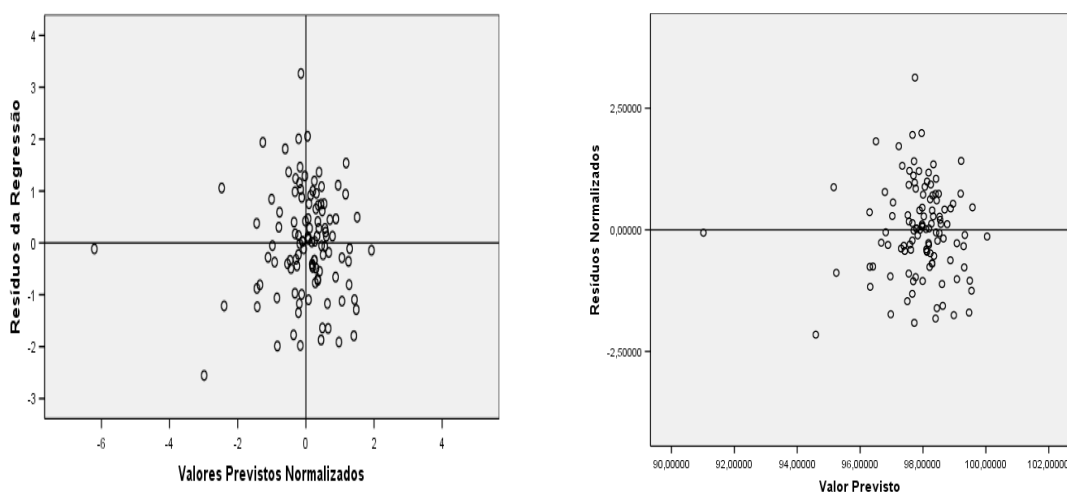
7.6.3 Validação dos pressupostos do modelo

7.6.3 Homocedasticidade

Os resíduos não apresentam tendências crescentes ou decrescentes, pelo que não se rejeita a hipótese da homocedasticidade, [Figura 29](#).

7.6.3.1 Covariância Nula

De acordo com o teste de Durbin-Watson, $d=2,184$, não existe auto-correlação entre os resíduos.



[Figura 29](#)- Avaliação da Homocedasticidade para o modelo M_{17} com 47 variáveis (2003).

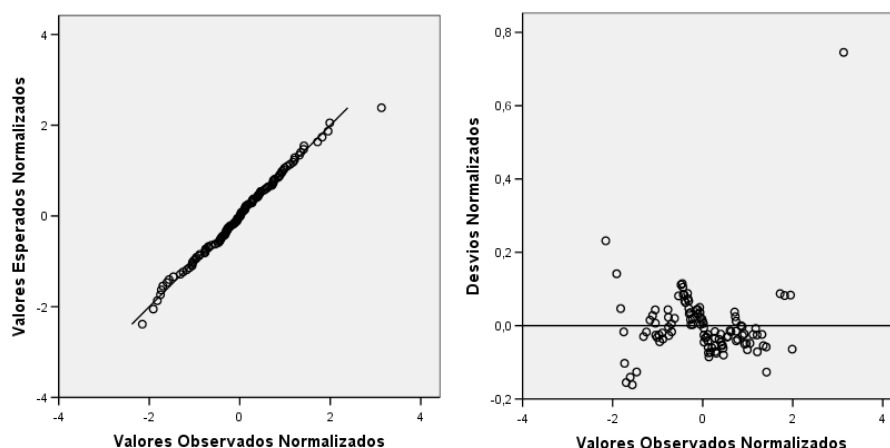
7.6.3.2 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, [Tabela 36](#); esta hipótese é

confirmada graficamente uma vez que as observações se dispõem ao longo das rectas oblíqua e horizontal, [Figura 30](#).

[Tabela 36](#)- Teste à Normalidade dos resíduos, modelo M_{17} com 47 variáveis (2003).

	Kolmogorov-Smirnov		
	Estatística	Graus Liberdade	Significância
Resíduos Padrão	0,047	112	0,200



[Figura 30](#)- Análise gráfica da normalidade dos resíduos, para o modelo M_{17} , com 47 variáveis (2003).

7.6.3.3 Multicolinearidade

Neste modelo verifica-se a existência de multicolinearidade para as variáveis independentes com 7 átomos de carbono (*i-Parafinas C₇*, *n-Parafinas C₇* e *Aromáticos C₇*), confirmado pelos valores da Tolerância e do VIF, [Tabela 35](#).

7.6.4 Previsão do RON usando o modelo

O modelo desenvolvido foi aplicado nas amostras de 2004 a 2006. O objectivo é verificar a resposta do mesmo quando aplicado a amostras que não contribuíram para o seu desenvolvimento, [Figura 31](#).

Embora este modelo seja o que o apresenta melhor ajuste, dos desenvolvidos até aqui, verifica-se que quando aplicado a outras amostras existe uma grande dispersão de resultados e uma diferença muito elevada entre o RON_{CFR} e RON_{modelo} , [Figura 32](#); em 43% dessas amostras a diferença é superior a $|0,7|$.

No entanto quando comparados os resultados obtidos pelos dois métodos pelos testes F e t , não são encontradas diferenças significativas nas variâncias, nem nas médias, respectivamente. Concluí-se portanto que não existem desvios significativos entre os resultados obtidos pelos dois métodos, Tabela 37.

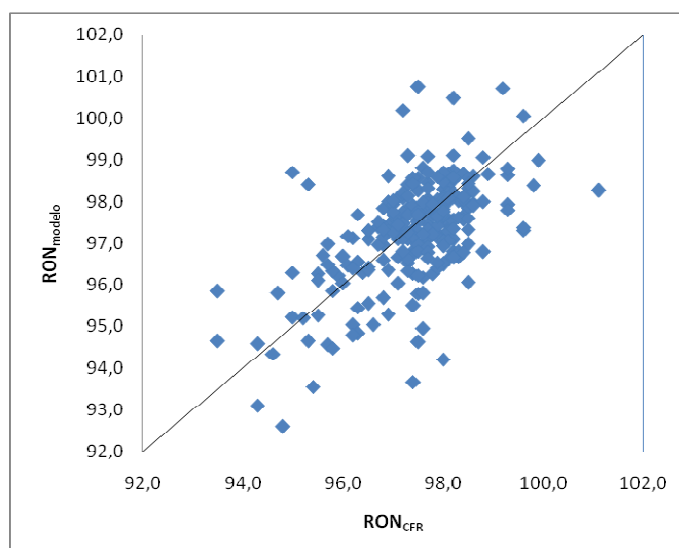


Figura 31- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis (2003).

Tabela 37- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis (2003).

	RON_{CFR}	RON_{modelo}
Média	97,4	97,3
Desvio Padrão	1,09	1,24
Variância	1,19	1,53
n	252	252
F_{calc}		0,78
F_{tab}		1,18
$F_{calc} < F_{tab}$		
t_{calc}		0,98
t_{tab}		1,96
$t_{calc} < t_{tab}$		

É de salientar que a diferença aumenta com o afastamento temporal das amostras, às quais é aplicado o modelo, relativamente às que foram usadas na sua elaboração (amostras de 2003). O que pode ser resultado de pequenas variação da matriz ao longo do tempo, Figura 32.

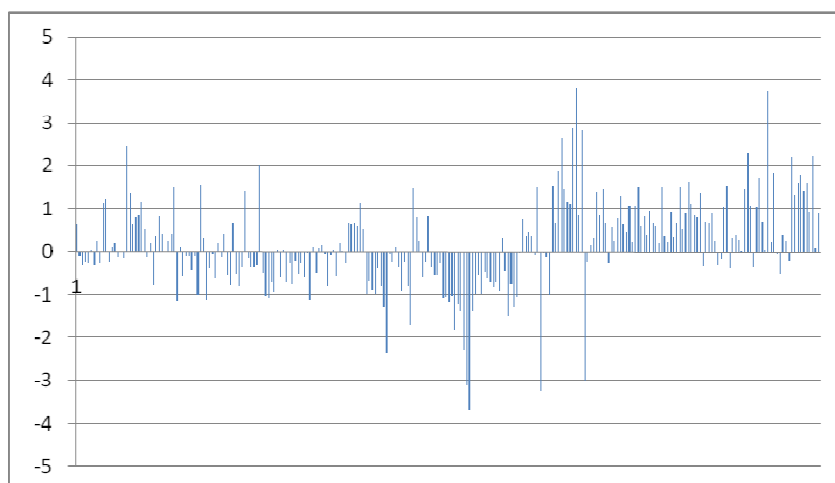


Figura 32- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} , quando aplicado às amostras de 2004 a 2006.

7.7 MODELO COM 47 VARIÁVEIS COM OS DADOS DE 2003 A 2004

7.7.1 Análise das Variáveis

Neste modelo foram consideradas novamente 47 variáveis, descritas na Tabela 38, sobre as quais foi calculada a média e o desvio padrão. As variáveis *n-Olefinas* C_4 , *c-Olefinas* C_9 e *c-Olefinas* C_{10} , foram excluídas por serem constantes com média e desvio padrão igual a zero. Do conjunto total foram retiradas as amostras atípicas e consideradas 252 amostras para desenvolver o modelo.

7.7.2 Estimação do Modelo

O modelo foi desenvolvido novamente pelo método de selecção de regressores *Stepwise*, em que a primeira variável a ser introduzida foi a *Nafténicos* C_6 ,

devido à forte correlação com a variável dependente RON. É de notar que esta também foi a primeira variável a ser introduzida, quando do desenvolvimento do primeiro modelo com 47 variáveis.

O melhor ajuste corresponde ao modelo M_{15} , com $r=0,925$, $R^2=0,856$ e $R_a^2=0,849$, Tabela 39. Este modelo escolheu como regressores 11 das 45 variáveis independentes inicialmente consideradas, modelo M_{15} :

$$RON = 106,119 - 1,071 * i\text{-Nafténicos}C_6 - 0,247 * i\text{-Parafinas}C_7 - 0,517 * n\text{-Parafinas}C_5 - 0,829 * n\text{-Parafinas}C_7 - 0,930 * n\text{-Parafinas}C_6 - 7,897 * n\text{-Parafinas}C_{10} + 0,056 * \text{Aromáticos}C_9 - 2,755 * i\text{-Olefinas}C_8 - 0,279 * n\text{-Parafinas}C_4 - 1,088 * n\text{-Parafinas}C_8 - 19,420 * n\text{-Parafinas}C_{11}$$

O número de regressores e os próprios regressores são diferentes dos usados nos modelos anteriores, também com 47 variáveis, Tabela 40.

O modelo é significativo em termos globais, logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

Tabela 38- Estatística descritiva das 47 variáveis (2003 a 2004).

	Média	Desvio Padrão	N		Média	Desvio Padrão	N
RON	97,608	1,0381	252	C nr 4 n olef	0,0000	0,00000	252
C nr 4 i par	0,0856	0,30456	252	C nr 5 n olef	0,0235	0,01233	252
C nr 5 i par	2,2975	0,48411	252	C nr 6 n olef	0,1043	0,02474	252
C nr 6 i par	7,7050	1,61151	252	C nr 7 n olef	0,0934	0,03094	252
C nr 7 i par	7,8151	0,82991	252	C nr 8 n olef	0,0346	0,02508	252
C nr 8 i par	2,6777	0,64169	252	C nr 9 n olef	0,0112	0,02093	252
C nr 9 i par	0,4806	0,29410	252	C nr 10 n olef	0,0013	0,00670	252
C nr 10i par	0,0908	0,07705	252	C nr 4 c olef	0,0004	0,00338	252
C nr 11 i par	0,0228	0,01826	252	C nr 5 c olef	0,0035	0,01110	252
C nr 4 n par	0,1454	0,35733	252	C nr 6 c olef	0,0403	0,01902	252
C nr 5 n par	1,4328	0,31781	252	C nr 7 c olef	0,0117	0,01787	252
C nr 6 n par	3,3933	0,57146	252	C nr 8 c olef	0,0007	0,00621	252
C nr 7 n par	2,1642	0,27820	252	C nr 9 c olef	0,0000	0,00000	252
C nr 8 n par	0,6483	0,19463	252	C nr 10 c olef	0,0000	0,00000	252
C nr 9 npar	0,2937	0,13861	252	C nr 5 Naft	0,2042	0,11099	252
C nr 10 npar	0,0077	0,01873	252	C nr 6 Naft	0,6150	0,21978	252
C nr 11 npar	0,0006	0,00351	252	C nr 7 Naft	0,2885	0,09135	252

C nr 4 i olef	0,0011	0,01138	252	C nr 8 Naft	0,1294	0,04294	252
C nr 5 i olef	0,0292	0,04344	252	C nr 9 Naft	0,0040	0,02216	252
C nr 6 i olef	0,2409	0,06500	252	C nr 10 Naft	0,0004	0,00693	252
C nr 7 i olef	0,3526	0,06000	252	C nr 6 Arom	6,3700	1,15233	252
C nr 8 i olef	0,1278	0,05759	252	C nr 7 Arom	18,1644	1,70049	252
C nr 9 i olef	0,0023	0,01665	252	C nr 8 Arom	23,3187	1,51487	252
C nr 10 i olef	0,0002	0,00378	252	C nr 9 Arom	16,7770	1,67006	252
				C nr 10 Arom	3,0210	0,85587	252

Tabela 39- Características dos modelos com 47 variáveis (2003 a 2004).

Modelo	r	R ²	R _a ²
M ₁	0,741	0,548	0,547
M ₂	0,816	0,665	0,662
M ₃	0,848	0,720	0,716
M ₄	0,861	0,741	0,737
M ₅	0,878	0,772	0,767
M ₆	0,903	0,815	0,811
M ₇	0,902	0,814	0,811
M ₈	0,908	0,825	0,820
M ₉	0,907	0,823	0,820
M ₁₀	0,912	0,832	0,828
M ₁₁	0,915	0,837	0,832
M ₁₂	0,918	0,843	0,838
M ₁₃	0,921	0,849	0,843
M ₁₄	0,923	0,853	0,846
M ₁₅	0,925	0,856	0,849

De acordo com o teste *t*, **Tabela 40**, todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes das variáveis independente são diferentes de zero, para o modelo M₁₅ [31].

Tabela 40- Coeficientes do modelo com melhor ajuste M₁₅, com 47 variáveis (2003 a 2004).

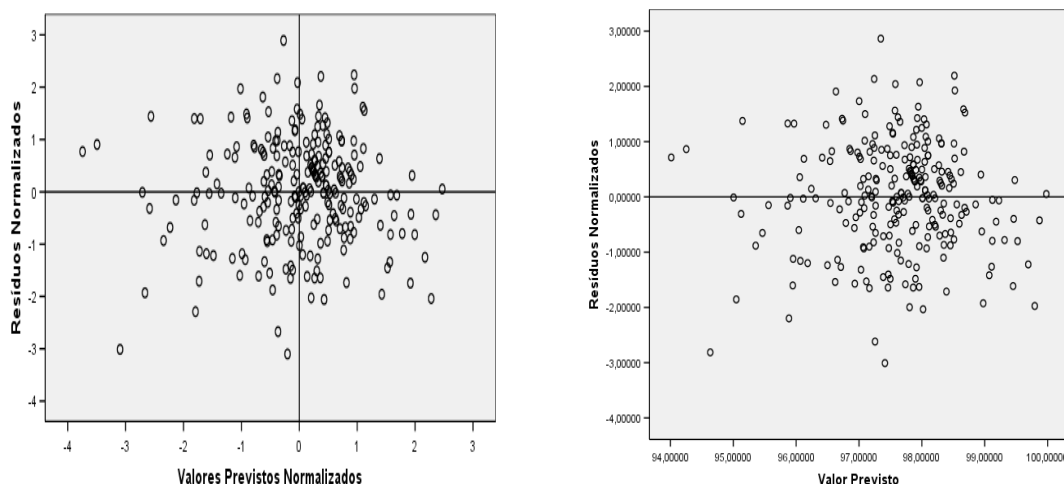
Modelo	Coeficientes		t	p-value	Estatísticas de Colinearidade		
	B	Erro Padrão			Tolerancia	VIF	
M ₁₅	Constante	106,119	0,506	209,839	0,000		
	C nr 6 Naft	-1,071	0,185	-5,795	0,000	0,392	2,551
	C nr 7 i par	-0,247	0,055	-4,525	0,000	0,314	3,186
	C nr 5 n par	-0,517	0,097	-5,334	0,000	0,680	1,471
	C nr 7 n par	-0,829	0,300	-2,768	0,006	0,093	10,744
	C nr 6 n par	-0,930	0,065	-14,387	0,000	0,474	2,112
	C nr 10 npar	-7,897	1,846	-4,278	0,000	0,541	1,849
	C nr 9 Arom	0,056	0,020	2,804	0,005	0,576	1,736
	C nr 8 i olef	-2,755	0,617	-4,467	0,000	0,512	1,952
	C nr 4 n par	-0,279	0,076	-3,651	0,000	0,869	1,151

C nr 8 n par	-1,088	0,394	-2,759	0,006	0,110	9,115
C nr 11 npar	-19,420	7,994	-2,429	0,016	0,819	1,220

7.7.3 Diagnóstico do Modelo Estimado

7.7.3.1 Homocedasticidade

De acordo com a distribuição dos resíduos, [Figura 33](#), não se rejeita a hipótese da homocedasticidade.



[Figura 33](#) – Avaliação da Homocedasticidade para o modelo M_{15} , com 47 variáveis (2003 a 2004).

7.7.3.1 Covariância Nula

De acordo com o teste de Durbin-Watson o modelo M_{15} apresenta um valor de teste $d=1,646$, logo existe auto-correlação entre os resíduos.

7.7.3.2 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, [Tabela 41](#); esta hipótese é confirmada graficamente uma vez que as observações se dispõem ao longo das rectas oblíqua e horizontal, [Figura 34](#).

[Tabela 41](#)- Teste à Normalidade dos resíduos, modelo M_{15} , com 47 variáveis (2003 a 2004).

	Kolmogorov-Smirnov		
	Estatística	Graus	Significância

		Liberdade	
Resíduos Padrão	0,047	252	0,200

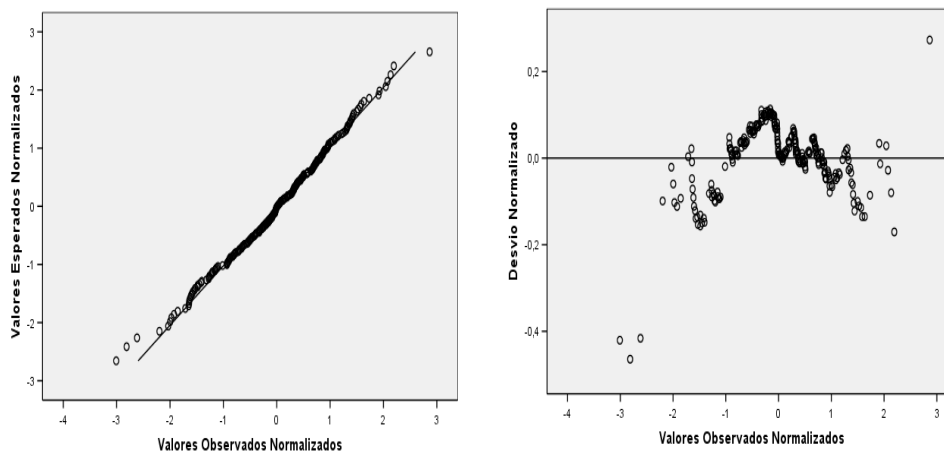


Figura 34- Análise gráfica da normalidade dos resíduos, para o modelo M_{15} com 47 variáveis (2003 a 2004).

7.7.3.4 Multicolinearidade

Não se verifica a existência de multicolinearidade entre as variáveis independentes. Esta análise é feita com base na matriz das correlações e nos valores da Tolerância e VIF, Tabela 40.

7.7.4 Previsão do RON usando o modelo

O modelo desenvolvido neste ponto foi aplicado nas amostras de 2005 a 2006, com o objectivo de verificar a resposta do modelo quando aplicado em amostras independentes ao seu desenvolvimento.

Embora o modelo apresente um bom ajuste quando aplicado a outras amostras apresenta uma grande dispersão de resultados, tal como no modelo anterior. Há uma diferença grande entre o RON_{CFR} e RON_{modelo} , Figura 35.

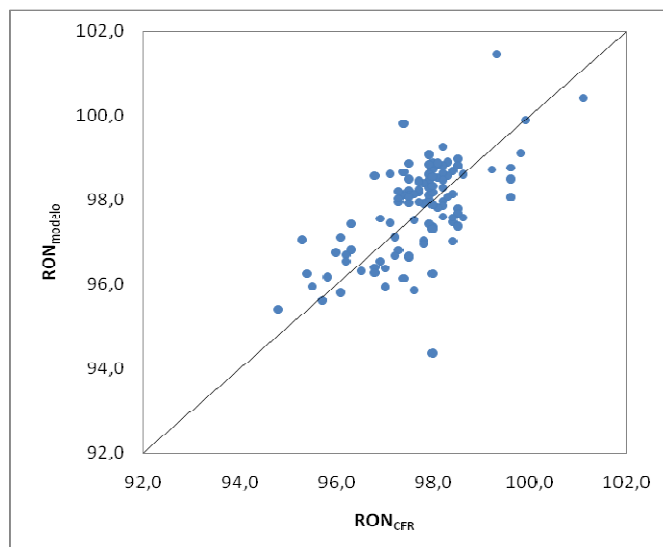


Figura 35- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis (2003 a 2004).

Tabela 42- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis (2003 a 2004).

	RON_{CFR}	RON_{modelo}
Média	97,7	97,7
Desvio Padrão	0,98	1,72
Variância	0,97	2,95
n	112	112
F _{calc}		1,15
F _{tab}		1,15
$F_{calc} = F_{tab}$		
t _{calc}		0,003
t _{tab}		1,96
$T_{calc} < t_{tab}$		

No entanto de acordo com o teste *t* das médias concluí-se que não existem desvios significativos entre os resultados obtidos por um ou outro método, Tabela 42. Relativamente às diferenças obtidas entre os dois métodos mantêm-se relativamente ao método desenvolvido apenas com os dados de 2003. A percentagem de resultados cuja diferença é superior a |0,7| manteve-se em 43% dos resultados, Figura 36.

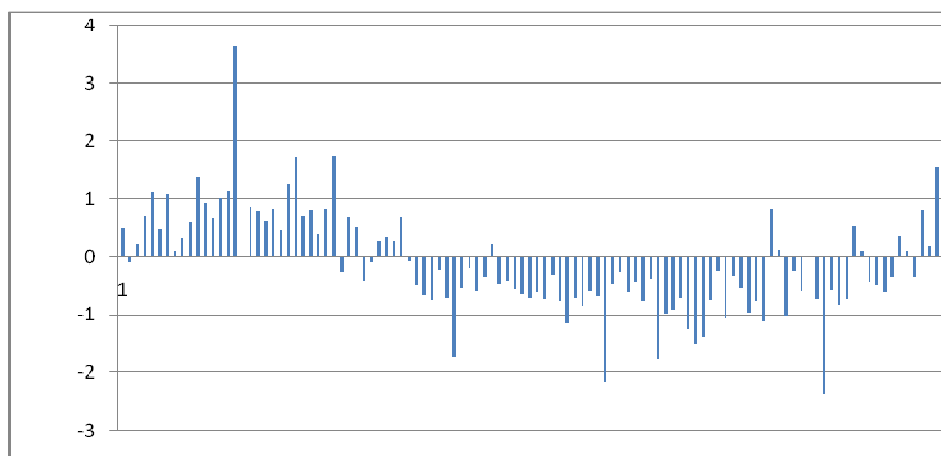


Figura 36- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} com 47 variáveis (2003 a 2004).

7.8 MODELO COM 47 VARIÁVEIS COM OS DADOS DE 2003 A 2005

7.8.1 Análise das Variáveis Independentes

Neste modelo foram consideradas novamente 47 variáveis, descritas na [Tabela 43](#), sobre as quais foi calculada a média e o desvio padrão. As variáveis *n-Olefinas* C_4 , *c-Olefinas* C_9 e *c-Olefinas* C_{10} , foram excluídas por serem constantes com média e desvio padrão igual a zero. Do conjunto total foram retiradas as amostras atípicas e consideradas 334 amostras para desenvolver o modelo.

7.8.2 Estimação do Modelo

O modelo foi desenvolvido novamente pelo método de selecção de regressores *Stepwise*, em que a primeira variável a ser introduzida foi a *Nafténicos* C_6 , devido à forte correlação com a variável dependente RON.

Tabela 43- Estatística descritiva das 47 variáveis (2003 a 2005).

	Média	Desvio Padrão	N		Média	Desvio Padrão	N
RON	97,667	0,9895	334	C nr 5 n olef	0,0203	0,01368	334
C nr 4 i par	0,0895	0,31036	334	C nr 6 n olef	0,1007	0,02626	334
C nr 5 i par	2,4061	0,62214	334	C nr 7 n olef	0,0887	0,03739	334
C nr 6 i par	8,2501	1,86453	334	C nr 8 n olef	0,0332	0,03026	334
C nr 7 i par	7,8428	0,80981	334	C nr 9 n olef	0,0093	0,01896	334
C nr 8 i par	2,5697	0,64926	334	C nr 10 n olef	0,0011	0,00596	334
C nr 9 i par	0,4323	0,29325	334	C nr 4 c olef	0,0003	0,00294	334

C nr 10i par	0,0742	0,08071	334	C nr 5 c olef	0,0026	0,00975	334
C nr 11 i par	0,0176	0,01859	334	C nr 6 c olef	0,0346	0,02197	334
C nr 4 n par	0,1435	0,34985	334	C nr 7 c olef	0,0089	0,01635	334
C nr 5 n par	1,4469	0,35722	334	C nr 8 c olef	0,0006	0,00550	334
C nr 6 n par	3,4405	0,60413	334	C nr 9 c olef	0,0000	0,00000	334
C nr 7 n par	2,1210	0,30593	334	C nr 10 c olef	0,0000	0,00000	334
C nr 8 n par	0,6327	0,29128	334	C nr 5 Naft	0,2067	0,10584	334
C nr 9 npar	0,2561	0,19210	334	C nr 6 Naft	0,5636	0,23654	334
C nr 10 npar	0,0090	0,04254	334	C nr 7 Naft	0,2745	0,09488	334
C nr 11 npar	0,0019	0,02696	334	C nr 8 Naft	0,1175	0,05043	334
C nr 4 i olef	0,0008	0,00989	334	C nr 9 Naft	0,0030	0,01932	334
C nr 5 i olef	0,0222	0,03973	334	C nr 10 Naft	0,0003	0,00602	334
C nr 6 i olef	0,2263	0,07250	334	C nr 6 Arom	6,3404	1,14254	334
C nr 7 i olef	0,3282	0,08070	334	C nr 7 Arom	18,0660	1,71770	334
C nr 8 i olef	0,1093	0,06567	334	C nr 8 Arom	23,6535	1,61433	334
C nr 9 i olef	0,0019	0,01455	334	C nr 9 Arom	16,5595	1,64405	334
C nr 10 i olef	0,0002	0,00328	334	C nr 10 Arom	2,7854	0,88023	334
C nr 4 n olef	0,0000	0,00000	334				

O melhor ajuste corresponde ao modelo M_{20} , com $r=0,900$, $R^2=0,811$ e $R_a^2=0,802$, Tabela 44. Este modelo escolheu como regressores 14 das 45 variáveis inicialmente consideradas. O número de regressores e os próprios regressores são diferentes dos usados nos modelos anteriores, também com 47 variáveis, Tabela 45.

O modelo é significativo em termos globais, logo rejeita-se a nulidade do(s) coeficiente(s) do parâmetro da(s) variável(s) independente(s) considerada(s) para cada modelo. E concluí-se que pelo menos uma das variáveis independentes tem um efeito significativo sobre a variação da variável dependente RON [31].

De acordo com o *teste t*, Tabela 45, todas as variáveis independentes têm poder explicativo do RON, uma vez que os coeficientes das variáveis independentes são diferentes de zero, para o modelo M_{20} [31].

Tabela 44- Características dos modelos com 47 variáveis (2003 a 2005).

Modelo	r	R^2	R^2 Ajustado
M_1	0,622	0,387	0,385
M_2	0,731	0,534	0,531
M_3	0,778	0,606	0,602
M_4	0,834	0,696	0,692
M_5	0,848	0,719	0,715

M ₆	0,866	0,750	0,745
M ₇	0,873	0,762	0,757
M ₈	0,880	0,774	0,769
M ₉	0,883	0,779	0,773
M ₁₀	0,885	0,783	0,776
M ₁₁	0,888	0,788	0,781
M ₁₂	0,892	0,795	0,787
M ₁₃	0,891	0,794	0,787
M ₁₄	0,894	0,800	0,792
M ₁₅	0,897	0,804	0,796
M ₁₆	0,898	0,807	0,798
M ₁₇	0,900	0,809	0,800
M ₁₈	0,899	0,809	0,800
M ₁₉	0,899	0,808	0,800
M ₂₀	0,900	0,811	0,802

Tabela 45- Coeficientes do modelo com melhor ajuste M₂₀, com 47 variáveis (2003 a 2005).

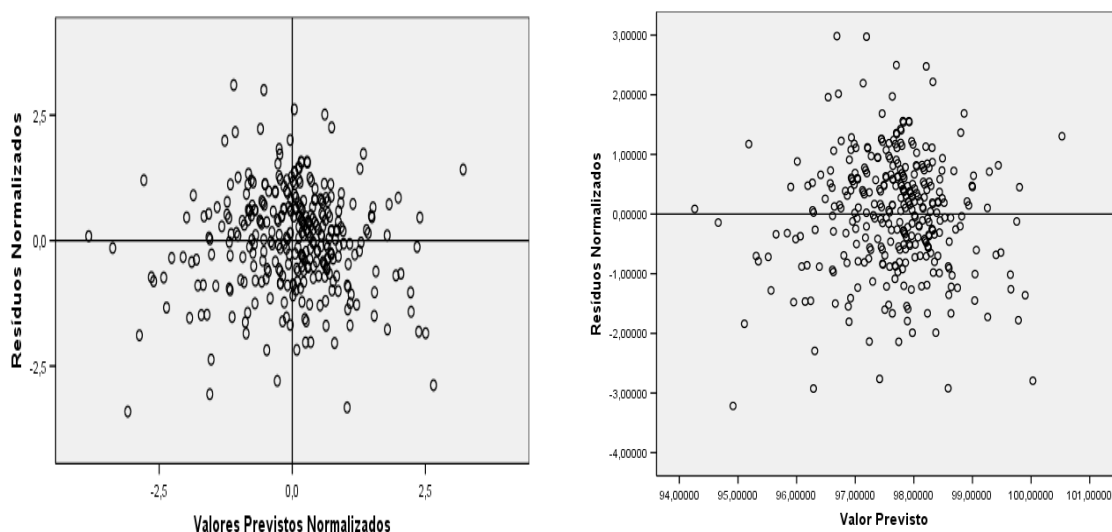
Modelo	Coeficientes		t	p-value	Estatística da Colinearidade		
	B	Erro Padrão			Tolerância	VIF	
M ₂₀	Constante	93,740	1,607	58,324	0,000		
	C nr 6 Naft	-0,600	0,201	-2,987	0,003	0,258	3,881
	C nr 7 n par	-1,010	0,179	-5,647	0,000	0,194	5,152
	C nr 6 n par	-0,248	0,086	-2,863	0,004	0,213	4,697
	C nr 11 npar	-5,040	1,276	-3,950	0,000	0,491	2,036
	C nr 9 i par	-0,748	0,207	-3,609	0,000	0,158	6,349
	C nr 5 n par	-0,494	0,102	-4,860	0,000	0,440	2,271
	C nr 5 i olef	2,915	0,690	4,225	0,000	0,774	1,292
	C nr 8 Arom	0,135	0,021	6,307	0,000	0,485	2,063
	C nr 9 Arom	0,217	0,030	7,134	0,000	0,232	4,314
	C nr 8 i par	-0,342	0,118	-2,894	0,004	0,099	10,138
	C nr 6 n olef	-5,101	1,639	-3,112	0,002	0,314	3,187
	C nr 7 i olef	1,218	0,476	2,558	0,011	0,394	2,538
	C nr 7 Arom	0,122	0,029	4,221	0,000	0,235	4,256
	C nr 8 Naft	1,959	0,978	2,002	0,046	0,239	4,186

$$\begin{aligned}
 \text{RON} = & 106,119 - 1,071 * i\text{-Nafténicos} C_6 - 0,247 * i\text{-Parafinas} C_7 - 0,517 * n\text{-} \\
 & \text{Parafinas} C_5 - 0,829 * n\text{-Parafinas} C_7 - 0,930 * n\text{-Parafinas} C_6 - 7,897 * n\text{-} \\
 & \text{Parafinas} C_{10} + 0,056 * \text{Aromáticos} C_9 - 2,755 * i\text{-Olefinas} C_8 - 0,279 * n\text{-} \\
 & \text{Parafinas} C_4 - 1,088 * n\text{-Parafinas} C_8 - 19,420 * n\text{-Parafinas} C_{11}
 \end{aligned}$$

7.8.3 Diagnóstico do Modelo Estimado

7.8.3.1 Homocedasticidade

De acordo com a distribuição dos resíduos, [Figura 37](#), não se rejeita a hipótese da homocedasticidade.



[Figura 37](#)– Avaliação da Homocedasticidade para o modelo M_{20} , com 47 variáveis (2003 a 2005).

7.8.3.2 Covariância Nula

De acordo com o teste de Durbin-Watson o modelo M_{20} apresenta um valor de teste $d=1,523$, logo existe auto-correlação entre os resíduos.

7.8.3.3 Normalidade dos Resíduos

De acordo com o teste de Kolmogorov-Smirnov, não se rejeita a hipótese dos resíduos seguirem uma distribuição normal, [Tabela 46](#). Esta hipótese é confirmada graficamente uma vez que as observações se dispõem ao longo das rectas oblíqua e horizontal, [Figura 38](#).

[Tabela 46](#)- Teste à Normalidade dos resíduos, modelo M_{20} com 47 variáveis (2003 a 2005).

	Kolmogorov-Smirnov		
	Estadística	Graus Liberdade	Significância
Resíduos Padrão	0,040	334	0,200

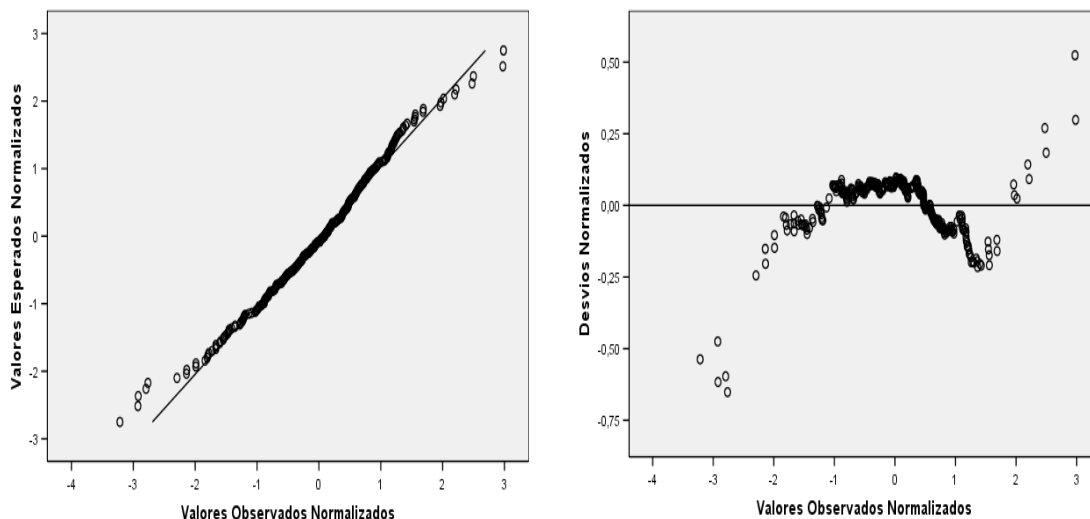


Figura 38- Análise gráfica da normalidade dos resíduos, para o modelo M₂₀, com 47 variáveis (2003 a 2005).

7.8.3.4 Multicolinearidade

Não se verifica a existência de multicolinearidade entre as variáveis independentes. Esta análise é feita com base na matriz das correlações e nos valores da tolerância e VIF, [Tabela 46](#).

7.8.4 Previsão do RON usando o modelo

Com base no modelo desenvolvido é possível calcular o valor de RON, a partir da composição cromatográfica, para as amostras de 2006 não usadas na sua elaboração. Os valores obtidos são comparados com os obtidos pelo CFR através de uma regressão linear, [Figura 39](#), e pelo teste *t de Student* das médias, [Tabela 47](#).

Da comparação dos resultados por regressão linear, verifica-se que os resultados para o RON obtidos por este modelo são mais próximos dos obtidos pelo método do motor CFR, do que os do modelo anterior. O coeficiente de correlação é superior e a recta obtida aproxima-se mais da relação ideal:

$$RON_{\text{modelo}} = 29,69 + 0,69RON_{\text{CFR}} \text{ com } R^2 = 0,85.$$

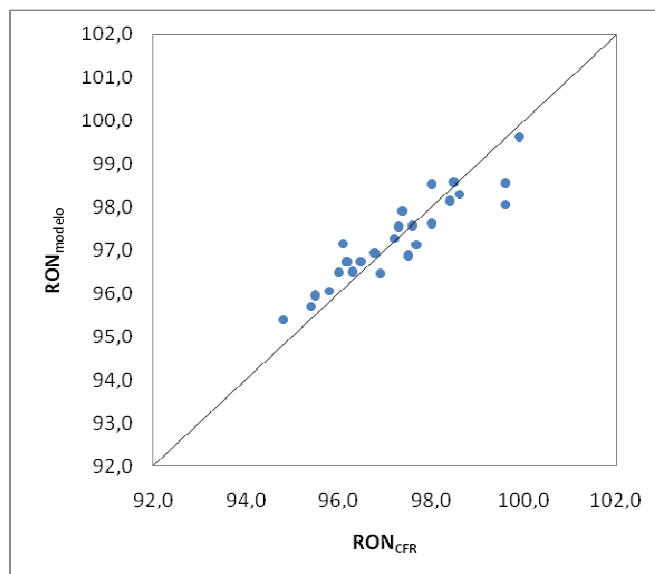


Figura 39- Representação gráfica do valor de RON_{CFR} vs RON_{modelo} com 47 variáveis (2003 a 2005).

Tabela 47- Resultados da comparação do RON_{CFR} e RON_{modelo} com 47 variáveis (2003 a 2005).

	RON_{CFR}	RON_{modelo}
Média	97,2	97,3
Desvio Padrão	1,35	1,02
Variância	1,83	1,04
n	26	26
F _{calc}		1,76
F _{tab}		1,96
$F_{calc} < F_{tab}$		
t _{calc}		0,09
t _{tab}		2,01
$t_{calc} < t_{tab}$		

Do teste *t* concluí-se que não existem desvios significativos entre os resultados obtidos por um ou outro método, Tabela 47.

Relativamente às diferenças obtidas entre os dois métodos estas também diminuíram e a percentagem de resultados cuja diferença é superior a $|0,7|$ diminuiu drasticamente de 43% para 12% dos resultados, Figura 40.

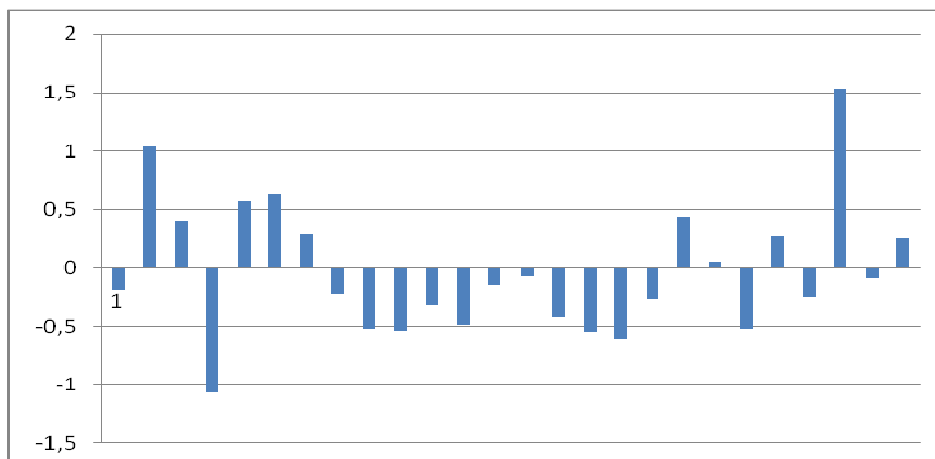


Figura 40- Diferença entre os resultados obtidos com o RON_{CFR} e RON_{modelo} com 47 variáveis 2003 a 2005.

8. DISCUSSÃO E CONCLUSÃO

O Número de Octano (RON), à semelhança de outras propriedades físicas e químicas das gasolinas e dos componentes de gasolina, está directamente relacionado com a sua composição [9].

Esta relação é confirmada pelos inúmeros estudos realizados com o objectivo de desenvolver métodos alternativos, na maioria dos casos modelos matemáticos, para a determinação do RON a partir da composição da amostra. De entre os vários métodos alternativos publicados salienta-se que os melhores resultados foram sempre obtidos para aqueles cuja composição foi determinada por cromatografia gasosa (GC) [9]. No entanto apesar desta técnica analítica permitir uma maior discriminação da composição da amostra, esta vantagem nem sempre foi usada uma vez que na maior parte dos modelos desenvolvidos a composição é agrupada, tendo em conta as estruturas dos vários compostos. Estes métodos para além da composição também usam os valores do RON dos compostos puros que constituem a mistura. Estes por sua vez são determinados em condições muito específicas, no entanto quando presentes numa mistura, com outros componentes, apresentam um comportamento bastante diferente [23]. Verifica-se ainda que cada um dos diversos métodos foi sempre desenvolvido para uma matriz específica sendo difícil aplicá-los a outras matrizes. Assim os vários métodos alternativos publicados apresentam um carácter muito específico no que diz respeito à sua aplicação, ao contrário do método de referência ASTM D 2699 que é universal.

A dificuldade de aplicar estes métodos a matrizes diferentes, daquelas para os quais foram desenvolvidos, foi sentida quando se tentaram aplicar alguns dos métodos às amostras de Reformado, matriz objecto de estudo nesta dissertação. Assim surgiu a necessidade de desenvolver um método para a determinação do RON, a partir da composição, direccionado para o Reformado. O método desenvolvido é também muito específico, uma vez que foi desenvolvido para o Reformado e o seu comportamento é ainda desconhecido para outras matrizes.

Inicialmente foi desenvolvida uma série de 3 modelos em que se aumentou gradualmente a discriminação da composição da amostra (4, 7 e 47 variáveis

independentes) no sentido de confirmar se o aumento da discriminação da composição da amostra melhora a capacidade de previsão do RON por parte dos modelos.

Confirma-se que com o aumento da discriminação da composição da amostra o R_a^2 do modelo melhora, passa de 0,704 do primeiro modelo (com 4 variáveis) para 0,866 no terceiro modelo (com 47 variáveis).

A percentagem de amostras em que a diferença entre os valores do RON determinado pelo motor CFR e pelo modelo é superior a $|0,7|$ (reprodutibilidade do método) passa de 22% no primeiro modelo (com 4 variáveis) para apenas 5% no terceiro modelo (com 47 variáveis).

Perante estes resultados conclui-se que o aumento da discriminação da composição da amostra melhora a capacidade de previsão do modelo e por isso os modelos seguintes foram desenvolvidos com as 47 variáveis.

Da análise de resíduos foram encontradas algumas amostras com valores atípicos. Estas observações extremas que apresentam resíduos superiores aos resíduos das restantes amostras [29], podem ser consequência de erros analíticos, de erros de transcrição de resultados, ou mesmo dias de produção anómalos que provocaram alterações na composição das amostras. No entanto como a esta distância não é possível rastrear qual o verdadeiro motivo destes valores, estas amostras não foram consideradas para a elaboração dos modelos.

Uma vez que é sempre indicado validar o modelo com outro conjunto de dados [29], os dados globais foram divididos em 3 grupos, que continham as amostras referentes aos períodos: 2003, 2003 a 2004 e 2003 a 2005.

Assim foi desenvolvido um primeiro modelo com os dados de 2003 e validado com os restantes dados 2004 a 2006, um segundo com os dados de 2003 a 2004 e validado com os dados de 2005 a 2006, e um terceiro modelo com os dados de 2003 a 2005 e validado com os dados de 2006. Os pressupostos do modelo respeitantes aos erros ou resíduos e à quasi(ortogonalidade) entre as variáveis independentes estão validados para os vários modelos [29].

Apesar de disponíveis as 47 variáveis, em nenhum dos modelos foram usadas todas as variáveis disponíveis, o número máximo de variáveis usado foi de 14.

A análise do coeficiente de determinação ajustado e a diferença entre o valor obtido para o RON pelo motor CFR e pelo modelo foram os critérios de qualidade do modelo. Assim verifica-se que o coeficiente de determinação é elevado nos três modelos no entanto diminui com o aumento do número de amostras usadas na sua elaboração: 0,942 no primeiro modelo (2003) e 0,802 no terceiro modelo (2003 a 2005).

Embora o coeficiente de determinação diminua, a capacidade de previsão do RON melhora bastante com a inclusão de mais amostras, uma vez que a diferença entre os valores obtidos pelo motor CFR e pelo modelo passa de 43% no primeiro modelo (2003) para 12% no terceiro modelo (2003 a 2005). No entanto de acordo com o definido no método de referência ASTM D 2699 a reprodutibilidade só deve ser ultrapassada 1 vez em cada 20 ensaios, isto é em 5% dos resultados, o que não acontece neste caso [16]. Este critério tem que ser cumprido para que o modelo possa ser usado como um método alternativo ao método de referência ASTM D 2699. Apesar disto, e de acordo com o teste *t* para as médias, concluímos que em todos os modelos os resultados obtidos por um e outro método não são estatisticamente diferentes, podem ser considerados como pertencentes à mesma população.

Deste modo, é de todo vantajoso num futuro trabalho prosseguir com a inclusão das amostras de 2006 no modelo e a validação do mesmo com as amostras de 2007, no sentido de confirmar a melhoria da capacidade de previsão do modelo. E também com o objectivo de alcançar diferenças de resultados para amostras analisadas pelos dois métodos inferiores à reprodutibilidade do método [0,7].

Nessa nova fase e uma vez que ainda se desconhece qual ou quais as variáveis que conduzem ao melhor modelo, o procedimento aconselhável é novamente a utilização dos três métodos (*Forward*, *Backward* e *Stepwise*) e identificação de quais as variáveis que são deixadas de fora pelos três métodos em simultâneo. Estas variáveis podem então ser eliminadas da análise e o modelo pode ser aplicado às outras variáveis [29].

Outro ponto que pode ser explorado é variar o nível de significância (α) usado como critério de saída e entrada das variáveis no modelo. Neste trabalho foram usados os valores por defeito do SPSS $\alpha_{\text{entrada}}=0,05$ e $\alpha_{\text{saída}}=0,10$. Note-se que a selecção final depende dos valores de α para a selecção de variáveis

e que diferentes valores de α podem conduzir a modelos completamente diferentes [29].

No caso de serem obtidos bons resultados seria interessante testar a aplicação do modelo desenvolvido noutras matrizes, como outros *Componentes de Gasolina*.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] – J. G. Pulido, P. Fonseca, *O Petróleo e Portugal - O Mundo do Petróleo e o seu impacto no nosso país*, Tribuna, 2004.
- [2] - www.galpennergia.com
- [3] - Soares Mota, *Refinação de Petróleos*, apresentação na Galp Energia em Lisboa a 11 de Março de 2008;
- [4] - www.setlaboratories.com
- [5] - www.intertek-cb.com
- [6] - Concawe- *The Oil Companies European Organization for Environmental and Health Protection*, Product. Dossier no. 92/103 gasolines, 1992.
- [7] – www.faqs.org, Bruce Hamilton, *Automotive Gasoline*.
- [8] - N. Pasadakis, V. Gaganis, C. Foteinopoulos, *Octane number prediction for gasoline blends*, Fuel Processing Technology 87 (2006) 505-509.
- [9] – N. Nikolaou, C. E. Papadopoulos, I. A. Gaglias, K. G. Pitarakis, *A new-linear calculation method of isomerisation gasoline research octane number based on gas chromatographic data*, Fuel 83 (2004) 517-523.
- [10] – R. L. Furey, A. M. Horowitz, N. J. Schroeder, *Manual on significance of Tests for Petroleum Products*, Automotive Gasoline, 6ª edição, 1993.
- [11] - Directiva Europeia 2003/17/CE, transposta pelo DL 235/2004.
- [12] – *Fabricar uma Gasolina Moderna*, Shell Ciência & Tecnologia, edição especial, Março de 1987.
- [13] – J. P. Wauquier, *Petroleum Refining – Crude Oil, Petroleum Products, Process Flowsheets*, Institut Français du Pétrole Publications, Éditions Technip, 1995.
- [14] – www.asifunciona.com/mecanica
- [15] – www.bp.com
- [16] - ASTM D 2699-06a, IP 237/87, *Standard Test Method for Research Octane Number of Spark-Ignition Engine Fuel*, 2006.
- [17]- ASTM D 2700-06. IP 236/87, *Standard Test Method for Motor Octane Number of Spark-Ignition Engine Fuel*, 2006.
- [18]- W. F. Schoen, A. V. Mrstik, *Calculating gasoline blend octane ratings*, Ind. and Engr. Chem., 47(9), 1955, pp 1740-1742.
- [19]- W. E. Stewart, *Predict octanes for gasoline blends*, Petroleum Refiner, 38(12), 1959, pp. 135-139.

- [20]- W.C. Healy, C. W. Jr Maassen, R. T. Peterson, *A new approach to blending octanes*, API Division of Refining, 24th midyear meeting, New York, May 27, 1959.
- [21]- M. H. T. Auckland, D. J. Charnock, *The development of linear blending indices for petroleum properties*, J. of the Inst. Of Petroleum, 55(545), 1969, pp. 322-329.
- [22]- P. C. Anderson, J. M. Sharkey, R. P. Walsh, *Calculation of Research Octane Number of Motor Gasolines from Gas Chromatographic Data and New Approach to Motor Gasoline Quality Control*, Journal of the Institute of Petroleum, Volume 58, N.º 560, Março 1972.
- [23]- *Knocking Characteristics of Pure Hydrocarbons*, API Research Project 45. American Society for Testing Materials, 1958.
- [24]- W. E. Morris, *Interaction approach to gasoline blending*, NPRA Paper AM-75-30, NPRA annual meeting, March 1975.
- [25]- M. H. Rusin, H. S. Chung, J. F. Marshall, *A transformation method for calculating the research and motor octane numbers of gasoline blends*, Ind. Eng. Chem. Fundam., 20(3), 1981, pp. 195-204.
- [26]- A. H. Zahed, S. A. Mullah, M. D. Bashir, *Predict octane number for gasoline bends*, Hydrocarbon Processing, May 1993, pp. 85-87.
- [27] – Luiz Alberto da Silva, Vera Maria Moura Torres da Costa, *Correlações de Octanagem por Cromatografia Gasosa e Análise PIANIO*, Petrobras.
- [28] - Manual do equipamento HP 5890- Tecnologia AC- Analisador PIONA;
- [29] - João Maroco, *Análise Estatística – com utilização do SPSS*, Edições Sílabo, 3^a edição, 2007.
- [30] – R. C. Guimarães, J. A. Sarsfield Cabral, *Estatística*, McGraw-Hill Lda, 1997.
- [31] – M. H. Pestana, J. N. Gageiro, *Análise de Dados para Ciências Sociais- A Complementaridade do SPSS*, Lisboa, Edições Silabo, 2003.
- [32] – James N. Miller & Jane C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Pearson Education Limited, 4th edition, 2000.