



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE
PETRÓLEO

PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE
GEOENGENHARIA DE RESERVATÓRIOS

WAGNER BARBOSA DE MELLO CASTRO

CORRELAÇÃO ROCHA-PERFIL ATRAVÉS DE REGRESSÃO LOGÍSTICA

Dissertação apresentada ao Instituto de Geociências como parte dos
requisitos para obtenção do título de Mestre em "Geociências - Área
de Geoengenharia de Reservatórios"

Orientador: Professora Doutora Regina Célia de Carvalho Pinto Moran

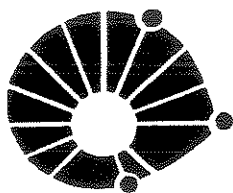
Este exemplar corresponde
redação final da dissertação
por Wagner Barbosa de Mello Castro
e aprovada por Regina Célia de Carvalho Pinto Moran
em 13/04/98
Regina Célia de Carvalho Pinto Moran
ORIENTADOR

CAMPINAS - SÃO PAULO

FEV - 1998

C279c

36271/BC



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

**PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE
GEOENGENHARIA DE RESERVATÓRIOS**

WAGNER BARBOSA DE MELLO CASTRO

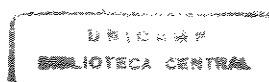
CORRELAÇÃO ROCHA-PERFIL ATRAVÉS DE REGRESSÃO LOGÍSTICA

Dissertação apresentada ao Instituto de Geociências
como parte dos requisitos para obtenção do título de
Mestre em "Geociências - Área de Geoengenharia de
Reservatórios"

Orientador: Professora Doutora Regina Célia de Carvalho Pinto Moran

CAMPINAS - SÃO PAULO

FEV - 1998



1903.309

UNIDADE	BC
N.º CHAMADA:	UNICAMP
V.	Er
TOMBO	60/36.271
PROG	229/99
C	☐
D	☒
PREÇO	R\$ 11,00
DATA	23/01/99
N.º CPD	

CM-00120369-8

FICHA CATALOGRÁFICA ELABORADA
PELA BIBLIOTECA I.G. UNICAMP

Castro, Wagner Barbosa de Mello
C279c Correlação rocha-perfil através de regressão logística / Wagner
Barbosa de Mello Castro.- Campinas,SP.: [s.n.], 1998.

Orientador: Regina Célia de Carvalho Pinto Moran
Dissertação (mestrado) Universidade Estadual de Campinas. Instituto de
Geociências

1. Petróleo - Geologia. 2. Estatística. I. Moran, Regina Célia de
Carvalho Pinto. II. Universidade Estadual de Campinas. Instituto de
Geociências. III. Título



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

**PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE
GEOENGENHARIA DE RESERVATÓRIOS**

AUTOR: Wagner Barbosa de Mello Castro

TÍTULO DA DISSERTAÇÃO/TESE: Correlação Rocha-Perfil Através de Regressão Logística

ORIENTADOR: Prof. Dra. Regina Célia de Carvalho Pinto Moran

Aprovada em: 17 / 4 / 98

PRESIDENTE: Profa. Dra. Regina Célia de Carvalho Pinto Moran

EXAMINADORES:

Prof. Dr. Regina Célia de Carvalho Pinto Moran

- Presidente

Prof. Dr. Chang Hung Kiang

Prof. Dr. Saul Barisnik Suslick

Campinas, 17 de abril de 98

Agradecimentos

À Professora Regina pela disponibilidade, motivação e oportuna orientação.

À Petrobras pelos recursos e pela liberação sem os quais não seria possível a conclusão deste trabalho.

Aos colegas Jorge A.D. Sarturi, Miguel A.B. de Castro, Ricardo Ruy Neto e Nelson P.F. Filho pelos comentários, sugestões e informações relevantes.

À Maria José e às pequenas Theresa e Cora pela nem sempre fácil compreensão e infinita paciência demonstradas pelas minhas frequentes ausências do convívio familiar.

Sumário

1. Introdução	1
2. Correlação Rocha-Perfil	5
2.1 Introdução	5
2.2 Tipos de Correlação	6
2.3 Natureza dos Dados de Perfis	8
2.4 Técnicas Multivariadas aplicadas à correlação rocha-perfil	9
3. Desenvolvimento do Modelo Teórico	11
3.1 O Modelo Logístico	11
3.2 Ajuste do Modelo	15
3.3 Avaliação do Modelo	17
3.4 Qualidade do ajuste/diagnóstico do modelo	18
4. Regressão Logística usando a Proc Logistic	20
4.1 Introdução	20
4.2 Descrição dos dados	21
4.2.1 Contexto geológico	23
4.2.2 Descrição das litofácies	23
4.2.2.1 Conglomerado arenoso (Ca)	23
4.2.2.2 Arenito conglomerático (Ac)	24
4.2.2.3 Arenito maciço (Am)	24
4.2.2.4 Interlaminação (I)	24
4.2.2.5 Arenito estratificado (Ae)	24
4.2.2.6 Arenito argiloso deformado (D)	25
4.2.3 Relacionamento entre as litofácies e respostas em perfis elétricos	25
4.2.4 Curvas de fácies e perfis elétricos utilizados	26
4.3 Análise exploratória dos dados de perfis	28
4.3.1 Poço P1	28
4.3.2 Poço P2	32
4.3.3 Comparação entre P1 e P2	35
4.4 Criação dos modelos com um conjunto fixo de variáveis	35
4.4.1 Criação do Modelo Global	38
4.5 Geração dos modelos com escolha de variáveis	42
4.6 Comentários sobre a avaliação dos modelos ajustados	45
4.7 Comparação com os resultados obtidos pela análise discriminante	46
5. Conclusões	48
Apêndice 1 - Listagens dos programas SAS	49
Apêndice 2 - Sidas dos programas SAS	56
Apêndice 3 - Análises sequenciais de testemunhos dos poços P1 e P2	88
Apêndice 4 - Gráficos e estatísticas descritivas das variáveis de perfis.	89
Referências Bibliográficas	103

Lista de Tabelas

<i>Tabela 4-1 Descrição das litofácies segundo Becker et alli (1989).</i>	26
<i>Tabela 4-2 Estatísticas descritivas para o poço P1 (geral e por fácies).</i>	29
<i>Tabela 4-3 Estatísticas descritivas para o poço P2 (geral e por fácies).</i>	33
<i>Tabela 4-4 Estatísticas descritivas para o poço Global (geral e por fácies)</i>	39
<i>Tabela 4-5 Modelos logísticos obtidos a partir do mesmo conjunto de variáveis da análise discriminante.</i>	39
<i>Tabela 4-6 Modelos logísticos obtidos a partir do método stepwise de seleção de variáveis.</i>	43
<i>Tabela 4-7 Tabela de classificação dos resultados obtidos pelos modelos logísticos construídos com base nas variáveis utilizadas pela análise discriminante, comparados aos valores da fácies-padrão (Realfac).</i>	44
<i>Tabela 4-8 Tabela de classificação dos resultados obtidos pelos modelos logísticos com seleção de variáveis comparados aos valores da fácies-padrão (Realfac).</i>	45
<i>Tabela 4-9 Tabela de classificação dos resultados obtidos pela análise discriminante (efac) e pela regressão logística (facies).</i>	47
<i>Tabela 4-10 Tabela de classificação dos resultados obtidos pela análise discriminante (efac) e pela regressão logística com escolha de variáveis (facies).</i>	47

Lista de Figuras

<i>Figura 3-1 Gráfico de $\text{logito}[\pi(x)]$ versus $\pi(x)$</i>	<i>14</i>
<i>Figura 4-1 - Mapa de Localização do Campo de Pargo.</i>	<i>22</i>
<i>Figura 4-2 Matriz de Correlação e gráficos de dispersão para o poço P1</i>	<i>31</i>
<i>Figura 4-3 Matriz de correlação e gráficos de dispersão para o poço P2.</i>	<i>34</i>
<i>Figura 4-4 Gráfico do perfil de raios-gama para o poço global com os dados originais e padronizados.</i>	<i>41</i>



UNIVERSIDADE ESTADUAL DE CAMPINAS/
INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE
PETRÓLEO

PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE
GEOENGENHARIA DE RESERVATÓRIOS

CORRELAÇÃO ROCHA-PERFIL ATRAVÉS DE REGRESSÃO LOGÍSTICA

RESUMO

DISSERTAÇÃO DE MESTRADO

Wagner Barbosa de Mello Castro

Com o objetivo principal de verificar-se a aplicabilidade da técnica na área de geoengenharia de reservatórios, foram gerados modelos logísticos de regressão a partir de litofácies descritas em testemunhos e dos perfis elétricos para dois poços da Bacia de Campos. Estes modelos foram posteriormente utilizados para se estimar a ocorrência de fácies reservatório nos poços.

Os resultados obtidos na estimativa de fácies foram então comparados aos resultados provenientes de uma análise discriminante prévia com o objetivo de averiguar a precisão das duas técnicas como ferramentas para se estimar fácies.

Embora a análise discriminante tenha se mostrado mais precisa na estimativa de fácies reservatório, não se deve descartar a utilização da regressão logística. A sua independência da pressuposição de normalidade torna esta técnica, ao menos em teoria, mais robusta do que a análise discriminante.



UNIVERSIDADE ESTADUAL DE CAMPINAS/
INSTITUTO DE GEOCIÊNCIAS/ÁREA DE GEOLOGIA DE
PETRÓLEO

PÓS-GRADUAÇÃO EM GEOCIÊNCIAS - ÁREA DE
GEOENGENHARIA DE RESERVATÓRIOS

CORRELAÇÃO ROCHA-PERFIL ATRAVÉS DE REGRESSÃO LOGÍSTICA

ABSTRACT

DISSERTAÇÃO DE MESTRADO

Wagner Barbosa de Mello Castro

Logistic regression models were generated starting from lithofacies described in cores and in well logs for two wells of Campos Basin. The main objective was verify the applicability of the technique in reservoir geology. The models were used to estimate the occurrence of reservoir facies in the wells.

Results obtained were compared to the results of a previous discriminant analysis with the objective of determinate the accuracy of the two techniques as tools to estimate reservoir facies.

Although discriminant analysis resulted more accurate in the estimate of reservoir facies, the use of logistic regression should not be discarded. Its independence of the normal distribution hypothesis make this technique, at least in theory, more robust than the discriminant analysis.

1. Introdução

Dentro do processo, na indústria do petróleo, das atividades de geoengenharia, ocupa a correlação rocha-perfil um papel muitas vezes injustamente relegado a um plano de importância secundária. É, entretanto, uma atividade cujos resultados podem influenciar fundamentalmente o resultado final dos modelos de simulação de fluxo, que, por sua vez, constituem a base dos projetos de investimento na exploração do campo. Chama-se atenção para o fato porque uma correlação rocha-perfil criteriosa, com modelos preditivos bem ajustados definirá, na verdade, a qualidade dos dados da informação geológica de detalhe que conseguimos obter do reservatório. Estas informações, conjugadas aos modelos geológicos de escala maior, cuja base por sua vez é definida pelo mapeamento sísmico, somadas às informações obtidas do histórico de produção do campo constituem o modelo global do reservatório.

O processo de correlação rocha-perfil envolve várias etapas, desde a descrição dos testemunhos, até o estudo dos perfis, passando pelas análises petrofísicas. A concatenação destas informações, com a geração de um modelo preditivo para as fácies descritas é onde situaremos o escopo do atual estudo. As técnicas estatísticas multivariadas vem sendo tradicionalmente utilizadas na construção deste modelo preditivo. Análises discriminantes, análises de agrupamentos e modelos lineares de regressão tem sido utilizadas com uma frequência cada vez maior pelo corpo técnico que atua na área, conforme veremos mais detalhadamente no próximo capítulo.

Muitas vezes o caráter dos dados disponíveis não permite a utilização adequada das técnicas multivariadas tradicionais devido às premissas básicas adotadas em seu desenvolvimento. Isto se evidencia de uma forma relativamente frequente na utilização dos modelos lineares de regressão. Estes modelos pressupõem que a variável-resposta seja normalmente distribuída, o que raramente se verifica na natureza. Geralmente adota-se algum tipo de transformação aos dados originais, algumas vezes com sucesso e outras tantas, não. É o caso, por exemplo, da lognormalidade assumida frequentemente como sendo a distribuição característica da permeabilidade em reservatórios. Deve-se notar, ainda, que dependendo da transformação, existe neste procedimento um problema adicional, pois o retorno aos dados

originais pode implicar em perda de informações. Jensen (1986) sugere um interessante procedimento a este respeito, situando a distribuição de permeabilidades de um determinado conjunto de dados em um ponto qualquer de uma família de distribuições cujos extremos são a distribuição normal e a lognormal.

As principais motivações para o desenvolvimento deste estudo, foram, primeiramente, a possibilidade de utilizar-se pioneiramente uma técnica estatística na solução de um problema específico de correlação rocha-perfil e, em segundo lugar, mas sem ordem relativa de importância, por acreditarmos, com base nas premissas assumidas no desenvolvimento da técnica e nas características dos dados analisados, ser a regressão logística um método teoricamente bem adequado à solução do problema específico que será enfocado.

Por conta disto, nos deteremos no caso mais geral de estimativa de fácies a partir dos perfis : estimar uma fácies binária, que é na realidade, uma variável indicatriz de fácies. Ou seja, ela pode assumir os valores 1, caso seja verificada a ocorrência de fácies reservatório e 0, caso contrário. A estimativa se dará a partir de um conjunto de curvas de perfis. Em outras palavras, faremos uma análise de regressão logística múltipla a partir de um conjunto de variáveis explanatórias contínuas, sem mistura de variáveis.

A utilização de modelos logísticos na indústria de petróleo é rara, ficando restrita a eventuais utilizações na área exploratória. Uma interessante aplicação é encontrada em Silva (1983), que, em uma aplicação do método denominado *creaming* (Meisner & Demirmen, 1981), ajustou o índice de sucesso de uma campanha exploratória através de um modelo logístico, com o objetivo de se estimar o potencial petrolífero de uma bacia.

Os modelos logísticos tem sido, principalmente nas últimas duas décadas, intensivamente utilizados em bioestatística. Em particular, a epidemiologia, em função do caráter dos seus dados e dos objetivos dos estudos, utiliza a regressão logística extensivamente, transformando-a em ferramenta padrão para a realização de determinados tipos de estudos (Barros, 1990). Digno de nota, ainda, é o caráter fortemente quantitativo da epidemiologia. Seus conceitos básicos e medidas de frequência de incidência de doenças são objetiva e claramente definidos, com base matemática e estatística. Os tipos de estudos mais frequentemente realizados, como os de caráter prospectivo e retrospectivo, por exemplo, encontram-se já bem estabelecidos, cada qual com sua própria aplicação do modelo logístico.

em função de sua particular estrutura de dados. Como estes estudos baseiam-se, em geral, em populações com incidência de algum tipo de doença, e os conjuntos de dados são construídos com base em observações diretas e/ou indiretas de características destas populações, é muito comum que o pesquisador desta área manipule com frequência conjuntos de dados formados por uma mistura de variáveis contínuas e discretas. Deve-se atentar para o fato, porque, em função dos tipos de variáveis explanatórias presentes, o modelo logístico pode apresentar-se de maneiras algo diferentes de um caso para outro. Esta bem definida estruturação estatística da epidemiologia contribuiu para a grande evolução verificada nos modelos logísticos em particular. Uma rápida verificação na bibliografia existente sobre o assunto mostra-nos de imediato que existe uma forte associação entre as duas ciências, sendo muito difícil determinar onde termina a epidemiologia e inicia a estatística.

Estas considerações tornaram-se necessárias para ilustrar o quão incipiente ainda está, como ciência formal, a área tecnológica de correlação rocha-perfil. Naturalmente, não se exige, em função de seus objetivos o mesmo nível de detalhamento e precisão, por exemplo, que um estudo epidemiológico às vezes precisa mostrar. Mas nota-se que faltam conceitos básicos definidos de maneira algo mais rígida, preferencialmente estatisticamente embasados, como veremos melhor no próximo capítulo. Nos últimos anos têm-se observado um crescente interesse na utilização de técnicas estatísticas multivariadas nos estudos de correlação rocha-perfil. O interesse maior continua recaindo sobre as técnicas tradicionais que são, sem dúvida, ótimas alternativas quando bem aplicadas. Espera-se que este estudo contribua para aumentar o leque de opções à disposição do intérprete, com a utilização de uma técnica consagrada em outras áreas, acreditando que ela possa ser bem aplicada na área de correlação rocha-perfil, com resultados satisfatórios.

Não pretende este estudo se transformar em referência obrigatória em nenhuma das duas áreas por onde transita. O enfoque adotado será essencialmente prático. Nosso objetivo será, como já assinalado, verificar a aplicabilidade da técnica a um problema específico de correlação rocha-perfil. Desta forma, leitores interessados em aprofundar-se tanto nos modelos logísticos quanto nos problemas de correlação rocha-perfil devem dirigir-se à bibliografia indicada e referenciada no texto. Os conceitos discutidos serão direcionados à solução do nosso problema específico de estimativa de fácies.

Neste sentido, o capítulo 2 discutirá de maneira geral o problema de correlação rocha-perfil, definirá alguns conceitos e comentará alguns exemplos de aplicação existentes na literatura. O capítulo 3 tratará do desenvolvimento teórico do modelo logístico, tendo em vista nosso problema específico. O capítulo 4 fornecerá um exemplo de aplicação da regressão logística a um conjunto real de dados provenientes de dois poços da Bacia de Campos, utilizando a `proc Logistic` do SAS. Seus principais recursos serão comentados no decorrer da análise. As estimativas de fácies obtidas com a regressão logística serão comparadas a estimativas oriundas de uma análise discriminante prévia, com o intuito de averiguar-se qual das técnicas fornece melhor capacidade preditiva para este conjunto de dados. Finalmente, no capítulo 5 comentaremos os resultados alcançados e analisaremos a viabilidade de aplicação da técnica ao problema proposto.

2. Correlação Rocha-Perfil

2.1 Introdução

Chamamos de correlação rocha-perfil a todo e qualquer estudo, dentro do processo da geoengenharia, que visa obter, a partir dos perfis elétricos, estimativas dos parâmetros que caracterizam uma rocha ou um determinado grupo de rochas. Estes parâmetros, ou atributos, normalmente referem-se às propriedades petrofísicas de maior interesse imediato no estudo de reservatórios, a porosidade e a permeabilidade, mas não estão restritos a eles. Na verdade, qualquer característica de um pacote rochoso que consiga-se verificar em perfis pode ser objeto de uma correlação rocha-perfil. Desta forma, a partir de dados de rocha que expressem alguma característica de interesse ao estudo, procura-se criar modelos quantitativos que expressem da melhor maneira possível o relacionamento entre estas características e suas respectivas respostas em perfis. Estes modelos serão utilizados para se fazer estimativas destas características em poços onde não se dispõe de dados diretos das rochas. Existe, portanto, uma forte justificativa econômica e operacional para a realização da correlação rocha-perfil. Os dados de rochas são provenientes quase sempre de testemunhos e são de obtenção consideravelmente mais cara e demorada que os dados de perfis elétricos. Por este motivo, nos estudos de reservatório, os poços testemunhados são considerados os poços-chave, pois é a partir deles que são obtidas as informações diretas das rochas, possibilitando a geração dos modelos quantitativos procurados.

O tema correlação rocha-perfil não se encontra, com relação à sua estruturação formal como área de conhecimento tecnológico, ainda bem estabelecido internacionalmente. Não se encontra, na literatura, correspondente em inglês para a expressão que estamos livremente utilizando. Os estudos de correlação rocha-perfil são costumeiramente relacionados ao que se convencionou chamar genericamente de aplicações geológicas de perfis. Isto pode incorrer em algum tipo de confusão conceitual, pois na área de reservatórios a correlação rocha-perfil tem objetivos bem claros e definidos. Em vista disto, Sarzenski & Toledo (1990) em trabalho já clássico no âmbito da indústria de petróleo no Brasil, fazem um extensivo levantamento

histórico de aplicações de técnicas de correlação rocha-perfil. Além disto, definem conceitos e propõem um modelo de classificação para os diversos tipos de correlação rocha-perfil identificados. A menos que explicitamente referenciado de forma contrária, adotaremos a classificação proposta por estes autores.

2.2 Tipos de Correlação

De acordo com a origem dos dados podem ser identificados três tipos de correlação :

1. **Correlação perfil-perfil** : quando são utilizados apenas dados de perfis, geralmente associados a feições e padrões bem definidos identificados nas curvas. Pode ser utilizada quando não se dispõe de dados das rochas ou quando o objetivo da correlação faz com que possam ser dispensados. Por exemplo, rastrear um determinado marco estratigráfico ou delimitar a geometria externa de um reservatório;
2. **Correlação rocha-rocha** : quando são utilizados exclusivamente dados de rocha ou informações derivadas das mesmas. Por exemplo, um estudo de porosidades, onde comparam-se porosidades obtidas a partir de análises petrofísicas com as observadas em lâminas delgadas das mesmas rochas;
3. **Correlação rocha-perfil** : é a mais rica em informações, pois utiliza dados dos dois tipos anteriores de correlação. É, também, com relação aos objetivos dos estudos de reservatório, a mais importante. Trata-se basicamente de comparar curvas de perfis a características observadas em um determinado tipo ou pacote rochoso. A este conjunto de características que definem um tipo litológico, diferenciando-o dos demais, dá-se o nome de fácies. Segundo o Novo Dicionário da Língua Portuguesa (Ferreira, 1986), o termo fácies exprime: “1. O aspecto de um corpo, tal como se apresenta à primeira vista; 2. Os caracteres de forma e configuração que distinguem um grupo; aspecto, em geral; e, 6. (Geol.) conjunto dos caracteres de uma rocha, considerados sob o aspecto de sua

formação.” Ou seja, fácies é o caráter distintivo de uma rocha ou grupo de rochas afins. Neste contexto, iremos trabalhar, neste estudo, com rochas cujas características definem uma fácies de rocha-reservatório. Esta fácies reservatório será identificada a partir das correspondentes assinaturas verificadas nos perfis elétricos. A este conjunto de assinaturas características chamamos eletrofácies. Este termo, originalmente empregado por Serra & Abbott (1980) é análogo ao conceito de fácies geológica e visa caracterizar em leituras de perfis, determinadas zonas através de suas respostas e possibilitar a diferenciação entre estas e outras zonas igualmente bem caracterizadas. Pode-se esperar, portanto, para cada ocorrência de fácies-reservatório uma respectiva eletrofácies.

Continuando com a classificação proposta por Sarzenski & Toledo (op. cit.), a correlação rocha-perfil, por sua vez, em função de seus objetivos e da natureza dos dados de rocha utilizados, pode ser subdividida em 4 tipos :

1. **Fácies litológica de correlação rocha-perfil** : quando objetiva-se obter modelos de correlação rocha-perfil que possibilitem a extrapolação de características litológicas, as litofácies, a partir das correspondentes eletrofácies. Por exemplo, a identificação de folhelhos capeadores em poços não testemunhados;
2. **Fácies reservatório de correlação rocha-perfil** : quando além, obviamente, da litologia, são consideradas também informações relativas à sua qualidade como reservatório, isto é, sua capacidade de armazenar e produzir hidrocarbonetos. Por exemplo, contrastando com o exemplo anterior, podemos ter um folhelho fraturado, representante de uma fácies-reservatório;
3. **Fácies diagenética de correlação rocha-perfil** : quando o estudo é direcionado a caracterizar fácies a partir de características derivadas de sua história diagenética. Por exemplo, rastrear níveis de arenito cimentado dentro de um reservatório;
4. **Fácies petrofísica de correlação rocha-perfil** : são de maneira geral as mais importantes ao estudo de detalhe de um reservatório. São aquelas onde propriedades importantes como porosidade efetiva, permeabilidade, volume de argilas, etc., são as características que,

associadas às rochas, definem as fácies (e correspondentes eletrofácies) que buscamos. Por exemplo, podemos querer caracterizar zonas de produção a partir de fácies petrofísicas.

Dentro deste contexto, iremos trabalhar com o reconhecimento de fácies reservatório a partir de suas correspondentes eletrofácies, usando para tal um modelo quantitativo definido através de regressão logística.

2.3 Natureza dos Dados de Perfis

Neste ponto, convém inserir uma breve discussão sobre a natureza dos dados oriundos dos perfis elétricos. Definimos perfis como sendo o registro contínuo, analógico ou digital, obtido em um poço a intervalos regulares de profundidade, de propriedades físicas do meio investigado. Existem ferramentas específicas para a leitura de raios gama, resistividade, tempo de trânsito de uma onda compressional, etc.. Cada uma destas propriedades tem por objetivo fornecer uma indicação de alguma propriedade da rocha e/ou do fluido contido nela. Trabalhamos, portanto, com medidas **indiretas** das propriedades de interesse. Daí a necessidade de **calibrar** estas leituras de perfis a partir de observações **diretas** propriedades, averiguadas em uma amostra da rocha. Estas amostras geralmente provêm de testemunhos, que, conforme comentamos anteriormente, são dispendiosos e raros. Torna-se óbvia, assim, a necessidade de se obter um modelo preditivo de correlação rocha-perfil. Um outro fato para o qual deve-se atentar com relação aos dados de perfis é que eles são, de maneira geral, muito afetados pelo ambiente dos poços. Temperatura, composição e salinidade do fluido de perfuração utilizado, por exemplo, influenciam sobremaneira as leituras efetuadas pelas ferramentas de perfilagem. Existem, naturalmente, métodos para filtrar estas influências, mas pode-se esperar uma perda ou, pelo menos, atenuação do sinal adquirido pela ferramenta.

Importante, também, é o suporte sobre o qual as informações são obtidas. Diferentes volumes de rocha investigados tendem a fornecer diferentes valores para a propriedade analisada. Ou seja, defrontamo-nos, também, com o problema de escala da medida

considerada. Diferentes perfis investigarão diferentes volumes de rocha e fornecerão distintas imagens do meio.

Os dados de perfis utilizados nas análises não são os dados brutos, isto é, as leituras diretamente realizadas pelas ferramentas. Eles passam por um processo de suavização, geralmente representado por uma média móvel, fazendo com que haja alguma perda de informações devido à atenuação do contraste verificado entre as camadas diferentes. Tenta-se, através de um processo conhecido por deconvolução, reverter esta suavização, recuperando os dados originais, mas frequentemente de maneira infrutífera. Uma outra maneira tentativa de recuperar-se um pouco da informação perdida é utilizar os perfis quadratizados, ou seja, considerando cada intervalo relativamente homogêneo como sendo representado por um valor constante (o valor médio do intervalo, por exemplo). Existem vários métodos de quadratização de perfis, sendo um dos mais simples o método das derivadas, onde os limites entre as zonas homogêneas são definidos pelos pontos de inflexão das curvas, entre os pontos de máximo e mínimo locais.

2.4 Técnicas Multivariadas aplicadas à correlação rocha-perfil

Caracterizados os dados quanto à sua natureza e definidos claramente os objetivos a atingir com a correlação rocha-perfil, passamos agora a discutir rapidamente os modelos quantitativos mais comumente utilizados em correlação rocha-perfil. Vamos nos ater a modelos decorrentes da utilização de métodos estatísticos multivariados, objetivo principal deste estudo, comentando brevemente alguns exemplos selecionados da literatura a partir de 1990. Até esta data uma ampla cobertura bibliográfica é encontrada em Sarzenski & Toledo (op.cit).

O procedimento mais comumente adotado tem sido a utilização da função discriminante isoladamente ou em conjunto com outras técnicas multivariadas como análise de agrupamentos, componentes principais, análise de fatores, etc. , para identificar fácies a partir de perfis. Hook et alli (1994) aplicaram a técnica discriminante a arenitos de um campo

de gás no Mar do Norte, com o intuito de reconhecer litofácies e obter uma previsão do comportamento da permeabilidade nos poços não testemunhados. Lin & Salish (1994), com o objetivo de avaliar as propriedades petrofísicas (porosidade e permeabilidade) de arenitos argilosos fizeram uso da discriminação. As eletrofácies foram previamente associadas às litofácies de conteúdo petrofísico conhecido via componentes principais e análise de agrupamentos sendo posteriormente discriminadas. Elek (1988) e Woronow (1987) discutem, respectivamente a análise de componentes principais e análise de agrupamentos como ferramentas úteis à correlação rocha-perfil. Uma interessante comparação entre a análise discriminante e a análise de regressão linear múltipla pode ser encontrada em Greder et alli (1995). Buscando estimar permeabilidade a partir de perfis em poços não testemunhados, estes autores chegaram à conclusão, que para o conjunto de dados analisado, a melhor característica preditiva é obtida a partir da análise discriminante.

No Brasil, principalmente no âmbito da Petrobras SA, as aplicações de técnicas multivariadas na correlação rocha-perfil tem sido cada vez mais frequentes. Pode-se citar, a este exemplo, os artigos de Bucheb (1991,1992) que tem trabalhado extensivamente com estas técnicas, notadamente com a análise de componentes principais. Castro (1991) propõe um algoritmo baseado em agrupamentos disjuntos, para reconhecimento de eletrofácies. Souza Jr. (1991) aplica as técnicas de agrupamento, discriminante e regressão múltipla à caracterização de reservatórios da Bacia Potiguar. Steagall et alli (1995) buscam estimar permeabilidades a partir de perfis em rochas carbonáticas da Bacia de Campos, utilizando análise discriminante e regressão linear múltipla.

Não foi encontrada, na literatura pesquisada, referência à utilização de modelos logísticos na correlação rocha-perfil.

3. Desenvolvimento do Modelo Teórico

3.1 O Modelo Logístico

A análise de regressão logística é utilizada para descrever o relacionamento entre uma variável-resposta categórica (discreta) e uma ou mais variáveis explanatórias que podem ser discretas e/ou contínuas. No caso específico deste estudo que analisa a correlação rocha-perfil, vamos nos ater exclusivamente ao problema tratado, ou seja, o relacionamento entre uma variável-resposta discreta, - mais especificamente, binária - representando a ocorrência ou não de fácies reservatório e um conjunto de variáveis explanatórias que representam as diversas curvas de perfis elétricos utilizadas.

Tome-se, por exemplo, o seguinte modelo :

$$Y = \beta_0 + \beta_i X_i + e \quad (3.1)$$

onde Y, a variável fácies, é estimada a partir da função linear simples assinalada acima, com β_0 e β_i representando os parâmetros a estimar e X_i o i-ésimo perfil. Os valores que pode assumir são 1, no caso da fácies representar a ocorrência de rocha reservatório e 0, caso contrário. Examinemos mais detalhadamente seu comportamento :

$$E[Y] = 1 \cdot P(Y=1) + 0 \cdot P(Y=0) = P(Y=1)$$

Chamemos de $\pi(x)$ esta probabilidade, $P(Y=1)$. Então,

$$P(Y=0) = 1 - P(Y=1) = 1 - \pi(x)$$

$$E[Y^2] = 1^2 \cdot \pi(x) + 0^2 \cdot [1 - \pi(x)] = \pi(x)$$

$$\text{Var}[Y] = E(Y^2) - [E(Y)]^2 = \pi(x) - \pi(x)^2 = \pi(x)[1 - \pi(x)]$$

Daqui decorrem algumas conclusões importantes :

1. Y é heterocedástica, ou seja, sua variância não é constante;
2. $\pi(x)$ é uma probabilidade, $0 \leq \pi(x) \leq 1$, significando a probabilidade de ocorrência da fácies reservatório. Neste caso, o modelo de regressão linear tradicional poderia fornecer valores estimados de Y fora do intervalo $[0,1]$, o que não faria sentido ;
3. Como variável discreta, Y não pode ser normalmente distribuída, como pressupõe a regressão linear.

Um modelo que atende estas restrições é o modelo logístico. De maneira diversa da regressão linear, a probabilidade de observarmos a ocorrência de uma determinada fácies é uma função não linear das variáveis explanatórias. Entretanto, é possível linearizá-lo através do logito e este fato possibilita o desenvolvimento do modelo. Vamos considerar $\pi(x)$ a probabilidade de ocorrência de fácies reservatório e $1 - \pi(x)$ a ocorrência de fácies não-reservatório. A razão

$\frac{\pi(x)}{1 - \pi(x)}$ é chamada *odds*. O logaritmo dos *odds*, $\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right)$ é o chamado logito, cuja transformação dá nome ao método.

Vamos considerar n observações de ocorrências de determinada fácies a partir das p curvas de perfis. Ou seja, para cada i fácies são medidos p perfis $x_{i1}, x_{i2}, \dots, x_{ip}$.

Assim, temos :

$$\text{logito}[\pi(x_i)] = \ln\left[\frac{\pi(x_i)}{1 - \pi(x_i)}\right] = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j \quad (3.2)$$

Exponenciando e explicitando $\pi(x_i)$, vem :

$$\frac{\pi(x_i)}{1 - \pi(x_i)} = e^{\beta_0 + \sum_{j=1}^p x_{ij}\beta_j}$$

$$\begin{aligned}
\pi(x_i) &= [1 - \pi(x_i)] e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j} \\
\pi(x_i) + \pi(x_i) e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j} &= e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j} \\
\pi(x_i) &= \frac{e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j}} \quad (3.3)
\end{aligned}$$

$\pi(x_i)$ é a probabilidade estimada da ocorrência de determinada fácies. Mais precisamente, é a probabilidade de ocorrência da fácies reservatório, dado um determinado conjunto de curvas de perfis.

Logo,

$$P(Y_i = 1 | x_i) = \pi(x_i) = \frac{e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^p x_{ij} \beta_j}} \quad (3.4)$$

Com $i=1,2,\dots,n$ observações e X_i o vetor que contém os valores das curvas de perfis para a i -ésima observação.

Vamos verificar o que acontece com estas probabilidades estimadas.

Consideremos, para simplificar, $g(x) = \beta_0 + \sum_{j=1}^p x_{ij} \beta_j$. Então.

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (3.5)$$

Para $g(x)=0$ temos $\pi(x)=0.5$.

Se $g(x)$ cresce muito, $\lim_{g(x) \rightarrow \infty} \pi(x) = 1$

Se $g(x)$ é muito pequeno, $\lim_{g(x) \rightarrow -\infty} \pi(x) = 0$

Verifica-se, portanto, que o modelo atende as restrições que o problema de estimativa de fácies impõe com respeito aos limites da estimativa de $\pi(x)$.

Analisemos, também, o comportamento da função logito,

$$\text{logito}[\pi(x)] = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) \quad (3.6)$$

com $\pi(x) \in (0,1)$, cujo gráfico é mostrado na figura 3.1-1.

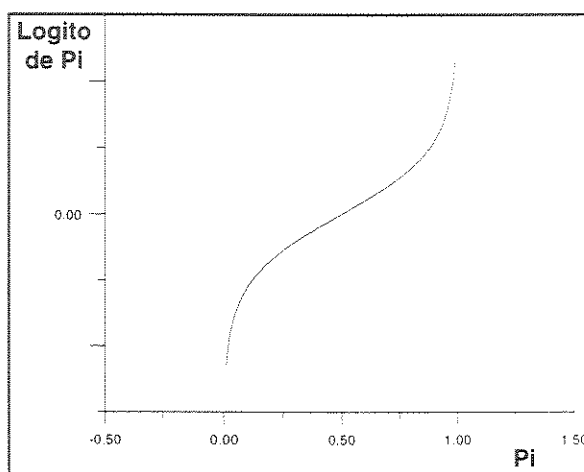


Figura 3-1 Gráfico de $\text{logito}[\pi(x)]$ versus $\pi(x)$

Verifica-se que seu domínio é o intervalo $(0,1)$ e sua imagem o intervalo $(-\infty, +\infty)$. Sua

derivada $\frac{d\{\text{logito}[\pi(x)]\}}{d\pi(x)} = \frac{1}{\pi(x)[1 - \pi(x)]}$ indica uma função crescente no intervalo de

domínio considerado, o que faz com que cada ponto no intervalo $(0,1)$ seja levado a apenas um ponto do intervalo $(-\infty, +\infty)$, fazendo com que esteja solucionada a talvez maior restrição à utilização da regressão linear a este tipo de problema. No caso do modelo de regressão logístico as probabilidades estimadas encontram-se coerentemente dentro do intervalo $(0,1)$.

3.2 Ajuste do Modelo

Ajustar um modelo de regressão logística a um conjunto de dados significa estimar os $p+1$ parâmetros β da equação (3.4). Na regressão linear isto é feito através dos quadrados mínimos, que minimiza a soma dos quadrados dos desvios entre os valores observados e os valores preditos a partir do modelo. No modelo logístico, entretanto, o método é diferente. A estimativa será feita através da função de máxima verossimilhança. O objetivo é selecionar os parâmetros que maximizam a probabilidade de obtenção de um determinado conjunto de dados. Uma descrição completa do método pode ser encontrada em, por exemplo, McCullagh & Nelder (1983).

O primeiro passo é construir uma função denominada função de verossimilhança, que exprime a probabilidade de ocorrências verificadas em um conjunto de dados em função dos parâmetros β_i . Os estimadores destes parâmetros são os valores que maximizam a função. As probabilidades associadas à ocorrência de fácies reservatório $P(Y = 1|x_i) = \pi_i$ e à ocorrência de fácies não-reservatório, $P(Y = 0|x_i) = 1 - \pi_i$ podem ser expressas da seguinte forma (para simplificar a notação usaremos π_i , deixando claro ser uma função dos x_i):

$$\ell_i = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (3.7)$$

Como assume-se que as observações são independentes, a função de verossimilhança é obtida a partir do produto dos termos ℓ_i :

$$\ell = \prod_{i=1}^n \ell_i \quad (3.8)$$

Portanto, devemos estimar os valores que maximizam a expressão acima. Estes valores são os β_i pertencentes ao vetor β de parâmetros estimados. Ou seja,

$$\ell(\beta) = \prod_{i=1}^n \ell_i \quad (3.9)$$

Este é um modelo não linear nos parâmetros, com tratamento computacional complicado. Vamos considerar o logaritmo da expressão acima,

$$\ln[\ell(\beta)] = L(\beta) = \sum_{i=1}^n (y_i \ln \pi_i) + (1 - y_i) \ln(1 - \pi_i) \quad (3.10)$$

Assim, nosso problema passa a ser encontrar os valores que maximizam a expressão acima. Isto é feito diferenciando $L(\beta)$ em relação aos $\beta_0, \beta_1, \dots, \beta_p$ parâmetros e igualando as expressões resultantes a zero, gerando as seguintes expressões, respectivamente para β_0 e para os demais parâmetros :

$$\sum_{i=1}^n (y_i - \pi_i) = 0 \quad (3.11)$$

e

$$\sum_{i=1}^n x_{ij} (y_i - \pi_i) = 0 \quad (3.12)$$

com j significando as p curvas de perfis. Estas $p+1$ equações não são lineares nos parâmetros, fazendo com que estes sejam obtidos a partir de algum método iterativo. Os pacotes comerciais usam comumente o método de Newton-Raphson. A equação (3.11) mostra uma relação interessante :

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{\pi}_i \quad (3.13)$$

Ou seja, a soma dos valores observados é igual à soma dos valores preditos, já dando uma informação preliminar a respeito da qualidade do ajuste do modelo. As variâncias e covariâncias dos coeficientes estimados são calculados a partir das derivadas segundas da função $L(\beta)$ e têm a forma geral :

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (3.14)$$

e

$$\frac{\partial^2 L(\beta)}{\partial \beta_i \partial \beta_j} = - \sum_{i=1}^n x_{ij} x_{iu} \pi_i (1 - \pi_i) \quad (3.15)$$

com $j, u = 1, 2, \dots, p$. A matriz $(p+1) \times (p+1)$ que contém os negativos dos termos fornecidos por estas equações chama-se matriz de informações, $\mathbf{I}(\beta)$. As variâncias e covariâncias procuradas são obtidas da inversa da matriz de informações, $\mathbf{V}(\beta) = \mathbf{I}^{-1}(\beta)$. O j -ésimo elemento da diagonal desta matriz é a variância de $\hat{\beta}_j$, $\sigma(\hat{\beta}_j)$ e os elementos fora da diagonal representam as covariâncias $\sigma(\hat{\beta}_j, \hat{\beta}_u)$ entre $\hat{\beta}_j$ e $\hat{\beta}_u$. A matriz dos estimadores das variâncias e covariâncias, $\hat{\mathbf{V}}(\hat{\beta})$ contém os valores $\hat{\sigma}(\hat{\beta}_j)$ e $\hat{\sigma}(\hat{\beta}_j, \hat{\beta}_u)$, $j, u = 0, 1, 2, \dots, p$, calculados resolvendo-se $\mathbf{V}(\beta)$ para $\hat{\beta}$. O erro padrão estimado destes coeficientes estimados é :

$$\hat{SE}(\hat{\beta}_j) = [\hat{\sigma}^2(\hat{\beta}_j)]^{1/2} \quad (3.16)$$

O leitor interessado em maiores detalhes dos procedimentos e definições acima pode se referir a Hosmer & Lemeshow (1989).

3.3 Avaliação do Modelo

Após estimados os coeficientes do modelo logístico é preciso avaliá-lo. O primeiro passo normalmente é verificar a significância das variáveis consideradas no modelo. Isto normalmente é feito a partir de testes sob hipótese nula, ou seja, queremos testar a hipótese do parâmetro em consideração ser nulo, $H_0: \beta_i = 0$. A significância a que nos referimos tem o significado clássico da significância estatística. Atenta-se para este detalhe porque o fato de os parâmetros serem significantes em um determinado nível estatístico não quer dizer, obrigatoriamente, que o modelo é adequado aos dados observados. O que irá nos dizer se os valores dos β estimados definem um modelo que seja representação acurada do conjunto de dados será o diagnóstico do modelo, a partir de suas estatísticas, criadas para se avaliar a qualidade do ajuste.

Com relação à significância, de maneira geral a avaliação é feita comparando-se os valores observados com os preditos a partir de modelos com e sem a variável analisada. Esta comparação é feita pela estatística da razão de verossimilhança :

$$G = -2 \ln \left[\frac{L(\text{modelo_sem_a_variavel})}{L(\text{modelo_com_a_variavel})} \right] = -2 \ln \left[\frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta})} \right] \quad (3.17)$$

Existem ainda duas outras estatísticas equivalentes à da razão de verossimilhança, a estatística de Wald e a dos escores eficientes (*score test*). Seu cálculo, embora relativamente simples no particular caso univariado, exibe um grande grau de complexidade no caso multivariado. No caso univariado, a título de exemplo, a estatística de Wald compara a estimativa do parâmetro à estimativa do erro padrão :

$$W = \frac{\hat{\beta}_1}{\hat{SE}(\hat{\beta}_1)} \quad (3.18)$$

Esta estatística tem distribuição normal padrão sob a hipótese nula. O *score test*, por sua vez, é baseado na distribuição condicional da derivada primeira em relação a β_1 da equação (3.12) dada a derivada primeira em relação a β_0 , equação (3.11). No caso multivariado estas estatísticas tem a seguinte forma :

$$W = \hat{\beta}' \left[\hat{V}(\hat{\beta}) \right]^{-1} \hat{\beta} \quad (3.19)$$

e

$$ST = U'(\hat{\beta}) I^{-1}(\hat{\beta}) U(\hat{\beta}) \quad (3.20)$$

onde U , o vetor dos escores eficientes, é o vetor que contém as derivadas primeiras da função de verossimilhança em relação aos parâmetros, e $I(\beta)$ é a matriz de informação já anteriormente referenciada. Sob a hipótese nula de que os p coeficientes das variáveis do modelo sejam iguais a zero, estas três estatísticas têm distribuição qui-quadrado com p graus de liberdade.

3.4 Qualidade do ajuste/diagnóstico do modelo

Ajustado o modelo, precisamos verificar o quão adequado ele está, isto é, se ele é realmente efetivo para descrever o conjunto de dados. Uma maneira simples de se verificar isto é

comparando a distância entre o valor observado y_i e o valor estimado, $\hat{y}_i$. Em outras palavras, fazer uma análise de resíduos, $(y_i - \hat{y}_i)$. Pregibon (1981), em trabalho clássico, discute e propõe diversas medidas e maneiras de se avaliar a qualidade do ajuste. Hosmer & Lemeshow (1989) são uma boa referência para uma discussão mais atualizada. Estas medidas incluem, por exemplo, o resíduo de Pearson e o resíduo da estatística de verossimilhança (chamada desvio, *deviance*, por Pregibon).

Como estes métodos têm caráter essencialmente prático - Pregibon (op. cit.) propõe o uso extensivo de gráficos índices, por exemplo - serão comentados, quando necessário, no próximo capítulo, que tratará de uma aplicação prática do modelo logístico.

4. Regressão Logística usando a Proc Logistic

4.1 Introdução

Este capítulo tratará de uma análise de regressão logística para estimar fácies a partir de um conjunto real de dados. O objetivo, além de ilustrar uma aplicação da técnica, é sugerir um enfoque alternativo a este tipo de problema de correlação rocha-perfil, podendo eventualmente vir a ser adotado como procedimento corriqueiro para conjuntos de dados de características semelhantes.

O problema que enfocaremos, como já referido anteriormente, será a estimativa de fácies reservatório e fácies não-reservatório a partir do cálculo de sua probabilidade de ocorrência. Esta probabilidade será fornecida por um modelo de regressão logística construído tomando-se por base as curvas de perfis elétricos e litofácies obtidas a partir da descrição de testemunhos.

Primeiramente serão gerados modelos logísticos para um conjunto pré determinado de variáveis. Isto se justifica porque pretende-se fazer uma comparação entre os resultados obtidos pela análise de regressão logística e os oriundos de uma análise discriminante prévia, que utilizou este mesmo conjunto de variáveis.

A seguir, criaremos modelos logísticos próprios, com a seleção das variáveis a serem incluídas no modelo se dando a partir do conjunto completo de variáveis. Será discutida a técnica stepwise de seleção de variáveis.

Finalmente, será feita uma comparação entre os resultados obtidos pelos diferentes modelos e também pela análise discriminante. Para a realização da modelagem será utilizada a proc Logistic do SAS. É possível o estudo de modelos logísticos com outras proes, por exemplo a Catmod para dados categóricos de maneira geral, mas a Logistic, para nossos

propósitos, é melhor solução por ser específica e fornecer os resultados de forma direta e objetiva. À medida que formos realizando a análise colocaremos eventualmente alguns trechos de programas SAS e/ou de suas saídas para facilitar os comentários e a interpretação dos resultados.

A listagem completa dos programas SAS utilizados encontra-se no apêndice 1 e suas respectivas saídas, no apêndice 2.

4.2 Descrição dos dados

Serão utilizados dados provenientes de 2 poços de um campo de Petróleo denominado Pargo, localizado na Bacia de Campos, no estado do Rio de Janeiro (ver mapa de localização, figura 4.1). Para este tipo de estudo, é considerado ser um campo de dificuldade mediana, onde os reservatórios principais são arenitos e conglomerados intercalados a folhelhos. Este conjunto de rochas, em geral, exibe um bom contraste nas leituras de perfis, facilitando a análise. Os dados estão indexados pela profundidade e compõem-se da própria profundidade (vertical), das fácies e das leituras efetuadas pelos perfis elétricos.

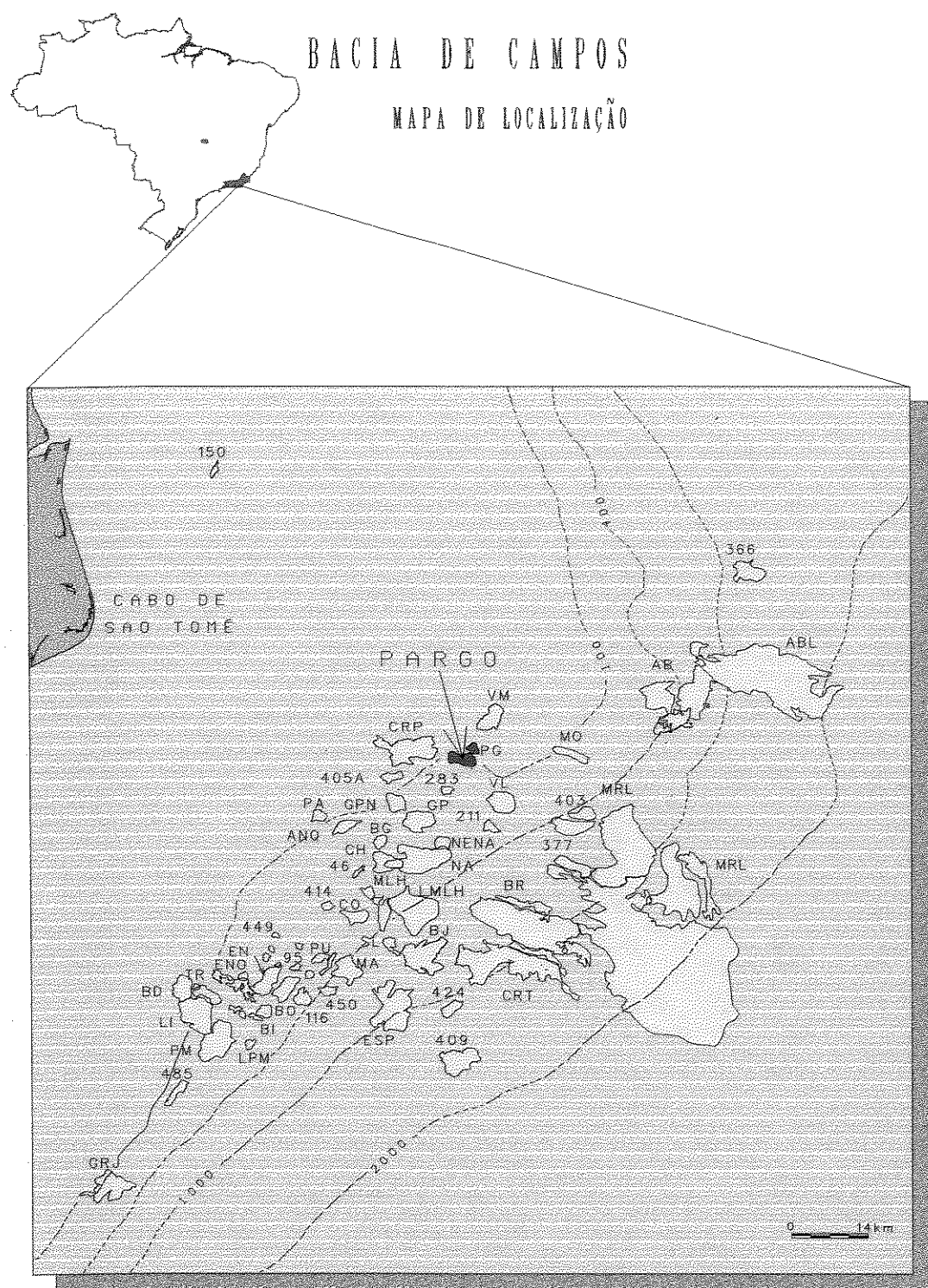


Figura 4-1 - Mapa de Localização do Campo de Pargo.

4.2.1 Contexto geológico

O campo de Pargo situa-se na porção norte da Bacia de Campos, sob uma lâmina d'água média de 100 metros. Seu reservatório principal, o Arenito Carapeba, de idade cretácea, é composto por sedimentos areno-conglomeráticos depositados por correntes de turbidez. A deposição deu-se segundo a direção ONO-ESE em uma calha alongada delimitada por dois altos estruturais gerados como efeito da halocinese sindeposicional que atuou na bacia. À medida que se depositavam os turbiditos o substrato acomodava-se a partir da movimentação das falhas lítricas de crescimento existentes na região limítrofe entre a calha e os altos estruturais. Isto possibilitou o acúmulo de uma expressiva pilha sedimentar que chega a até 260 metros de espessura.

Com relação à geometria externa, estes sedimentos constituem um complexo turbidítico de lobos canalizados recobertos por um complexo de canais e diques marginais (Bruhn & Moraes, 1988). Segundo Bruhn (1993), toda a sedimentação se deu em um sistema retrogradante onde a diminuição da taxa de sedimentação e consequente decréscimo da granulometria da área fonte tornou os eventos turbidíticos menos frequentes e com granulometria mais fina, dando origem a sequências *fining* e *thinning upward* bem definidas.

4.2.2 Descrição das litofácies

A partir da descrição dos testemunhos pode-se caracterizar, conforme Becker et alli (1989), seis litofácies : Conglomerado arenoso (Ca), Arenito conglomerático (Ac), Arenito maciço (Am), Interlaminação (I), Arenito estratificado (Ae) e Arenito argiloso deformado (D). As análises sequenciais de testemunhos para os dois poços encontram-se no apêndice 3

4.2.2.1 Conglomerado arenoso (Ca)

Rocha formada por grânulos e pequenos seixos (até 1 cm) dispersos em matriz arenosa de granulometria média a grossa, arcoseana. Os clastos mais frequentes são fragmentos de rochas graníticas, quartzo e feldspatos e compõem de 30 a 60% da rocha. Ocorrem eventualmente níveis de cimentação calcítica intensa. Exibe comumente gradação normal.

4.2.2.2 Arenito conglomerático (Ac)

Arenitos de granulometria média a grosseira com grânulos e, secundariamente, seixos dispersos de maneira aleatória. Os clastos podem atingir percentuais de 20 a 30% e, como a litofácies anterior, compõem-se de fragmentos de rochas graníticas, quartzo e feldspatos. Mostram, ainda, uma gradação preferencialmente normal, com o tamanho e percentual dos clastos diminuindo em direção ao topo das camadas. Podem ocorrer níveis de cimentação calcítica ou silicosa.

4.2.2.3 Arenito maciço (Am)

Arenitos de granulometria média a fina, eventualmente grosseira a muito fina, seleção boa a moderada, aspecto maciço, com gradação normal. Na porção fina pode ocorrer laminação plano-paralela. Ocorrem níveis com cimentação calcítica.

4.2.2.4 Interlaminado (I)

Porções de alternância de camadas e lentes de arenito muito fino, siltito, folhelho argiloso, marga e calcilito. Espessuras de milimétricas a centimétricas para as camadas e lentes, com a transição entre os níveis podendo dar-se de maneira gradacional ou brusca. Estratificação plano-paralela, às vezes difusa e micro estratificação cruzada. Presença de bioturbação em grau variável.

4.2.2.5 Arenito estratificado (Ae)

Arenitos finos ou muito finos gradando normalmente a siltitos e intercalações de siltitos e folhelhos para o topo das camadas. Comumente encontra-se toda cimentada por calcita. Relativamente rica em estruturas sedimentares, com laminação plano-paralela, estratificações cruzadas de pequeno porte, microestratificações cruzadas, etc. Ocorrência muito pouco frequente nos poços considerados.

4.2.2.6 Arenito argiloso deformado (D)

Engloba as rochas sem estruturas primárias bem definidas, decorrentes da homogeneização dos sedimentos por movimentação plástica e por bioturbação. Pode se mostrar total ou parcialmente homogeneizada, neste caso sob a forma de camadas com intercalações centimétricas de arenitos muito finos com siltitos arenosos e folhelhos intensamente deformados. É comum a presença de diques e soleiras de arenito nos níveis mais finos. Também muito pouco frequente nos poços analisados.

4.2.3 Relacionamento entre as litofácies e respostas em perfis elétricos

Dentre as litofácies descritas e que encontram-se sumarizadas na tabela 4.1, apenas as litofácies Ca, Ac e Am constituem reservatório. Formam camadas cíclicas, granodecrescentes para o topo, de base comumente erosiva, eventualmente capeadas pelos siltitos e folhelhos da litofácies I. A porção basal do reservatório é comumente mais grosseira e espessa, contrastando com a porção superior, de ciclos menos espessos e granulometria mais fina, com predominância da litofácies Am. Estas três litofácies (Ca, Ac e Am) normalmente mostram passagens transicionais entre si, podendo, entretanto, apresentar contatos bruscos, devidos a erosão. Para o topo a litofácies Am pode transicionar para a litofácies I - que é mais frequentemente observada no topo do reservatório - intercalada aos estratos arenosos. As litofácies D e Ae são de ocorrência restrita, normalmente associadas à litofácies I.

De maneira geral, os perfis elétricos, em particular o de densidade, respondem razoavelmente bem às variações petrográficas. Os perfis de raios gama identificam bem os níveis silticos e argilosos servindo para delimitar de maneira satisfatória as litofácies com características de reservatório. As litofácies Am e Ac/Ca exibem boas associações com o perfil de densidade, com um aumento da densidade e consequente diminuição da porosidade indicada pelo perfil sendo associada à diminuição do grau de seleção e do aumento da proporção de minerais densos da litofácies Am em direção às litofácies Ac e Ca. Dentro da litofácies Am o perfil também corrobora com a granodecrescência ascendente verificada nas

rochas, com densidades maiores na base (menores porosidades) e menores no topo (maiores porosidades).

LITO FÁCIES		AMBIENTE	LITOLOGIA	ESTRUT. SEDIM.	SELEÇÃO	DISTRIBUIÇÃO	GRANULOMETRIA
CONGL. ARENOSO	Ca*	LOBOS CANALIZADOS	GRANULOS E SEIXOS DISPERSOS EM MATRIZ ARENOSA	GRAD. NORMAL E MACIÇO	MUITO BAIXA	BASE DE CICLOS	GROSSEIRA/LÍTICOS SFRIÁVEL
ARENITO CONGL.	Ac*	LOBOS CANALIZADOS	GRANULOS E SEIXOS ALEATÓRIOS	MACIÇA. DISH. LAM. PARALELA	POBRE	BASE/MEIO CICLO	MED/GROSSA FRIÁVEL
ARENITO MACIÇO	Am*	LOBOS CANALIZADOS	INTRACLASTOS DE PELITOS E MARGAS	MACIÇA. DISH	BOA A MODERADA	TOPO DE CICLO	MED/FINA FRIÁVEL
INTERLAMINA DO	I	FÁCIES DISTAIS	INTERC. DE LENTES DE ARN MFNO. STO, FLH. MRG E CLU	LAM PARA LELA. RIPPLES		TOPO	
ARENITO ESTRAT.	Ae	FÁCIES DISTAIS	ARN, STO. STO/FLH. INTRACL., CIMENT. CAL	GRAD. NORMAL. LAM PAR.		RARO	MFNO/FNO
ARENITO ARGILOSO	D	FÁCIES DISTAIS	ARGILOSO. COM INTRACL. ARGILOSOS	MOVIMENTO TAÇÃO PLÁSTICA		RARO	

Tabela 4-1 Descrição das litofácies segundo Becker et alli (1989).

4.2.4 Curvas de fácies e perfis elétricos utilizados

Para representar as litofácies reservatório (Ca, Ac e Am) será criada a variável binária Realfac que assume os valores 1, caso as litofácies sejam de rochas reservatório e 0, caso contrário. Esta variável será utilizada como parâmetro, representando a nossa fácies-padrão, para fins de comparação com os valores a serem estimados. Esta simplificação será feita

porque a análise discriminante realizada tinha por meta identificar as fácies reservatório. Desta forma teremos conjuntos de dados semelhantes, possibilitando uma comparação entre os resultados obtidos pelas duas técnicas como será visto adiante.

Um segundo tipo de fácies, a variável Efac, acrônimo de eletrofácies, representa a fácies estimada a partir da análise discriminante. É, também, uma variável binária, codificada 1 caso a ocorrência seja de fácies reservatório e 0, caso contrário. A análise discriminante realizada (Franco Filho, 1993), contemplou o conjunto de variáveis Grco/Phin/Drdn/Nplt como sendo a suite que melhor explica o conjunto de dados. As variáveis Grco e Phin são leituras de perfis e exprimem, respectivamente, a radioatividade natural da rocha através da leitura dos raios-gama emitidos e o índice de neutrons contido nos fluidos que integram o volume de sistema rocha-fluido investigado pela ferramenta de perfilagem. Postula-se que este índice de neutrons é uma medida direta do volume de fluidos contido na rocha estando, portanto, diretamente relacionado à sua porosidade. Já as variáveis Drdn e Nplt não são leituras diretas dos perfis, mas combinações destas. Adota-se corriqueiramente este artifício pois as características expressas pelas curvas utilizadas tornam-se mais facilmente visíveis, facilitando o trabalho do intérprete de perfis. Por exemplo, valores negativos de DRDN costumam assinalar arenitos que no nosso caso são os reservatórios procurados. As transformações são as seguintes :

$$Nplt=(100-Phin)/(Rhoc-1) \quad (4.1)$$

$$Drdn=((Rhoc-2)/0.05)-((45-Phin)/3) \quad (4.2)$$

A variável Rhoc é, como as primeiras, uma leitura de propriedades das rochas, mais especificamente, de sua densidade.

Adicionalmente a estas curvas de perfis serão também consideradas, principalmente na etapa de geração dos modelos logísticos com escolha de variáveis, as variáveis Dlt, Rt, Phit e Phie. A primeira é uma leitura do tempo de trânsito de uma onda compressional no poço e é diretamente correlacionável à porosidade da rocha. A segunda, Rt, é uma medida da resistividade do meio e está diretamente relacionada ao tipo de fluido presente no sistema. As duas últimas, Phit e Phie são curvas de perfis secundárias, obtidas a partir do processamento

de outros perfis e exprimem respectivamente a porosidade total da rocha e a porosidade efetiva ao fluxo de fluidos.

4.3 Análise exploratória dos dados de perfis

Antes de passarmos ao detalhamento dos dados de perfis é necessário que sejam feitos alguns comentários a respeito da natureza dos dados analisados. Via de regra, qualquer conjunto de dados referente a reservatórios tende, por uma série de motivos, o principal deles econômico, a privilegiar algumas características em detrimento de outras. Isto é bem natural, considerando-se que os investimentos de tempo e recursos financeiros visam alcançar objetivos claramente definidos no sentido de obtenção de máximo retorno do investimento. Assim, no processo que culmina com a exploração de um campo de petróleo, teremos sempre uma amostragem seletiva. O caso aqui considerado não é, naturalmente, diferente. O poço P1 era originalmente um poço de desenvolvimento e, portanto, foi perfurado em uma época onde o conhecimento do campo já era consideravelmente maior do que o correspondente à época da perfuração do poço P2, originalmente um poço exploratório. Desta forma, no conjunto dos dados, o poço P1 exibe características “melhores” com relação às propriedades de reservatório que o poço P2. Por características melhores assinalamos quantidades como por exemplo a razão fácies reservatório/fácies não-reservatório por poço, coluna de óleo, etc. Isto fica bem evidente na análise a seguir. Os histogramas, box-plots, q-q plots, bem como as estatísticas calculadas para cada variável considerada encontram-se no apêndice 4.

4.3.1 Poço P1

De maneira geral as variáveis diretamente relacionadas à porosidade (Dlt , $Phin$, $Rhoc$) exibem distribuições razoavelmente centradas e simétricas. As transformadas $Nplt$ e $Drdn$ mostram comportamento semelhante entre si, cada qual ligeiramente assimétrica em direção aos maiores e menores valores, respectivamente. A variável $Grco$ mostra um interessante comportamento diretamente relacionado ao problema de amostragem tendenciosa referenciado acima. Exibe uma forte assimetria positiva, concentrando-se nos baixos valores.

Era-se de esperar este comportamento, pois valores baixos de Grco estão associados às fácies reservatório, que neste poço ocorrem em maior proporção. Um outro aspecto interessante é que as variáveis Phin, Dlt e Rhoc (associadas à porosidade), apesar da amostragem seletiva, mantiveram-se aproximadamente simétricas. Tal comportamento deve-se ao fato de que elas estão exprimindo, na verdade, a porosidade total da rocha e não sua porosidade efetiva. A tabela 4.2 ilustra bem isto. Nas fácies reservatório, Phie exibe um valor médio bem próximo ao verificado para Phit (e também Phin, Dlt e Rhoc), caindo para algo em torno da metade nas fácies não-reservatório, enquanto as demais permanecem grosseiramente constantes.

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	304	3039.66	17.6562667	3005.00	3070.00
DLT	Sonico	265	81.0644151	4.9219183	66.0100000	94.8400000
RT	Resistividade	265	2.4429811	1.0464553	0.5700000	4.7700000
PHIN	Poros. Neutrons	301	20.6455934	3.3785140	10.7588000	29.4786000
RHOC	Densidade	301	2.3191385	0.0808332	2.0403000	2.5481000
EFAC	EletroFacies	304	0.7664474	0.4237884	0	1.0000000
PHIT	Poros. Total	265	22.1377736	3.0119802	12.4200000	31.7600000
GRCO	Raios-Gama	271	58.8451292	15.8599650	40.9600000	104.0700000
PHIE	Poros. Efetiva	265	18.8329506	5.1240782	0	31.7648000
REALFAC	Facies Padrao	304	0.7138158	0.4527218	0	1.0000000
NPLT	Razao Phin,Rhoc	301	60.3386204	3.8076038	49.9226482	74.1606267
DRDN	Diferenca Rhoc,Phin	301	-1.7353648	1.7534231	-6.5771000	3.3641133

Facies Padrao=0						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	87	3039.63	23.0935584	3005.00	3070.00
DLT	Sonico	48	82.5602083	3.9083062	74.9600000	92.1600000
RT	Resistividade	48	2.0943750	0.5999287	0.5900000	2.8500000
PHIN	Poros. Neutrons	84	21.7492024	3.7421098	12.2493000	28.1834000
RHOC	Densidade	84	2.3985798	0.0784304	2.2804000	2.5481000
EFAC	EletroFacies	87	0.1839080	0.3896551	0	1.0000000
PHIT	Poros. Total	48	21.1181250	2.2336638	17.0600000	25.6100000
GRCO	Raios-Gama	54	81.2762963	13.9671904	48.6100000	104.0700000
PHIE	Poros. Efetiva	48	11.6054271	4.2852199	5.0149000	19.7518000
NPLT	Razao Phin,Rhoc	84	56.1405138	4.3782264	49.9226482	63.9478107
DRDN	Diferenca Rhoc,Phin	84	0.2213294	2.0935644	-3.3545667	3.3641133

Facies Padrao=1						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	217	3039.67	14.9989819	3013.00	3064.40
DLT	Sonico	217	80.7335484	5.0673267	66.0100000	94.8400000
RT	Resistividade	217	2.5200922	1.1077559	0.5700000	4.7700000
PHIN	Poros. Neutrons	217	20.2183899	3.1329454	10.7588000	29.4786000
RHOC	Densidade	217	2.2883871	0.0575069	2.0403000	2.4597000
EFAC	EletroFacies	217	1.0000000	0	1.0000000	1.0000000
PHIT	Poros. Total	217	22.3633180	3.1177330	12.4200000	31.7600000
GRCO	Raios-Gama	217	53.2631797	10.4625018	40.9600000	100.4600000
PHIE	Poros. Efetiva	217	20.4316654	3.7308500	0	31.7648000
NPLT	Razao Phin,Rhoc	217	61.9636940	1.8064265	56.3652525	74.1606267
DRDN	Diferenca Rhoc,Phin	217	-2.4927948	0.7215532	-6.5771000	-0.0448000

Tabela 4-2 Estatísticas descritivas para o poço P1 (geral e por fácies).

A variável R_t (resistividade) a partir do primeiro quartil, aproximadamente, mostra-se centrada e simétrica. Até este ponto, entretanto, exibe um comportamento anômalo atribuído à influência dos baixos valores observados para as fácies não-reservatório. De fato, a média para as fácies não-reservatório está bem próxima do valor correspondente à região limite do primeiro quartil. Deve-se ainda chamar atenção para o fato de que os baixos valores de resistividade observados para o poço devem-se a uma particular característica do campo, conforme descrevem Lemos et alli, (1989).

Para as fácies reservatório nota-se um claro predomínio de arenitos, como esperado, evidenciado pelo valor médio negativo da variável Dr_{dn} para esta fácies.

A figura 4.2 exibe a matriz de correlação e gráficos de dispersão para este poço. Chamam atenção os relativamente altos coeficientes de correlação verificados entre Gr_{co} e as variáveis transformadas N_{plt} e Dr_{dn} e a baixa correlação observada entre Ph_{in} e as variáveis transformadas. Na primeira associação (Gr_{co}) comprova-se ser o perfil de raios gama um bom indicador de fácies. À medida que Gr_{co} cresce (piores reservatórios) coerentemente também cresce Dr_{dn} e diminui N_{plt} . Com relação a Ph_{in} esperar-se-ia uma maior magnitude para o coeficiente de correlação, visto que as transformadas são criadas com sua participação. No gráfico de dispersão correspondente verifica-se a existência de dois grupos, mascarando, desta forma o coeficiente para o poço.

Matriz de Correlações / Poço P1							
Variável	dlt	rt	phin	rhoc	grco	nplt	drdn
dlt	1,0000	0,1737	0,6684	-0,3034	0,1126	-0,1477	0,1522
rt	0,1737	1,0000	0,2281	-0,3050	-0,1550	0,1423	-0,1312
phin	0,6684	0,2281	1,0000	-0,2076	0,4270	-0,4382	0,4530
rhoc	-0,3034	-0,3050	-0,2076	1,0000	0,4544	-0,7840	0,7775
grco	0,1126	-0,1550	0,4270	0,4544	1,0000	-0,6689	0,6909
nplt	-0,1477	0,1423	-0,4382	-0,7840	-0,6689	1,0000	-0,9964
drdn	0,1522	-0,1312	0,4530	0,7775	0,6909	-0,9964	1,0000

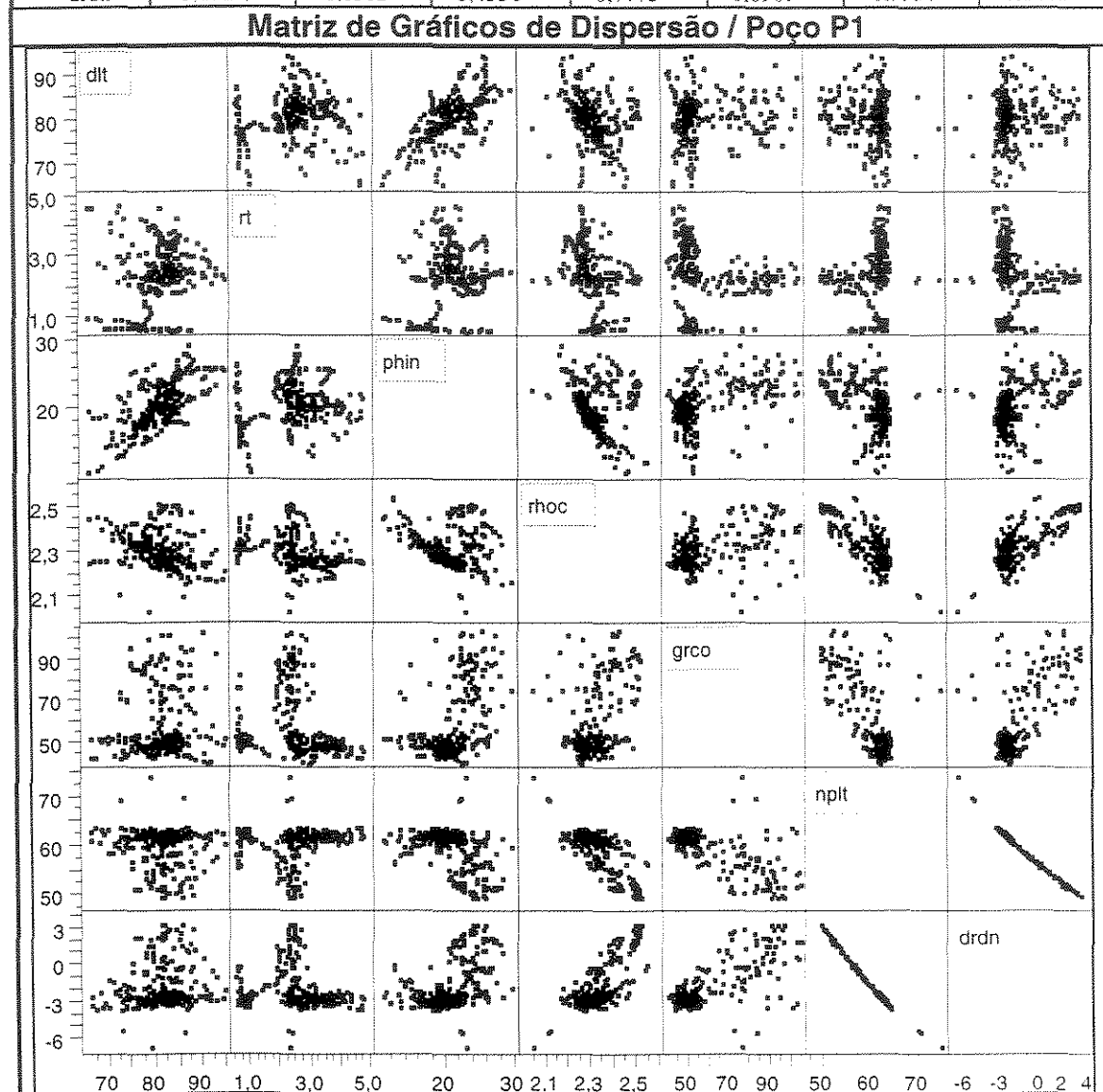


Figura 4-2 Matriz de Correlação e gráficos de dispersão para o poço P1

4.3.2 Poço P2

Mostra, em geral, comportamento semelhante a P1 (tabela 4.3). Adicionalmente pode-se verificar na análise por fácies que a variável $Phin$ responde melhor, exibindo valores médios próximos a $Phie$.

Constata-se também uma interessante inversão nos valores médios de Rt para este poço. Em condições normais deveríamos esperar valores maiores de Rt nas fácies reservatório, mas verifica-se o contrário. Isto se deve à relativamente menor coluna com óleo verificada para este poço quando comparado ao poço P1. Consequentemente a zona de água da formação (muito condutiva) faz abaixar o valor médio de Rt , o que explica a inversão verificada.

As correlações observadas entre as curvas (figura 4.3) exibem também um comportamento similar ao do poço P1. Desperta atenção a relativamente alta correlação direta verificada entre Rt e $Rhoc$, indicando que Rt está também respondendo a características da rocha. Novamente pode-se atribuir este comportamento à influência da zona de água da formação. À medida que a densidade cresce, menor fica a porosidade, consequentemente armazenando menos água e dando origem a maiores valores de Rt .

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	226	3057.50	13.0769517	3035.00	3080.00
DLT	Sonico	207	81.8788406	8.0810368	58.9800000	102.9400000
RT	Resistividade	207	1.9623671	0.8763099	0.4600000	4.8300000
PHIN	Poros. Neutrons	226	19.8721429	5.4975655	3.4179000	36.6002000
RHOC	Densidade	226	2.3937779	0.1016543	2.1723000	2.6002000
EFAC	EletroFacies	226	0.5929204	0.4923805	0	1.0000000
PHIT	Poros. Total	207	20.1040097	4.4806272	7.5300000	30.5100000
GRCO	Raios-Gama	221	72.8249321	19.4720205	45.7300000	126.4900000
PHIE	Poros. Efetiva	207	14.2942058	6.3452418	0.2558000	27.7328000
REALFAC	Facies Padrao	226	0.5663717	0.4966753	0	1.0000000
NPLT	Razao Phin,Rhoc	226	57.7380960	5.0791896	44.8276886	64.9721430
DRDN	Diferenca Rhoc,Phin	226	-0.5003948	2.4634071	-4.1060000	5.7421667

----- Facies Padrao=0 -----						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	98	3055.44	13.6799207	3035.00	3080.00
DLT	Sonico	79	82.0479747	7.0469636	64.3800000	102.9400000
RT	Resistividade	79	2.2364557	0.7604318	0.6000000	4.4600000
PHIN	Poros. Neutrons	98	22.3700888	5.0725118	10.8263000	36.6002000
RHOC	Densidade	98	2.4710357	0.0689539	2.2673000	2.6002000
EFAC	EletroFacies	98	0.0612245	0.2409742	0	1.0000000
PHIT	Poros. Total	79	19.8107595	3.6702420	13.0300000	29.6400000
GRCO	Raios-Gama	93	91.1141935	14.6113339	56.9200000	126.4900000
PHIE	Poros. Efetiva	79	8.8084810	3.9805137	0.2558000	21.0942000
NPLT	Razao Phin,Rhoc	98	52.8350788	3.5794006	44.8276886	63.0394432
DRDN	Diferenca Rhoc,Phin	98	1.8774105	1.7768785	-3.0203333	5.7421667

----- Facies Padrao=1 -----						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	128	3059.07	12.4208073	3038.00	3079.20
DLT	Sonico	128	81.7744531	8.6831083	58.9800000	102.3700000
RT	Resistividade	128	1.7932031	0.9027063	0.4600000	4.8300000
PHIN	Poros. Neutrons	128	17.9596531	5.0389902	3.4179000	29.5329000
RHOC	Densidade	128	2.3346273	0.0808945	2.1723000	2.5311000
EFAC	EletroFacies	128	1.0000000	0	1.0000000	1.0000000
PHIT	Poros. Total	128	20.2850000	4.9197983	7.5300000	30.5100000
GRCO	Raios-Gama	128	59.5366406	8.9017507	45.7300000	98.5100000
PHIE	Poros. Efetiva	128	17.6799266	5.0302100	5.2070000	27.7328000
NPLT	Razao Phin,Rhoc	128	61.4919686	1.7796197	56.9295746	64.9721430
DRDN	Diferenca Rhoc,Phin	128	-2.3209021	0.7977083	-4.1060000	-0.1654333

Tabela 4-3 Estatísticas descritivas para o poço P2 (geral e por fácies).

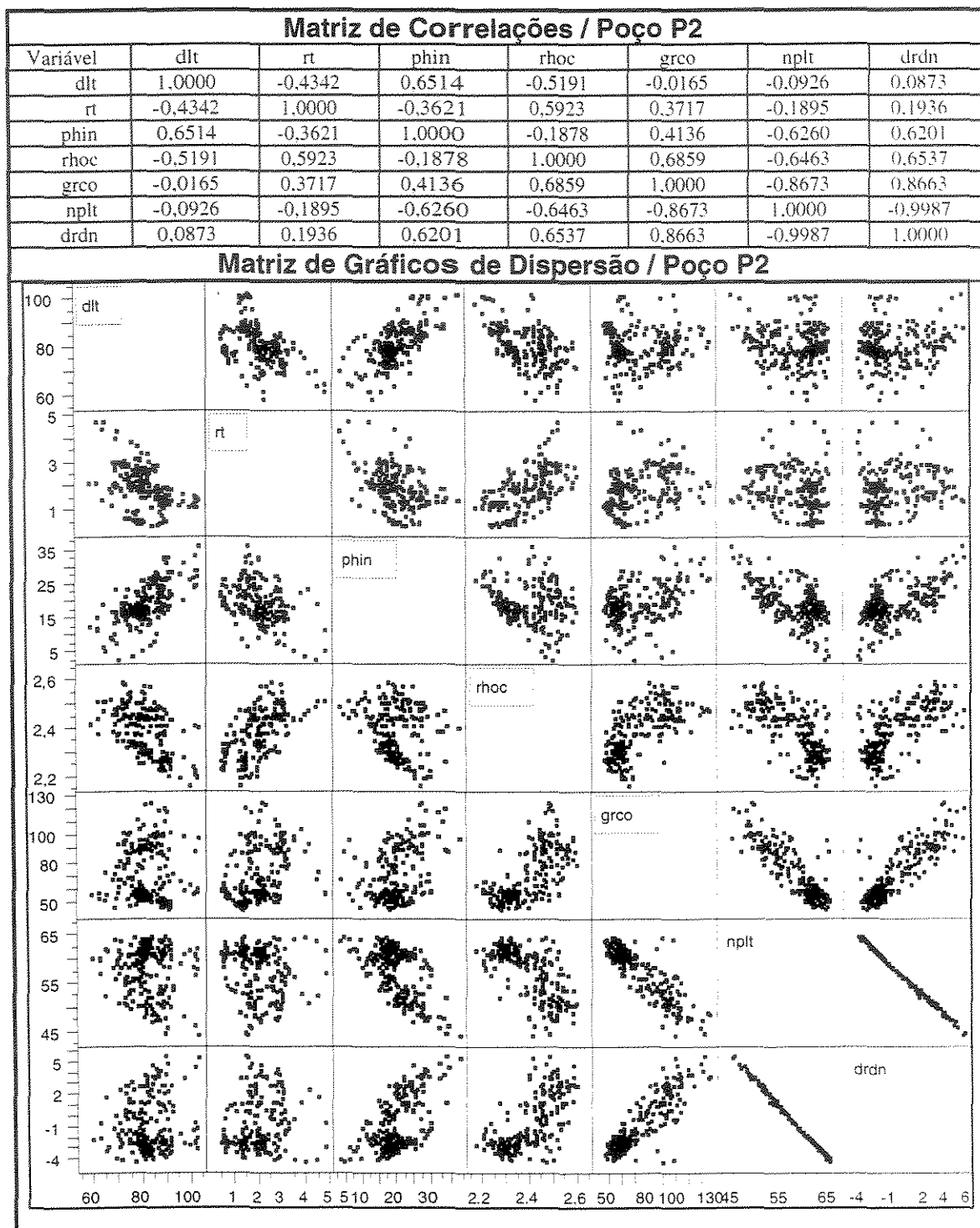


Figura 4-3 Matriz de correlação e gráficos de dispersão para o poço P2.

4.3.3 Comparação entre P1 e P2

Comparando os dois poços verifica-se de maneira geral que as distribuições das variáveis contidas em um e outro (apêndice 4) encontram-se deslocadas entre si de maneira coerente. Considerando que P1 é “melhor” que P2 (maior razão óleo/água, maior proporção de fácies reservatório, etc) temos as seguintes situações (sempre comparando as médias) : $Rt(P1>P2)$, $Phin(P1>P2)$, $Rhoc(P1<P2)$, $Grco(P1<P2)$, $Nplt(P1>P2)$, $Drdn(P1<P2)$. É importante caracterizar este deslocamento pois ele influirá de maneira importante nos parâmetros a serem estimados para os modelos logísticos. Para conjuntos de dados deslocados entre si pode-se esperar grandes variações na magnitude da estimativa de β_0 (termo independente), alterando pouco os β_i demais parâmetros.

Para todas estas variáveis, além do deslocamento, existe associada, embora de maneira bem sutil às vezes, uma mudança de escala de um poço em relação ao outro. Esta mudança de escala exercerá um efeito multiplicativo na estimativa dos parâmetros, quando comparados os dois poços. Um caso interessante e típico de mudança de escala é verificado com relação à variável Dlt. A média para os dois poços é praticamente a mesma, mas P2 exibe um range bem maior que P1. Pode-se fazer com que os dois poços fiquem na mesma escala utilizando-se um fator multiplicativo. A mudança de escala também afeta os parâmetros estimados, mas diferentemente do deslocamento, que sobrecarrega o termo independente, ela influencia todos os β_0, β_i .

4.4 Criação dos modelos com um conjunto fixo de variáveis

Utilizaremos a proc logistic para modelar a variável-resposta Realfac em função das variáveis de perfis.

Ajustaremos primeiramente um modelo equivalente ao modelo utilizado na análise discriminante.

```

proc logistic data=p1 descending;
  model REALFAC=grco phin drdn nplt;
run;

```

O que fornece a seguinte saída :

```

The LOGISTIC Procedure

Data Set: WORK.P1
Response Variable: REALFAC   Facies Padrao
Response Levels: 2
Number of Observations: 271
Link Function: Logit

Response Profile

Ordered
Value   REALFAC   Count
1         1       217
2         0        54

WARNING: 33 observation(s) were deleted due to missing values for the response or explanatory
variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion          Intercept
                   Only
AIC                272.663
SC                 276.265
-2 LOG L          270.663
Score              .

Intercept
and
Covariates
Chi-Square for Covariates
59.780
77.791
49.780
220.883 with 4 DF (p=0.0001)
206.059 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable DF   Parameter Estimate   Standard Error   Wald Chi-Square   Pr > Chi-Square   Standardized Estimate   Odds Ratio Variable Label
INTERCPT 1    -69.5193              287.0           0.0587           0.8086           .                      .                      Intercept
GRCO      1    -0.00446             0.0321          0.0193           0.8895          -0.039001             0.996 Raioes-Gama
PHIN      1     0.2106             0.1380          2.3302           0.1269           0.392871             1.234 Poros. Neutrons
DRDN      1    -0.5457             11.0879         0.0024           0.9607          -0.523293             0.579 Diferenca Phin/Phin
NPLT      1     1.1272             5.0687          0.0495           0.8240           2.361498             3.087 Razao Phin Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 98.4%
Discordant = 1.2%
Tied = 0.4%
(11718 pairs)

Somers' D = 0.971
Gamma = 0.975
Tau-a = 0.311
c = 0.986

```

Na proc logistic usamos a opção descending porque esta proc, por *default*, ordena os valores da variável resposta em ordem crescente o que faria que estimássemos a probabilidade $P(Y=0|x_i)$, ou seja, a ocorrência de fâcies não-reservatório.

Neste ponto convém discutir alguns dos itens constantes da saída, pois serão úteis no decorrer da análise.

A estatística Score fornece uma estimativa da significância das variáveis independentes no modelo. Podemos verificar que o conjunto de variáveis escolhido é bastante significativo para o modelo ($p=0.0001$). O valor p pode ser interpretado como a probabilidade dos parâmetros estimados serem nulos ($\beta = 0$). De maneira semelhante, -2LogL dá uma estimativa da significância das variáveis independentes, baseada na função de verossimilhança. Também indica que as variáveis são significantes neste modelo.

A seguir são exibidos os parâmetros estimados e as estatísticas associadas a eles. A estimativa é realizada pela máxima verossimilhança e, caso seja verificada alguma combinação linear com outra variável da análise, a estimativa é nula ($\beta = 0$). De maneira semelhante, na coluna DF (graus de liberdade), quando se verifica dependência linear entre as variáveis, $DF=0$. O erro-padrão é uma estimativa do desvio-padrão do parâmetro estimado. A estatística de Wald é a estatística qui-quadrado para testar a hipótese de que o parâmetro estimado seja igual a zero. É expressa pelo quadrado da razão entre o parâmetro estimado e o erro-padrão. Um valor pequeno ($p<0.05$, por exemplo) indica que o parâmetro não é nulo. A probabilidade $\text{pr}>\text{chi-square}$ é a probabilidade de se obter uma estatística qui-quadrado maior, em valor absoluto, que o valor observado, dado que o parâmetro verdadeiro seja nulo.

Finalmente têm-se uma série de medidas para verificar-se a qualidade das estimativas efetuadas por este modelo tomando-se como base os valores observados, isto é, auxiliar na avaliação da qualidade preditiva do modelo. São valores computados com base no número de pares $(p, \hat{p})$ discordantes e concordantes. As estatísticas D e Gamma variam de -1 a 1 e, quanto maiores, melhor a qualidade preditiva do modelo. A estatística C varia de 0 a 1, 1 indicando a melhor capacidade preditiva e a Tau-a varia com o tamanho da amostra, até um máximo de 0,5 com valores tanto maiores quanto melhor for a qualidade do ajuste. No exemplo discutido, sob qualquer dos critérios acima, pode-se esperar previsões confiáveis a partir do modelo ajustado.

Voltando ao modelo : a partir, portanto, dos parâmetros estimados, ajustamos o seguinte modelo :

$$\text{logito}(\pi) = -69.5193 - 0.00446 * \text{GRCO} + 0.2106 * \text{PHIN} - 0.5457 * \text{DRDN} + 1.1272 * \text{NPLT}$$

Notemos que este modelo não é ainda o que procuramos. Na verdade desejamos obter a probabilidade de ocorrência de fácies reservatório, $P(Y = 1 | x_i)$. É preciso, portanto, inverter a função logito para a obtenção de π_i , a probabilidade estimada para um determinado vetor x_i . Digamos que este vetor x_i contém os seguintes valores : GRCO=47.90, PHIN=21.6031, DRDN=-1.92097 e NPLT=60.5896 (observação 21 do dataset p1). Logo, $\text{logito}(\pi_i) = 4.1615$ e a probabilidade estimada, π_i , será :

$$\pi_i = \frac{e^{4.1615}}{1 + e^{4.1615}} = 0.9847 = P(Y = 1 | x_i)$$

Ou seja, temos aproximadamente 98% de probabilidade de ocorrência de fácies reservatório para estas leituras de perfis.

Este cálculo foi apenas para ilustrar o método. Naturalmente, na prática o cálculo é imediato, bastando apenas estipular a opção `p=` na declaração Output da Proc Logistic.

4.4.1 Criação do Modelo Global

Com finalidade de obtermos um modelo que contenha informação tanto de um como de outro poço, vamos criar, agora, um novo dataset que denominaremos de global e que representa simplesmente a concatenação dos dois pré-existentes. A tabela 4.4 relaciona seu conteúdo :

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	304	3058.73	11.6882069	3035.00	3080.00
DLT	Sonico	257	81.0187549	7.7209340	58.9800000	102.9400000
RT	Resistividade	257	1.7247471	0.9249316	0.4600000	4.8300000
PHIN	Poros. Neutrons	304	19.3424862	5.0805854	3.4179000	36.6000000
RHOC	Densidade	304	2.3797286	0.0954146	2.1723000	2.6000000
EFAC	EletroFacies	304	0.6052632	0.4896000	0	1.0000000
PHIT	Poros. Total	257	20.0394942	4.2360700	7.5300000	30.5100000
GRCO	Raios-Gama	274	69.0425547	19.3842782	45.7300000	126.4900000
PHIE	Poros. Efetiva	257	14.9317062	5.9797490	0.2558000	27.7320000
REALFAC	Facies Padrao	304	0.5657895	0.4964701	0	1.0000000
NPLT	Razao Phin,Rhoc	304	58.6836750	4.7702793	44.8276226	64.9721430
DRDN	Diferenca Rhoc,Phin	304	-0.9579322	2.3052839	-4.1060000	5.7421667

Facies Padrao=0						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	132	3058.20	12.7592025	3035.00	3080.00
DLT	Sonico	85	82.2848235	6.8597394	64.3800000	102.9400000
RT	Resistividade	85	2.1229412	0.8418232	0.5900000	4.4600000
PHIN	Poros. Neutrons	132	21.4557379	4.9737893	10.8263000	36.6000000
RHOC	Densidade	132	2.4369742	0.0898082	2.2673000	2.6000000
EFAC	EletroFacies	132	0.0909091	0.2885750	0	1.0000000
PHIT	Poros. Total	85	20.1236471	3.7260687	13.0300000	29.6400000
GRCO	Raios-Gama	102	88.6592157	16.4607874	48.6100000	126.4900000
PHIE	Poros. Efetiva	85	9.4720753	4.5426519	0.2558000	21.0940000
NPLT	Razao Phin,Rhoc	132	54.8613766	4.7986290	44.8276226	63.9473067
DRDN	Diferenca Rhoc,Phin	132	0.8913975	2.3414841	-3.3545667	5.7421667

Facies Padrao=1						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	172	3059.14	10.8151278	3038.00	3079.20
DLT	Sonico	172	80.3930814	8.0584831	58.9800000	102.9400000
RT	Resistividade	172	1.5279651	0.9026983	0.4600000	4.8300000
PHIN	Poros. Neutrons	172	17.7206884	4.5489529	3.4179000	29.5300000
RHOC	Densidade	172	2.3357959	0.0740384	2.1723000	2.6311000
EFAC	EletroFacies	172	1.0000000	0	1.0000000	1.0000000
PHIT	Poros. Total	172	19.9979070	4.4764604	7.5300000	30.5100000
GRCO	Raios-Gama	172	57.4094186	8.6129934	45.7300000	28.5100000
PHIE	Poros. Efetiva	172	17.6297797	4.6104138	5.2070000	27.7320000
NPLT	Razao Phin,Rhoc	172	61.6170669	1.6576619	56.9295746	64.9721430
DRDN	Diferenca Rhoc,Phin	172	-2.3771853	0.7419378	-4.1060000	-0.1654331

Tabela 4-4 Estatísticas descritivas para o poço Global (geral e por fácies)

Seguindo os mesmos passos anteriores, chegamos a um modelo global de regressão. Este modelo, juntamente com os dois anteriores encontra-se relacionado na tabela 4.5.

Poço	Modelo
P1	$\text{logito}(\pi) = -69.5193 - 0.00446 * \text{GRCO} + 0.2106 * \text{PHIN} - 0.5457 * \text{DRDN} + 1.1272 * \text{NPLT}$
P2	$\text{logito}(\pi) = 237.2 - 0.0688 * \text{GRCO} - 0.0487 * \text{PHIN} - 10.4285 * \text{DRDN} - 4.1113 * \text{NPLT}$
Global	$\text{logito}(\pi) = 81.1490 - 0.0241 * \text{GRCO} + 0.0559 * \text{PHIN} - 5.3622 * \text{DRDN} - 1.4507 * \text{NPLT}$

Tabela 4-5 Modelos logísticos obtidos a partir do mesmo conjunto de variáveis da análise discriminante.

O procedimento seguido, de concatenar poços, não é muito recomendado. As curvas de perfis, por serem medidas indiretas de propriedades das rochas, estão às vezes muito sujeitas a influências externas, as chamadas variações ambientais. O mais desejável, no caso, seria, antes de mais nada, colocar os dados em uma mesma escala, isto é, padronizá-los. Na verdade, faríamos isto em uma situação real. Como o intuito deste trabalho, no entanto, é exemplificar a utilização da técnica, optamos por não padronizar os dados e preservar o seu sentido original. Apenas a título de ilustração, apresentaremos dois gráficos da curva de raios-gama do arquivo global (figura 4-4), o primeiro com os dados brutos e o segundo, com os dados padronizados. A curva de raios-gama foi escolhida por ser uma das mais sensíveis às variações ambientais. No primeiro gráfico pode-se notar uma ligeira elevação no valor médio da curva a partir aproximadamente da observação 300. Já no gráfico seguinte, com os valores padronizados, verificamos um padrão mais equilibrado com relação ao valor médio para esta curva, com uma ligeira elevação na média da primeira metade.

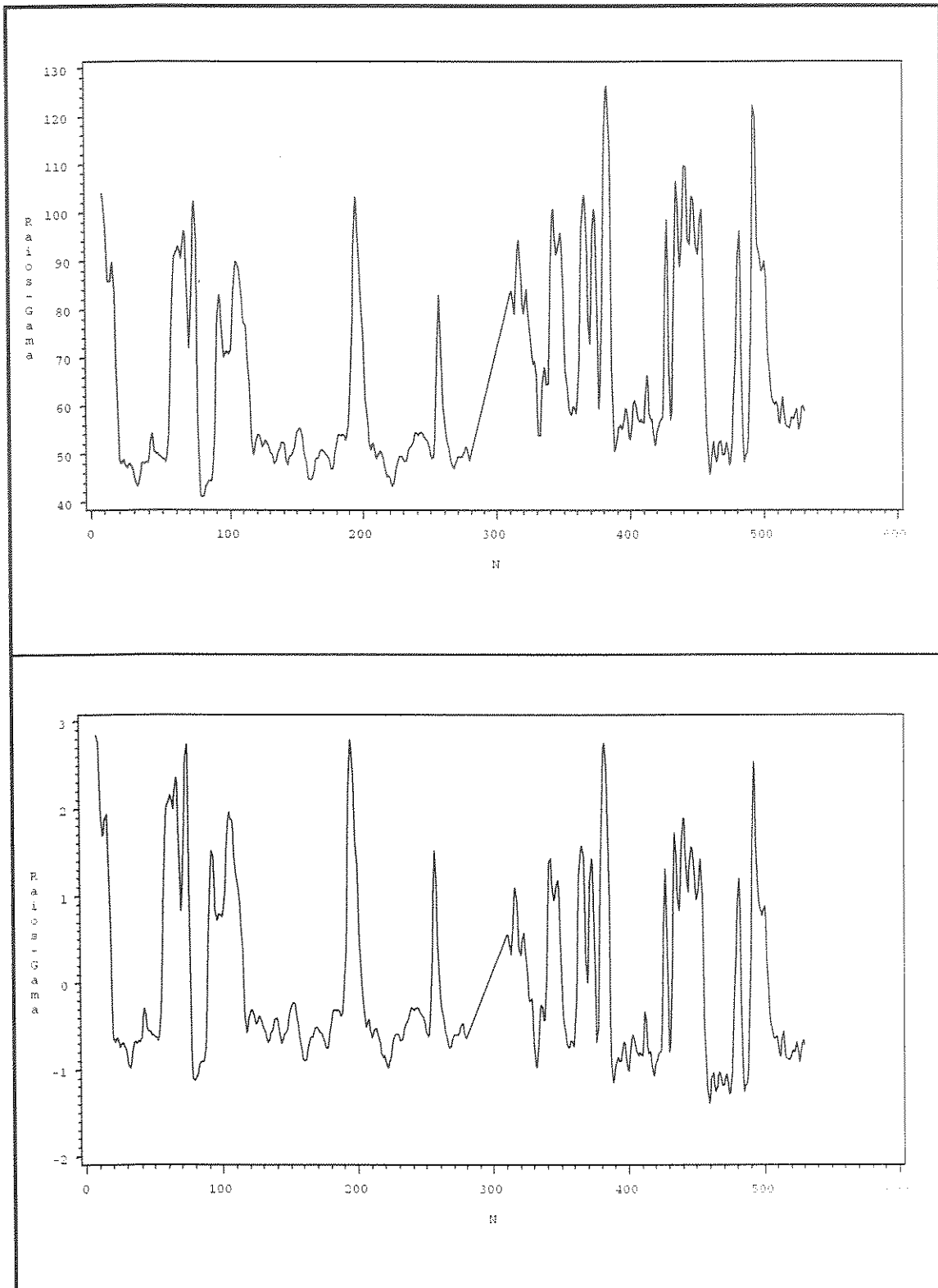


Figura 4-4 Gráfico do perfil de raios-gama para o poço global com os dados originais (acima) e padronizados (abaixo).

4.5 Geração dos modelos com escolha de variáveis

Até este momento trabalhamos com um conjunto fixo de variáveis pelo motivo já exposto. A partir de agora definiremos nossos próprios modelos logísticos, com base no conjunto original de variáveis. Como na regressão linear, um passo fundamental na análise é a escolha das variáveis que serão utilizadas no modelo. Uma técnica de seleção de variáveis versátil e bem implementada nos pacotes computacionais é a seleção stepwise. Trata-se de um método iterativo, onde as variáveis vão sendo checadas passo a passo. Primeiramente é estimado o termo independente do ajuste. A seguir, o conjunto de variáveis é averiguado e é escolhida a variável que exibe a maior importância relativa dentre as demais. No caso da regressão logística esta importância relativa é definida pelo seu escore qui-quadrado. Escolhida a variável, é estimado um parâmetro para ela e as demais variáveis são agora analisadas. De maneira similar, as variáveis vão entrando no modelo até que não exista mais nenhuma que atenda ao critério de seleção. O critério de seleção é basicamente a probabilidade obtida com o teste qui-quadrado. No SAS, a opção `SLENTRY=` define este valor de probabilidade. O valor default, de $p=0,05$ é muito restritivo, como já assinalaram Constanza & Afifi (1979, in Hosmer & Lemeshow, 1989). Estes autores sugerem a utilização de valores entre 0,15 e 0,20. Em situações especiais como onde não se conhece bem o relacionamento entre as variáveis, pode-se, por exemplo, em caráter prospectivo, adotar valores bem maiores, forçando a inclusão de variáveis no modelo para posteriormente refiná-lo. Adotaremos 0,20. O mesmo critério é adotado para verificar-se se uma variável deve ou não permanecer no modelo. São calculados os valores correspondentes de qui-quadrado para as variáveis constantes do modelo definido até o momento. Se alguma delas não satisfizer aos critérios de permanência no modelo, ela é automaticamente eliminada. No SAS, a opção `SLSTAY=` define o valor da probabilidade para o critério da eliminação. O algoritmo prossegue desta maneira até que não existam mais variáveis a serem incluídas e/ou eliminadas.

De maneira análoga à geração dos modelos anteriores, os comandos SAS

```
proc logistic data=p1 descending;
  model REALFAC= DLT RT PHIN RHOC GRCO NPLT DRDN/
    selection=stepwise
    slentry=0.2
    slstay=0.2
    details;
run;
```

produzem saídas de maneira geral idênticas às anteriores. A cada variável selecionada são calculadas as informações relativas ao ajuste do modelo e à análise dos parâmetros. Adicionalmente é feita uma análise das variáveis não constantes do modelo, com base na qual será escolhida, caso atenda aos critérios, a próxima variável.

Os modelos obtidos foram os seguintes (tabela 4.6):

Poço	Modelo
P1	$\text{logito}(\pi) = -6,3776 + 2,2014 \cdot \text{RT} - 5,8800 \cdot \text{DRDN}$
P2	$\text{logito}(\pi) = -169,2 - 0,1513 \cdot \text{GRCO} + 3,1450 \cdot \text{NPLT}$
Global	$\text{logito}(\pi) = 395,7 + 3,9314 \cdot \text{RT} - 146,0 \cdot \text{RHOC} - 0,1795 \cdot \text{GRCO} - 3,6386 \cdot \text{DRDN} - 2,1787 \cdot \text{PHIT}$

Tabela 4-6 Modelos logísticos obtidos a partir do método stepwise de seleção de variáveis.

Maiores detalhes dos ajustes e das estatísticas calculadas podem ser obtidas das saídas dos programas (apêndice 2).

Neste ponto convém fazer alguns comentários a respeito do conjunto de dados utilizado. Em uma situação real, não dispomos de um conjunto completo como o que estamos utilizando. As informações de rocha, principalmente, costumam ser bastante escassas. A grande maioria dos poços, geralmente, dispõem apenas de informações de perfis. As informações de rocha são escassas, existem muito poucos poços testemunhados. Daí a importância de se ajustar com o melhor critério possível um modelo a estes poços-chave. Eles

serão utilizados para prever-se a ocorrência de fácies reservatório nos demais poços da área estudada.

A título de ilustração, geraremos uma curva de fácies estimada pelo modelo do poço P1 no poço P2 e a compararemos, tão bem como às fácies preditas pelo modelo logístico para este poço à fácies padrão, Realfac. As classificações das duas em relação à fácies padrão estão mostradas na tabela 4.7.

Realfac	Modelo de P1		Modelo de P2		Modelo Global	
	0	1	0	1	0	1
0	94	4	92	6	94	4
1	4	124	2	126	3	125
Totais	98	128	94	132	97	129

Tabela 4-7 Tabela de classificação dos resultados obtidos pelos modelos logísticos construídos com base nas variáveis utilizadas pela análise discriminante, comparados aos valores da fácies-padrão (Realfac).

Deve-se ressaltar que este conjunto de dados mostra uma situação praticamente ideal. Os perfis exibem grande contraste entre as fácies reservatório e não-reservatório levando-nos a um percentual de acertos anormalmente elevado. No entanto, embora os percentuais de acerto globais não tenham se modificado, nota-se uma muito sutil melhora na classificação obtida com o modelo do próprio poço, com dois pontos a mais sendo reclassificados como reservatório, em relação à fácies padrão. Isto já era de se esperar, pois em princípio as melhores estimativas são sempre obtidas em modelos estimados a partir dos próprios dados.

Continuando por esta linha de raciocínio vamos agora verificar o comportamento das estimativas obtidas a partir do modelo global. Para efeito de comparação vamos estimar as fácies também para o poço P2. De maneira semelhante, verificamos (tabela 4.7) uma melhora muito sutil em relação à estimativa feita a partir do modelo do poço P1, com um ponto a mais sendo classificado como fácies reservatório pela estimativa a partir do modelo global.

Os dados para as estimativas efetuadas a partir dos modelos ajustados com seleção de variáveis stepwise encontram-se relacionados na tabela 4.8.

Realfac	P1 Próprio versus P2		P2 Próprio versus P2		Global Próprio versus P2	
	0	1	0	1	0	1
0	93	5	93	5	98	0
1	1	127	2	126	1	127
Totais	94	132	95	131	99	127

Tabela 4-8 Tabela de classificação dos resultados obtidos pelos modelos logísticos com seleção de variáveis comparados aos valores da fácies-padrão (Realfac).

4.6 Comentários sobre a avaliação dos modelos ajustados

Os primeiros três modelos ajustados (tabela de modelos com variáveis fixas) em função da comparação com os resultados obtidos pela análise discriminante que será feita adiante tiveram fixado um determinado conjunto de variáveis. Assim, modelos que de outra forma seriam descartados, como por exemplo o ajustado para o poço P1 estão sendo considerados nesta análise. Se admitirmos como valor de corte $p=0.05$, no caso do poço P1, nenhuma das variáveis deveria integrar o modelo. Mesmo para o critério mais brando de $p=0.20$ sugerido por Constanza & Afifi (op. cit.), apenas a curva de porosidade neutrônica permaneceria no modelo. De maneira similar, o mesmo ocorre para os conjuntos P2 e Global, embora, para o primeiro, de maneira mais atenuada. Pelo critério $p=0.20$, apenas PHIN sairia. Não obstante, todos modelos alcançaram uma proporção muito elevada de valores preditos concordantes. Isto se deve, em especial, às características dos dados, como já assinalado anteriormente.

A observação dos modelos obtidos pela seleção stepwise ilustra bem isto. As variáveis primeiramente selecionadas para integrar os modelos foram as compostas DRDN e NPLT.

Dentre todas as variáveis consideradas são as que obtiveram escore qui-quadrado mais elevado. Isto é reflexo não apenas do fato de serem elas variáveis que têm um alto valor agregado derivado de duas importantes variáveis de elevado teor de informação litológica. O modelo ajustado para o poço P2, por exemplo, tem uma forte componente de informação litológica. As variáveis selecionadas, NPLT e GRCO estão diretamente relacionadas a propriedades físicas das rochas em questão. NPLT, composta por PHIN, porosidade neutrônica e RHOC, a densidade da rocha, se associam, neste caso, a GRCO, uma leitura da radioatividade natural das rochas. O poço P1, embora exiba também uma forte componente litológica, expressa pela presença da variável DRDN, mostra uma dependência interessante evidenciada pela presença de RT. RT, a resistividade do meio, está relacionada ao fluido presente no sistema. Em caso de hidrocarbonetos, meios não-condutivos de eletricidade, mostram valores elevados. Os dados que originam o conjunto P2 são provenientes de um poço exploratório da época do descobrimento do campo. A coluna geológica perfurada foi mais extensa e o fluido presente em grande parte dela foi a água, salgada, boa condutora e, conseqüentemente, com baixos valores associados de resistividade. Já os dados que deram origem ao conjunto P1 são oriundos de um poço de desenvolvimento do campo, onde os objetivos já encontravam-se bem definidos, mostrando uma maior proporção da coluna geológica com hidrocarbonetos, com um padrão mais elevado de valores de RT. Neste contexto, onde existe rocha reservatório, pela presença de óleo, a curva RT costuma ser fortemente associada à ocorrência de fácies reservatório.

4.7 Comparação com os resultados obtidos pela análise discriminante

Os resultados obtidos pelos modelos logísticos desenvolvidos a partir do mesmo conjunto de variáveis da análise discriminante estão relacionados na tabela 4-9. Em geral a análise discriminante, por este critério, foi mais precisa, pois para a ocorrência de fácies reservatório (Realfac=1), todas as suas estimativas foram classificadas corretamente. Embora o erro verificado nas estimativas efetuadas pelos modelos logísticos sejam desprezíveis do ponto de vista prático (3 observações para o poço P1 e 2 para o poço P2), devemos considerá-lo para fins de comparação. Assinala-se que, para o poço P1, o índice de acertos dos modelos

logísticos na estimativa de fácies não-reservatório foi bem maior que o obtido pela análise discriminante.

	Poço P1				Poço P2			
	Efac		Facies		Efac		Facies	
Realfac	0	1	0	1	0	1	0	1
0	71	16	82	5	92	6	92	6
1	0	217	3	214	0	128	2	126

Tabela 4-9 Tabela de classificação dos resultados obtidos pela análise discriminante (efac) e pela regressão logística (facies).

Comparando-se os modelos logísticos construídos com escolha de variáveis (tabela 4-10) pode-se verificar que houve uma ligeira melhora na estimativa global, com uma observação a mais sendo classificada corretamente, para a fácies não-reservatório, em cada poço. Para a fácies reservatório, a situação manteve-se inalterada em relação aos resultados alcançados pelos modelos de conjunto fixo de variáveis. Consequentemente, para este conjunto de dados, pode-se concluir que é mais recomendável a utilização dos modelos gerados com escolha de variáveis.

	Poço P1				Poço P2			
	Efac		Facies		Efac		Facies	
Realfac	0	1	0	1	0	1	0	1
0	71	16	83	4	92	6	93	5
1	0	217	3	214	0	128	2	126

Tabela 4-10 Tabela de classificação dos resultados obtidos pela análise discriminante (efac) e pela regressão logística com escolha de variáveis (facies).

Efron (1975) em uma comparação teórica entre os dois métodos chega à conclusão que a regressão logística é menos eficiente que a análise discriminante. Entretanto considera que sua utilização é justificada pelo fato dela ser, pelo menos em teoria, mais robusta que a discriminação normal.

5. Conclusões

- Foram utilizados nas análises dados correspondentes apenas ao reservatório principal do campo. Desta forma Os resultados obtidos podem ter sido em parte mascarados devido à evidente seletividade da amostragem. Uma demonstração clara disto é o elevado percentual de acertos verificado com a utilização de qualquer das técnicas, fato de difícil constatação no dia a dia do intérprete.
- A utilização dos modelos logísticos de regressão mostrou ser teórica e praticamente adequada à solução de problemas de correlação rocha-perfil envolvendo a estimativa de fácies a partir de perfis.
- As particularidades da aquisição dos dados de perfis e o efeito que o ambiente do poço exerce neles, podem fortemente influenciar as estimativas de parâmetros para os modelos. Há sempre, portanto, a necessidade de identificar e corrigir, caso possível, os deslocamentos e mudanças de escala porventura verificados.
- Os modelos obtidos a partir da seleção stepwise de variáveis mostraram-se ligeiramente melhores que os obtidos a partir do conjunto fixo de variáveis.
- A análise discriminante mostrou-se algo mais precisa que a regressão logística, para este caso particular, embora deva-se considerar que os erros de estimativa verificados na regressão logística foram, do ponto de vista prático, desprezíveis.

Apêndice 1 - Listagens dos programas SAS

```
libname pocos 'c:\usuario\wagner\tese\dados';
```

```
data p1;  
set pocos.p1;  
proc means;
```

```
data p2;  
set pocos.p2;  
proc means;
```

```
run;
```

```
/*
```

```
modelos logicos com mesmas variaveis da  
discriminante - pocos p1 e p2
```

```
*/
```

```
proc logistic data=p1 descending;  
model reafac=grco phin drdn nplt;  
run;
```

```
proc logistic data=p2 descending;  
model reafac=grco phin drdn nplt;  
run;
```

```
/*
```

```
criacao do dataset global (com os 2 pocos)  
modelo logistico (=discrim) global
```

```
*/
```

```
data global(drop=prof);  
set p1 p2;  
proc means;
```

```
proc logistic data=global descending;  
model reafac=grco phin drdn nplt;  
run;
```

```
/*
```

```
comparacao dos dados brutos e padronizados
```

```
*/
```

```
symbol i=join v=none;
```

```
data um;  
set global;
```

```

n=_n_;
run;

proc gplot data=um;
plot grco*n/frame;
run;

proc standard data=p1 out=p1std mean=0 std=1;
var dlt rt phin rhoc efac phit
    grco phie realfac nplt drdn;

proc standard data=p2 out=p2std mean=0 std=1;
var dlt rt phin rhoc efac phit
    grco phie realfac nplt drdn;

data globstd;
set p1std p2std;
run;

data um;
set globstd;
n=_n_;

proc gplot data=um;
plot grco*n/frame;
quit;
run;
/*
*/

/*
Criacao do modelo proprio
selecao de variaveis stepwise

Inclusao de phit e phie
*/

proc logistic data=p1 descending;
model realfac= dlt rt phin rhoc grco nplt drdn phit phie/
    selection=stepwise
    slentry=0.2
    slstay=0.2
    details;
proc logistic data=p2 descending;
model realfac= dlt rt phin rhoc grco nplt drdn phit phie/
    selection=stepwise

```

```

    slentry=0.2
    slstay=0.2
    details;
proc logistic data=global descending;
  model realfac= dlt rt phin rhoc grco nplt drdn phit phie/
    selection=stepwise
    slentry=0.2
    slstay=0.2
    details;
run;

/*
  Criacao da facies estimada pelo modelo do poço P1
  no poco P2.
*/

data p;
title 'Facies obtidas pelo modelo de P1';
set p2;
logitpi=-69.5193-0.00446*grco+0.2106*phin-0.5457*drdn+1.1272*nplt;
prob=exp(logitpi)/(1+exp(logitpi));
if prob>=.5 then facp=1; else facp=0;
label logitpi='Logito'
    prob='Prob. estim.'
    facp='fac estimada por P1';
proc freq data=p;
table realfac*facp;
run;

/*
  Comparacao entre as facies estimadas pelo modelo do poco p1
  e a estimada pela regressao logistica no poço p2.
*/

proc logistic data=p2 descending noprint;
  model realfac=grco phin drdn nplt;
  output out=outp2 p=predp2;
run;

data complp2;
merge outp2 p;
by prof;
if predp2>=.5 then faclog=1; else faclog=0;
label predp2='Prob pred'

```

```
run;
```

```
/*  
  Criacao da facies estimada pelo modelo global  
  no poço P2.  
*/
```

```
data p;  
title 'Facies Estimada pelo Modelo Global';  
set p2;  
logitpi=395.7+3.9314*rt-146.0*rhoc-0.1795*grco-3.6386*drdn-2.1787*phit;  
probg=exp(logitpi)/(1+exp(logitpi));  
if probg>=.5 then facg=1; else facg=0;  
label logitpi='Logito'  
      probg='Prob estim. Global'  
      facg='fac estimada global';  
proc freq data=p;  
table realfac*facg;  
run;
```

```
/*  
  Comparacao entre as facies estimadas pelo modelo global  
  e a estimada pela regressao logistica no poço p2.  
*/
```

```
proc logistic data=p2 descending noprint;  
model realfac=grco phin drdn nplt;  
output out=outg p=predg;  
run;
```

```
data comp1p2;  
merge outg p;  
by prof;  
if predg>=.5 then faclog=1; else faclog=0;  
label predp2='Prob pred'  
      faclog='Facies estim logistica';  
proc freq data=comp1p2;  
table realfac*faclog;  
run;
```

```
/*  
  Criacao da curva de facies estimada  
  pelo modelo p1 proprio no poço P2.
```

```

/*
  Criacao da curva de facies estimada
  pelo modelo p1 proprio no poco P2.
*/

data p;
title 'P1 proprio x P2';
set p2;
logitpi=-6.3776+2.2014*rt-5.8800*drdn;
prob=exp(logitpi)/(1+exp(logitpi));
if prob>=.5 then fac=1; else fac=0;
label logitpi='Logito'
      prob='Prob calculada'
      fac='facies calculada';
proc freq data=p;
table realfac*fac;
run;

/*
  Criacao da curva de facies estimada
  pelo modelo p2 proprio no poco P2.
*/

data p;
title 'P2 proprio x P2';
set p2;
logitpi=-169.2-0.1513*grco+3.1450*nplt;
prob=exp(logitpi)/(1+exp(logitpi));
if prob>=.5 then fac=1; else fac=0;
label logitpi='Logito'
      prob='Prob calculada'
      fac='facies calculada';
proc freq data=p;
table realfac*fac;
run;

/*
  Criacao da curva de facies estimada
  pelo modelo global proprio no poco P2.
*/

data p;

```

```

title 'Global proprio x P2';
set p2;
logitpi=120.3-0.2733*dlt+4.1180*rt-39.9990*rhoc-0.1910*grco-6.5924*drdn;
prob=exp(logitpi)/(1+exp(logitpi));
if prob>=.5 then fac=1; else fac=0;
label logitpi='Logito'
      prob='Prob calculada'
      fac='facies calculada';
proc freq data=p;
table realfac*fac;
run;

```

```

/*
  Comparacao entre as facies estimadas pela regressao
  logistica(modelos com escolha de variaveis) e as
  estimadas pela analise discriminante
*/

```

```

/*
  Poco P1
*/

```

```

data p1disc;
set p1;
logitpi=-6.3776+2.2014*rt-5.88*drdn;
prob=exp(logitpi)/(1+exp(logitpi));
if prob>=0.5 then facies=1; else facies=0;
label facies='Facies Est. Logist.';
proc freq data=p1disc;
table realfac*(facies efac);
run;

```

```

/*
  Poco P2
*/

```

```

data p2disc;
set p2;
logitpi=-169.2-0.1513*grco+3.145*nplt;
prob=exp(logitpi)/(1+exp(logitpi));
if prob>=0.5 then facies=1; else facies=0;
label facies='Facies Est. Logist.';
proc freq data=p2disc;
table realfac*(facies efac);

```

```

run;

/*
  global x Poco P1
*/
data glp1;
set p1;
  logitpi=120.3-0.2733*dlt+4.118*rt-39.999*rhoc-0.1910*grco-6.5924*drdn;
  prob=exp(logitpi)/(1+exp(logitpi));
  if prob>=0.5 then facies=1; else facies=0;
label facies='Facies Est. Logist.';
proc freq data=glp1;
table realfac*(facies efac);
run;

/*
  global x Poco P2
*/

data glp2;
set p2;
  logitpi=120.3-0.2733*dlt+4.118*rt-39.999*rhoc-0.1910*grco-6.5924*drdn;
  prob=exp(logitpi)/(1+exp(logitpi));
  if prob>=0.5 then facies=1; else facies=0;
label facies='Facies Est. Logist.';
proc freq data=glp1;
table realfac*(facies efac);
run;

/*
*/

```

Apêndice 2 - Saidas dos programas SAS

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	304	3039.66	17.6562667	3005.00	3070.00
DLT	Sonico	265	81.0644151	4.9219183	66.0100000	94.8400000
RT	Resistividade	265	2.4429811	1.0464553	0.5700000	4.7700000
PHIN	Poros. Neutrons	301	20.6455934	3.3785140	10.7588000	29.4786000
RHOC	Densidade	301	2.3191385	0.0808332	2.0403000	2.5481000
EFAC	EletroFacies	304	0.7664474	0.4237884	0	1.0000000
PHIT	Poros. Total	265	22.1377736	3.0119802	12.4200000	31.7600000
GRCO	Raios-Gama	271	58.8451292	15.8599650	40.9600000	104.0700000
PHIE	Poros. Efetiva	265	18.8329506	5.1240782	0	31.7648000
REALFAC	Facies Padrao	304	0.7138158	0.4527218	0	1.0000000
NPLT	Razao Phin,Rhoc	301	60.3386204	3.8076038	49.9226482	74.1606267
DRDN	Diferenca Rhoc,Phin	301	-1.7353648	1.7534231	-6.5771000	3.3641333

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PROF	Profundidade	226	3057.50	13.0769517	3035.00	3080.00
DLT	Sonico	207	81.8788406	8.0810368	58.9800000	102.9400000
RT	Resistividade	207	1.9623671	0.8763099	0.4600000	4.8300000
PHIN	Poros. Neutrons	226	19.8721429	5.4975655	3.4179000	36.6002000
RHOC	Densidade	226	2.3937779	0.1016543	2.1723000	2.6002000
EFAC	EletroFacies	226	0.5929204	0.4923805	0	1.0000000
PHIT	Poros. Total	207	20.1040097	4.4806272	7.5300000	30.5100000
GRCO	Raios-Gama	221	72.8249321	19.4720205	45.7300000	126.4900000
PHIE	Poros. Efetiva	207	14.2942058	6.3452418	0.2558000	27.7328000
REALFAC	Facies Padrao	226	0.5663717	0.4966753	0	1.0000000
NPLT	Razao Phin,Rhoc	226	57.7380960	5.0791896	44.8276886	64.9721430
DRDN	Diferenca Rhoc,Phin	226	-0.5003948	2.4634071	-4.1060000	5.7421667

The LOGISTIC Procedure

Data Set: WORK.P1
 Response Variable: REALFAC Facies Padrao
 Response Levels: 2
 Number of Observations: 271
 Link Function: Logit

Response Profile

Ordered Value	REALFAC	Count
1	1	217
2	0	54

WARNING: 33 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	272.663	59.780	.
SC	276.265	77.791	.
-2 LOG L	270.663	49.780	220.883 with 4 DF (p=0.0001)
Score	.	.	206.059 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	-69.5193	287.0	0.0587	0.8086	.	.	Intercept
GRCO	1	-0.00446	0.0321	0.0193	0.8895	-0.039001	0.996	Raios-Gama
PHIN	1	0.2106	0.1380	2.3302	0.1269	0.392871	1.234	Poros. Neutrons
DRDN	1	-0.5457	11.0879	0.0024	0.9607	-0.523893	0.579	Diferenca Rhoc,Phin
NPLT	1	1.1272	5.0687	0.0495	0.8240	2.361498	3.087	Razao Phin,Rhoc

Association of Predicted Probabilities and Observed Responses

Concordant = 98.4%	Somers' D = 0.971
Discordant = 1.2%	Gamma = 0.975
Tied = 0.4%	Tau-a = 0.311
(11718 pairs)	c = 0.986

The LOGISTIC Procedure

Data Set: WORK.P2
 Response Variable: REALFAC Facies Padrao
 Response Levels: 2
 Number of Observations: 221
 Link Function: Logit

Response Profile

Ordered Value	REALFAC	Count
1	1	128
2	0	93

WARNING: 5 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	302.805	73.616	.
SC	306.203	90.607	.
-2 LOG L	300.805	63.616	237.189 with 4 DF (p=0.0001)
Score	.	.	163.005 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	237.2	140.3	2.8565	0.0910	.	.	Intercept
GRCO	1	-0.0688	0.0372	3.4218	0.0643	-0.738567	0.934	Raios-Gama
PHIN	1	-0.0487	0.0801	0.3693	0.5434	-0.148876	0.953	Poros. Neutrons
DRDN	1	-10.4285	5.3610	3.7841	0.0517	-14.024252	0.000	Diferenca Rhoc, Phin
NPLT	1	-4.1113	2.4656	2.7804	0.0954	-11.419995	0.016	Razao Phin, Rhoc

Association of Predicted Probabilities and Observed Responses

Concordant = 96.8%	Somers' D = 0.938
Discordant = 3.0%	Gamma = 0.940
Tied = 0.2%	Tau-a = 0.459
(11904 pairs)	c = 0.969

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
DLT	Sonico	472	81.4215890	6.5041241	58.9800000	102.9400000
RT	Resistividade	472	2.2322034	1.0033222	0.4600000	4.8300000
PHIN	Poros. Neutrons	527	20.3139049	4.4255035	3.4179000	36.6002000
RHOC	Densidade	527	2.3511471	0.0975399	2.0403000	2.6002000
EFAC	EletroFacies	530	0.6924528	0.4619139	0	1.0000000
PHIT	Poros. Total	472	21.2458475	3.8582572	7.5300000	31.7600000
GRCO	Raios-Gama	492	65.1246748	18.8854801	40.9600000	126.4900000
PHIE	Poros. Efetiva	472	16.8424417	6.1163074	0	31.7648000
REALFAC	Facies Padrao	530	0.6509434	0.4771222	0	1.0000000
NPLT	Razao Phin,Rhoc	527	59.2234050	4.5786090	44.8276886	74.1606267
DRDN	Diferenca Rhoc,Phin	527	-1.2057572	2.1733789	-6.5771000	5.7421667

The LOGISTIC Procedure

Data Set: WORK.GLOBAL
 Response Variable: REALFAC Facies Padrao
 Response Levels: 2
 Number of Observations: 492
 Link Function: Logit

Response Profile

Ordered Value	REALFAC	Count
1	1	345
2	0	147

WARNING: 38 observation(s) were deleted due to missing values for the response or explanatory variables.

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	602.070	130.021	.
SC	606.269	151.014	.
-2 LOG L	600.070	120.021	480.049 with 4 DF (p=0.0001)
Score	.	.	370.739 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	81.1490	119.4	0.4618	0.4968	.	.	Intercept
GRCO	1	-0.0241	0.0214	1.2585	0.2619	-0.250511	0.976	Raios-Gama
PHIN	1	0.0559	0.0640	0.7637	0.3822	0.139163	1.058	Poros. Neutrons
DRDN	1	-5.3622	4.6141	1.3506	0.2452	-6.392689	0.005	Diferenca Rhoc,Phin
NPLT	1	-1.4507	2.1077	0.4738	0.4913	-3.651364	0.234	Razao Phin,Rhoc

Association of Predicted Probabilities and Observed Responses

Concordant = 97.3%	Somers' D = 0.950
Discordant = 2.3%	Gamma = 0.954
Tied = 0.4%	Tau-a = 0.399
(50715 pairs)	c = 0.975

The LOGISTIC Procedure

Data Set: WORK.F1
 Response Variable: REALFAC Facies Padrao
 Response Levels: 2
 Number of Observations: 265
 Link Function: Logit

Response Profile

Ordered Value	REALFAC	Count
1	1	217
2	0	48

WARNING: 39 observation(s) were deleted due to missing values for the response or explanatory variables.

Stepwise Selection Procedure

Step 0. Intercept entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	1.5087	0.1595	89.4662	0.0001	.	.	Intercept

Residual Chi-Square = 215.4607 with 8 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	5.4343	0.0197	Sonico
RT	6.5298	0.0106	Resistividade
PHIN	49.8770	0.0001	Poros. Neutrons
RHOC	112.8083	0.0001	Densidade
GRCO	138.1709	0.0001	Raios-Gama
NPLT	195.6231	0.0001	Razao Phin,Rhoc
DRDN	202.3034	0.0001	Diferenca Rhoc,Phin
PHIT	6.7432	0.0094	Poros. Total
PHIE	117.0622	0.0001	Poros. Efetiva

The LOGISTIC Procedure

Step 1. Variable DRDN entered:

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	252.746	27.723	.
SC	256.326	34.883	.
-2 LOG L	250.746	23.723	227.023 with 1 DF (p=0.0001)
Score	.	.	202.303 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Variable Ratio Label
INTERCPT	1	-1.8708	0.7368	6.4476	0.0111	.	. Intercept
DRDN	1	-4.3832	1.1623	14.2218	0.0002	-4.136737	0.012 Diferenca Rhoc,Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 99.8%	Somers' D = 0.996
Discordant = 0.2%	Gamma = 0.996
Tied = 0.0%	Tau-a = 0.296
(10416 pairs)	c = 0.998

Residual Chi-Square = 10.3771 with 7 DF (p=0.1682)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.6313	0.4269	Sonico
RT	5.1011	0.0239	Resistividade
PHIN	0.0033	0.9544	Poros. Neutrons
RHOC	0.0033	0.9544	Densidade
GRCO	0.0949	0.7581	Raios-Gama
NPLT	0.0802	0.7770	Razao Phin,Rhoc
PHIT	0.0277	0.8677	Poros. Total
PHIE	0.0019	0.9652	Poros. Efetiva

Step 2. Variable RT entered:

The LOGISTIC Procedure

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	252.746	23.730	.
SC	256.326	34.469	.
-2 LOG L	250.746	17.730	233.016 with 2 DF (p=0.0001)
Score	.	.	202.789 with 2 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	-6.3776	2.8829	4.8939	0.0270	.	.	Intercept
RT	1	2.2014	1.1677	3.5544	0.0594	1.270076	9.038	Resistividade
DRDN	1	-5.8800	2.0334	8.3617	0.0038	-5.549350	0.003	Diferenca Rhoc,Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 99.9%	Somers' D = 0.997
Discordant = 0.1%	Gamma = 0.997
Tied = 0.0%	Tau-a = 0.297
(10416 pairs)	c = 0.999

Residual Chi-Square = 8.0769 with 6 DF (p=0.2325)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.4137	0.5201	Sonico
PHIN	0.2270	0.6338	Poros. Neutrons
RHOC	0.2270	0.6338	Densidade
GRCO	1.4887	0.2224	Raios-Gama
NPLT	1.0628	0.3026	Razao Phin,Rhoc
PHIT	0.0647	0.7993	Poros. Total
PHIE	0.2217	0.6377	Poros. Efetiva

NOTE: No (additional) variables met the 0.2 significance level for entry into the model.

The LOGISTIC Procedure

Summary of Stepwise Procedure

Step	Variable		Number In	Score	Wald	Pr >	Variable Label
	Entered	Removed		Chi-Square	Chi-Square	Chi-Square	
1	DRDN		1	202.3	.	0.0001	Diferenca Rhoc, Phin
2	RT		2	5.1011	.	0.0239	Resistividade

The LOGISTIC Procedure

Data Set: WORK.P2
 Response Variable: REALFAC Facies Padrao
 Response Levels: 2
 Number of Observations: 207
 Link Function: Logit

Response Profile

Ordered Value	REALFAC	Count
1	1	128
2	0	79

WARNING: 19 observation(s) were deleted due to missing values for the response or explanatory variables.

Stepwise Selection Procedure

Step 0. Intercept entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	0.4826	0.1431	11.3765	0.0007	.	.	Intercept

Residual Chi-Square = 186.6700 with 8 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.0562	0.8125	Sonico
RT	12.5591	0.0004	Resistividade
PHIN	45.3753	0.0001	Poros. Neutrons
RHOC	87.1574	0.0001	Densidade
GRCO	143.8442	0.0001	Raios-Gama
NPLT	160.3767	0.0001	Razao Phin,Rhoc
DRDN	160.0018	0.0001	Diferenca Rhoc,Phin
PHIT	0.5499	0.4584	Poros. Total
PHIE	95.9538	0.0001	Poros. Efetiva

The LOGISTIC Procedure

Step 1. Variable NPLT entered:

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	277.253	21.002	.
SC	280.586	27.668	.
-2 LOG L	275.253	17.002	258.251 with 1 DF (p=0.0001)
Score	.	.	160.377 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	-163.7	52.8311	9.5958	0.0020	.	.	Intercept
NPLT	1	2.8498	0.9206	9.5823	0.0020	7.929116	17.285	Razao Phin, Rhoc

Association of Predicted Probabilities and Observed Responses

Concordant = 99.9%	Somers' D = 0.998
Discordant = 0.1%	Gamma = 0.998
Tied = 0.0%	Tau-a = 0.473
(10112 pairs)	c = 0.999

Residual Chi-Square = 6.7152 with 7 DF (p=0.4591)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.0647	0.7992	Sonico
RT	0.0143	0.9049	Resistividade
PHIN	0.9136	0.3391	Poros. Neutrons
RHOC	0.9131	0.3393	Densidade
GRCO	5.1316	0.0235	Raios-Gama
DRDN	0.8317	0.3618	Diferenca Rhoc, Phin
PHIT	0.5040	0.4778	Poros. Total
PHIE	0.8053	0.3695	Poros. Efetiva

Step 2. Variable GRCO entered:

The LOGISTIC Procedure

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	277.253	16.569	.
SC	280.586	26.567	.
-2 LOG L	275.253	10.569	264.684 with 2 DF (p=0.0001)
Score	.	.	164.492 with 2 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	-169.2	72.8595	5.3926	0.0202	.	.	Intercept
GRCO	1	-0.1513	0.0808	3.5032	0.0613	-1.651187	0.860	Raios-Gama
NPLT	1	3.1450	1.3091	5.7719	0.0163	8.750306	23.219	Razao Phin,Rhoc

Association of Predicted Probabilities and Observed Responses

Concordant	=100.0%	Somers' D	= 0.999
Discordant	= 0.0%	Gamma	= 0.999
Tied	= 0.0%	Tau-a	= 0.474
(10112 pairs)		c	= 1.000

Residual Chi-Square = 4.4289 with 7 DF (p=0.7293)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.3228	0.5699	Sonico
RT	1.2766	0.2585	Resistividade
PHIN	0.0096	0.9220	Poros. Neutrons
RHOC	0.0086	0.9259	Densidade
DRDN	0.0002	0.9895	Diferenca Rhoc,Phin
PHIT	0.0127	0.9104	Poros. Total
PHIE	0.0069	0.9336	Poros. Efetiva

NOTE: No (additional) variables met the 0.2 significance level for entry into the model.

The LOGISTIC Procedure

Summary of Stepwise Procedure

Step	Variable		Number In	Score	Wald	Pr >	Variable Label
	Entered	Removed		Chi-Square	Chi-Square	Chi-Square	
1	NPLT		1	160.4	.	0.0001	Razao Phin,Rhoc
2	GRCO		2	5.1316	.	0.0235	Raios-Gama

The LOGISTIC Procedure

Data Set: WORK.GLOBAL
 Response Variable: REALFAC Facies Padrao
 Response Levels: 2
 Number of Observations: 472
 Link Function: Logit

Response Profile

Ordered Value	REALFAC	Count
1	1	345
2	0	127

WARNING: 58 observation(s) were deleted due to missing values for the response or explanatory variables.

Stepwise Selection Procedure

Step 0. Intercept entered:

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	0.9994	0.1038	92.7091	0.0001	.	.	Intercept

Residual Chi-Square = 385.7862 with 9 DF (p=0.0001)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	2.7675	0.0962	Sonico
RT	0.4229	0.5155	Resistividade
PHIN	81.5201	0.0001	Poros. Neutrons
RHOC	212.8958	0.0001	Densidade
GRCO	296.9105	0.0001	Raios-Gama
NPLT	363.5058	0.0001	Razao Phin,Rhoc
ORDN	367.1913	0.0001	Diferenca Rhoc,Phin
PHIT	10.3565	0.0013	Poros. Total
PHIE	226.5622	0.0001	Poros. Efetiva

The LOGISTIC Procedure

Step 1. Variable DRDN entered:

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	551.719	44.558	.
SC	555.876	52.872	.
-2 LOG L	549.719	40.558	509.161 with 1 DF (p=0.0001)
Score	.	.	367.191 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	-2.1855	0.5948	13.4996	0.0002	.	.	Intercept
DRDN	1	-5.0004	1.0207	23.9985	0.0001	-5.886060	0.007	Diferenca Rhoc,Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 99.9%	Somers' D = 0.997
Discordant = 0.1%	Gamma = 0.997
Tied = 0.0%	Tau-a = 0.393
(43815 pairs)	c = 0.999

Residual Chi-Square = 13.2247 with 7 DF (p=0.0668)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.3046	0.5810	Sonico
RT	2.1840	0.1395	Resistividade
PHIN	0.1538	0.6949	Poros. Neutrons
RHOC	0.1538	0.6949	Densidade
GRCO	2.9167	0.0877	Raios-Gama
NPLT	0.2431	0.6220	Razao Phin,Rhoc
PHIT	0.0156	0.9006	Poros. Total
PHIE	0.1148	0.7348	Poros. Efetiva

Step 2. Variable GRCO entered:

The LOGISTIC Procedure

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	551.719	43.518	.
SC	555.876	55.989	.
-2 LOG L	549.719	37.518	512.201 with 2 DF (p=0.0001)
Score	.	.	375.846 with 2 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	1.8979	2.4531	0.5986	0.4391	.	.	Intercept
GRCO	1	-0.0544	0.0328	2.7584	0.0967	-0.562546	0.947	Raios-Gama
DRDN	1	-4.8533	1.1179	18.8490	0.0001	-5.712911	0.008	Diferenca Rhoc,Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 99.9%	Somers' D = 0.998
Discordant = 0.1%	Gamma = 0.998
Tied = 0.0%	Tau-a = 0.393
(43815 pairs)	c = 0.999

Residual Chi-Square = 10.7251 with 6 DF (p=0.0973)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	1.0633	0.3025	Sonico
RT	5.2277	0.0222	Resistividade
PHIN	0.0186	0.8916	Poros. Neutrons
RHOC	0.0186	0.8916	Densidade
NPLT	0.0086	0.9262	Razao Phin,Rhoc
PHIT	0.1455	0.7029	Poros. Total
PHIE	0.0266	0.8705	Poros. Efetiva

Step 3. Variable RT entered:

The LOGISTIC Procedure

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	551.719	38.210	.
SC	555.876	54.838	.
-2 LOG L	549.719	30.210	519.509 with 3 DF (p=0.0001)
Score	.	.	375.846 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	2.4884	2.8116	0.7833	0.3761	.	.	Intercept
RT	1	1.7676	0.7867	5.0485	0.0246	0.977778	5.857	Resistividade
GRCO	1	-0.1154	0.0473	5.9627	0.0146	-1.193177	0.891	Raios-Gama
DRDN	1	-5.5102	1.4199	15.0592	0.0001	-6.486185	0.004	Diferenca Rhoc,Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 99.9%	Somers' D = 0.998
Discordant = 0.1%	Gamma = 0.998
Tied = 0.0%	Tau-a = 0.394
(43815 pairs)	c = 0.999

Residual Chi-Square = 8.5586 with 5 DF (p=0.1280)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.0155	0.9008	Sonico
PHIN	3.5855	0.0583	Poros. Neutrons
RHOC	3.5855	0.0583	Densidade
NPLT	3.1819	0.0745	Razao Phin,Rhoc
PHIT	2.2625	0.1325	Poros. Total
PHIE	3.4870	0.0619	Poros. Efetiva

Step 4. Variable RHOC entered:

The LOGISTIC Procedure

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	551.719	36.666	.
SC	555.876	57.451	.
-2 LOG L	549.719	26.666	523.053 with 4 DF (p=0.0001)
Score	.	.	376.123 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	39.6906	23.1874	2.9300	0.0869	.	.	Intercept
RT	1	2.8567	1.2157	5.5214	0.0188	1.580188	17.403	Resistividade
RHOC	1	-15.8545	9.3974	2.8464	0.0916	-0.820783	0.000	Densidade
GRCO	1	-0.1394	0.0644	4.6909	0.0303	-1.442067	0.870	Raios-Gama
DRDN	1	-5.7758	1.6150	12.7896	0.0003	-6.798864	0.003	Diferenca Rhoc,Phin

Association of Predicted Probabilities and Observed Responses

Concordant = 99.9%	Somers' D = 0.999
Discordant = 0.1%	Gamma = 0.999
Tied = 0.0%	Tau-a = 0.394
(43815 pairs)	c = 0.999

Residual Chi-Square = 5.3609 with 4 DF (p=0.2522)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	5.2186	0.0223	Sonico
PHIN	.	.	Poros. Neutrons
NPLT	0.1667	0.6830	Razao Phin,Rhoc
PHIT	5.3110	0.0212	Poros. Total
PHIE	0.2835	0.5944	Poros. Efetiva

Step 5. Variable PHIT entered:

The LOGISTIC Procedure

Model Fitting Information and Testing Global Null Hypothesis BETA=0

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	551.719	33.381	.
SC	555.876	58.323	.
-2 LOG L	549.719	21.381	528.338 with 5 DF (p=0.0001)
Score	.	.	377.911 with 5 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio	Variable Label
INTERCPT	1	395.7	180.0	4.8291	0.0280	.	.	Intercept
RT	1	3.9314	1.6355	5.7782	0.0162	2.174704	50.979	Resistividade
RHOC	1	-146.0	66.0687	4.8860	0.0271	-7.560491	0.000	Densidade
GRCO	1	-0.1795	0.0776	5.3533	0.0207	-1.856029	0.836	Raios-Gama
DRDN	1	-3.6386	1.8480	3.8768	0.0490	-4.283120	0.026	Diferenca Rhoc,Phin
PHIT	1	-2.1787	1.0657	4.1795	0.0409	-4.634524	0.113	Poros. Total

Association of Predicted Probabilities and Observed Responses

Concordant	=100.0%	Somers' D	= 0.999
Discordant	= 0.0%	Gamma	= 0.999
Tied	= 0.0%	Tau-a	= 0.394
(43815 pairs)		c	= 1.000

Residual Chi-Square = 0.3765 with 3 DF (p=0.9451)

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square	Variable Label
DLT	0.2680	0.6047	Sonico
PHIN	.	.	Poros. Neutrons
NPLT	0.0255	0.8732	Razao Phin,Rhoc
PHIE	0.0497	0.8236	Poros. Efetiva

NOTE: No (additional) variables met the 0.2 significance level for entry into the model.

The LOGISTIC Procedure

Summary of Stepwise Procedure

Step	Variable		Number	Score	Wald	Pr >	Variable
	Entered	Removed	In	Chi-Square	Chi-Square	Chi-Square	Label
1	DRDN		1	367.2	.	0.0001	Diferenca Rhoc,Phin
2	GRCO		2	2.9167	.	0.0877	Raios-Gama
3	RT		3	5.2277	.	0.0222	Resistividade
4	RHOC		4	3.5855	.	0.0583	Densidade
5	PHIT		5	5.3110	.	0.0212	Poros. Total

Facies obtidas pelo modelo de P1

TABLE OF REALFAC BY FACP

REALFAC(Facies Padrao)		FACP(fac estimada por P1)		
Frequency				
Percent				
Row Pct				
Col Pct	0	1	Total	
0	94	4	98	
	41.59	1.77	43.36	
	95.92	4.08		
	95.92	3.13		
1	4	124	128	
	1.77	54.87	56.64	
	3.13	96.88		
	4.08	96.88		
Total	98	128	226	
	43.36	56.64	100.00	

Facies obtidas pelo modelo de P1

TABLE OF REALFAC BY FACLOG

REALFAC(Facies Padrao)

FACLOG(Facies estim logistica)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	92	6	98
	40.71	2.65	43.36
	93.88	6.12	
	97.87	4.55	
1	2	126	128
	0.88	55.75	56.64
	1.56	98.44	
	2.13	95.45	
Total	94	132	226
	41.59	58.41	100.00

Facies Estimada pelo Modelo Global

TABLE OF REALFAC BY FACG

REALFAC(Facies Padrao)
FACG(fac estimada global)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	94	4	98
	41.59	1.77	43.36
	95.92	4.08	
	96.91	3.10	
1	3	125	128
	1.33	55.31	56.64
	2.34	97.66	
	3.09	96.90	
Total	97	129	226
	42.92	57.08	100.00

Facies Estimada pelo Modelo Global

TABLE OF REALFAC BY FACLOG

REALFAC(Facies Padrao)
FACLOG(Facies estim logistica)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	92	6	98
	40.71	2.65	43.36
	93.88	6.12	
	97.87	4.55	
1	2	126	128
	0.88	55.75	56.64
	1.56	98.44	
	2.13	95.45	
Total	94	132	226
	41.59	58.41	100.00

P1 proprio x P2

TABLE OF REALFAC BY FAC

REALFAC(Facies Padrao)
FAC(facies calculada)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	93	5	98
	41.15	2.21	43.36
	94.90	5.10	
	98.94	3.79	
1	1	127	128
	0.44	56.19	56.64
	0.78	99.22	
	1.06	96.21	
Total	94	132	226
	41.59	58.41	100.00

P2 proprio x P2

TABLE OF REALFAC BY FAC

REALFAC(Facies Padrao)
FAC(facies calculada)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	93	5	98
	41.15	2.21	43.36
	94.90	5.10	
	97.89	3.82	
1	2	126	128
	0.88	55.75	56.64
	1.56	98.44	
	2.11	96.18	
Total	95	131	226
	42.04	57.96	100.00

Global proprio x P2

TABLE OF REALFAC BY FAC

REALFAC(Facies Padrao)
FAC(facies calculada)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	98	0	98
	43.36	0.00	43.36
	100.00	0.00	
	98.99	0.00	
1	1	127	128
	0.44	56.19	56.64
	0.78	99.22	
	1.01	100.00	
Total	99	127	226
	43.81	56.19	100.00

Global proprio x P2
TABLE OF REALFAC BY FACIES

REALFAC(Facies Padrao)
FACIES(Facies Est. Logist.)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	83	4	87
	27.30	1.32	28.62
	95.40	4.60	
	96.51	1.83	
1	3	214	217
	0.99	70.39	71.38
	1.38	98.62	
	3.49	98.17	
Total	86	218	304
	28.29	71.71	100.00

TABLE OF REALFAC BY EFAC

REALFAC(Facies Padrao)
EFAC(EleetroFacies)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	71	16	87
	23.36	5.26	28.62
	81.61	18.39	
	100.00	6.87	
1	0	217	217
	0.00	71.38	71.38
	0.00	100.00	
	0.00	93.13	
Total	71	233	304
	23.36	76.64	100.00

Global proprio x F2

TABLE OF REALFAC BY FACIES

REALFAC(Facies Padrao)
FACIES(Facies Est. Logist.)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	93	5	98
	41.15	2.21	43.36
	94.90	5.10	
	97.89	3.82	
1	2	126	128
	0.88	55.75	56.64
	1.56	98.44	
	2.11	96.18	
Total	95	131	226
	42.04	57.96	100.00

TABLE OF REALFAC BY EFAC

REALFAC(Facies Padrao)
EFAC(EleetroFacies)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	92	6	98
	40.71	2.65	43.36
	93.88	6.12	
	100.00	4.48	
1	0	128	128
	0.00	56.64	56.64
	0.00	100.00	
	0.00	95.52	
Total	92	134	226
	40.71	59.29	100.00

Global proprio x P2

TABLE OF REALFAC BY FACIES

REALFAC(Facies Padrao)
FACIES(Facies Est. Logist.)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	86	1	87
	28.29	0.33	28.62
	98.85	1.15	
	98.85	0.46	
1	1	216	217
	0.33	71.05	71.38
	0.46	99.54	
	1.15	99.54	
Total	87	217	304
	28.62	71.38	100.00

TABLE OF REALFAC BY EFAC

REALFAC(Facies Padrao)
EFAC(EleetroFacies)

Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	71	16	87
	23.36	5.26	28.62
	81.61	18.39	
	100.00	6.87	
1	0	217	217
	0.00	71.38	71.38
	0.00	100.00	
	0.00	93.13	
Total	71	233	304
	23.36	76.64	100.00

Global proprio x P2

TABLE OF REALFAC BY FACIES

REALFAC(Facies Padrao)			
FACIES(Facies Est. Logist.)			
Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	86	1	87
	28.29	0.33	28.62
	98.85	1.15	
	98.85	0.46	
1	1	216	217
	0.33	71.05	71.38
	0.46	99.54	
	1.15	99.54	
Total	87	217	304
	28.62	71.38	100.00

TABLE OF REALFAC BY EFAC

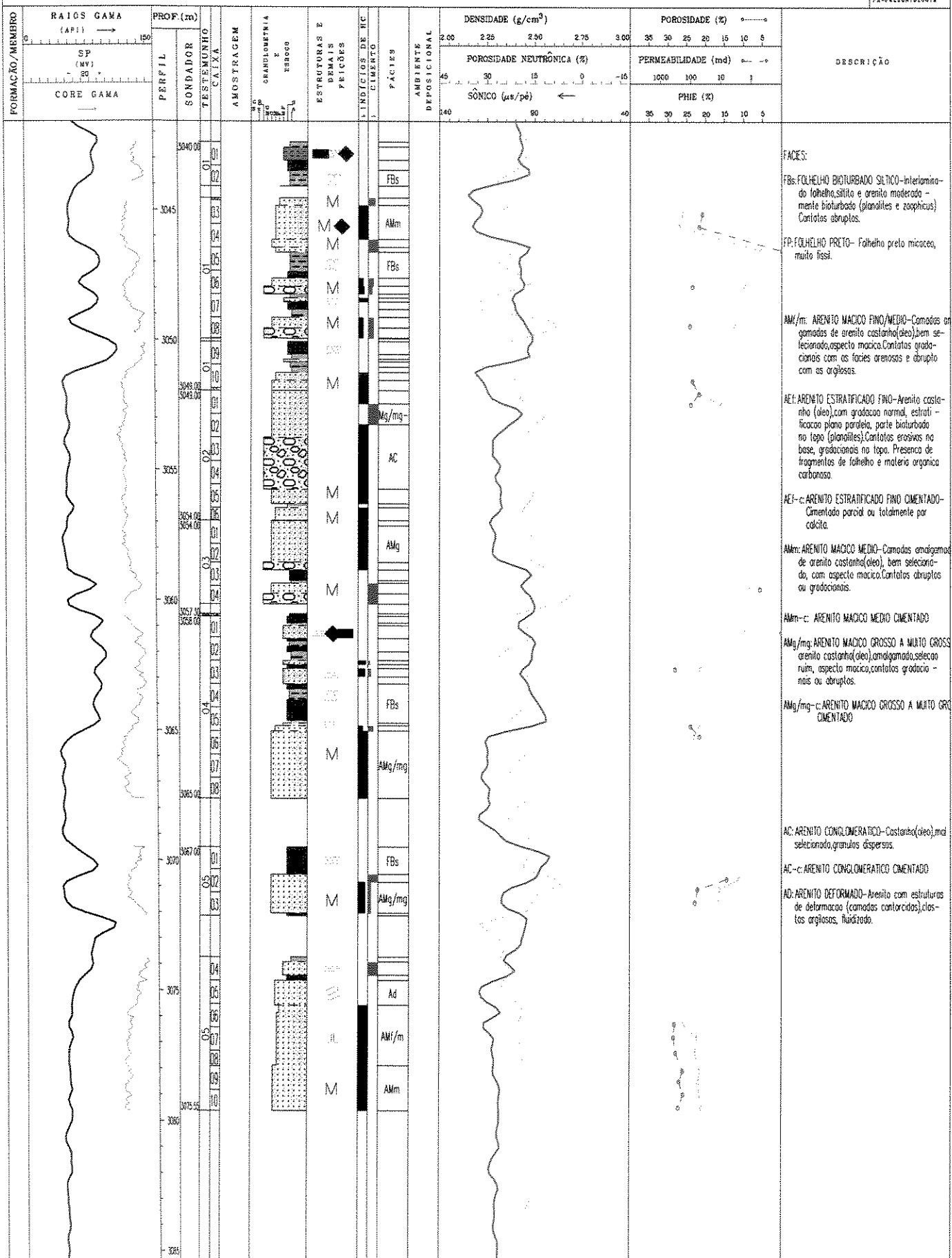
REALFAC(Facies Padrao)			
EFAC(EletroFacies)			
Frequency			
Percent			
Row Pct			
Col Pct	0	1	Total
0	71	16	87
	23.36	5.26	28.62
	81.61	18.39	
	100.00	6.87	
1	0	217	217
	0.00	71.38	71.38
	0.00	100.00	
	0.00	93.13	
Total	71	233	304
	23.36	76.64	100.00

Apêndice 3 - Análises sequenciais de testemunhos dos poços P1 e P2

LEGENDA:

- Bioturbacao intensa ■ Estrat. lam. plano-paral. M Macico
 ■ Bioturbacao moderada ■ Fragmentos de corvao ■ Pilar
 ■ Escorregamento ■ Fragmentos de folhelho

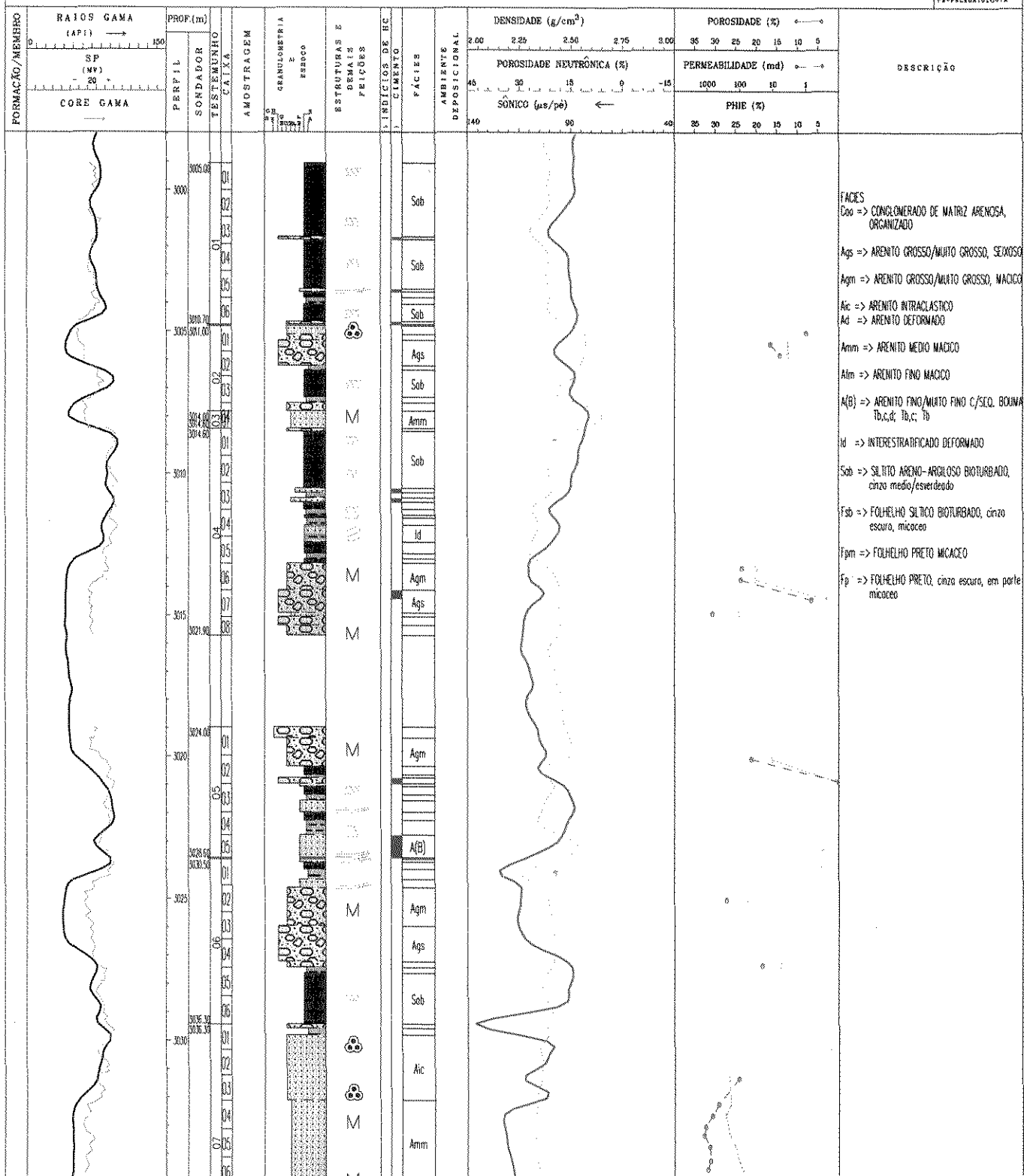
- L-LANINA
O-DRE
M-MEF
C-CRANULOGESTERA
C-CALCINEPTIA
F-FLUG
F-FOTO DE DETALHO
F-FALCANTOLOGIA

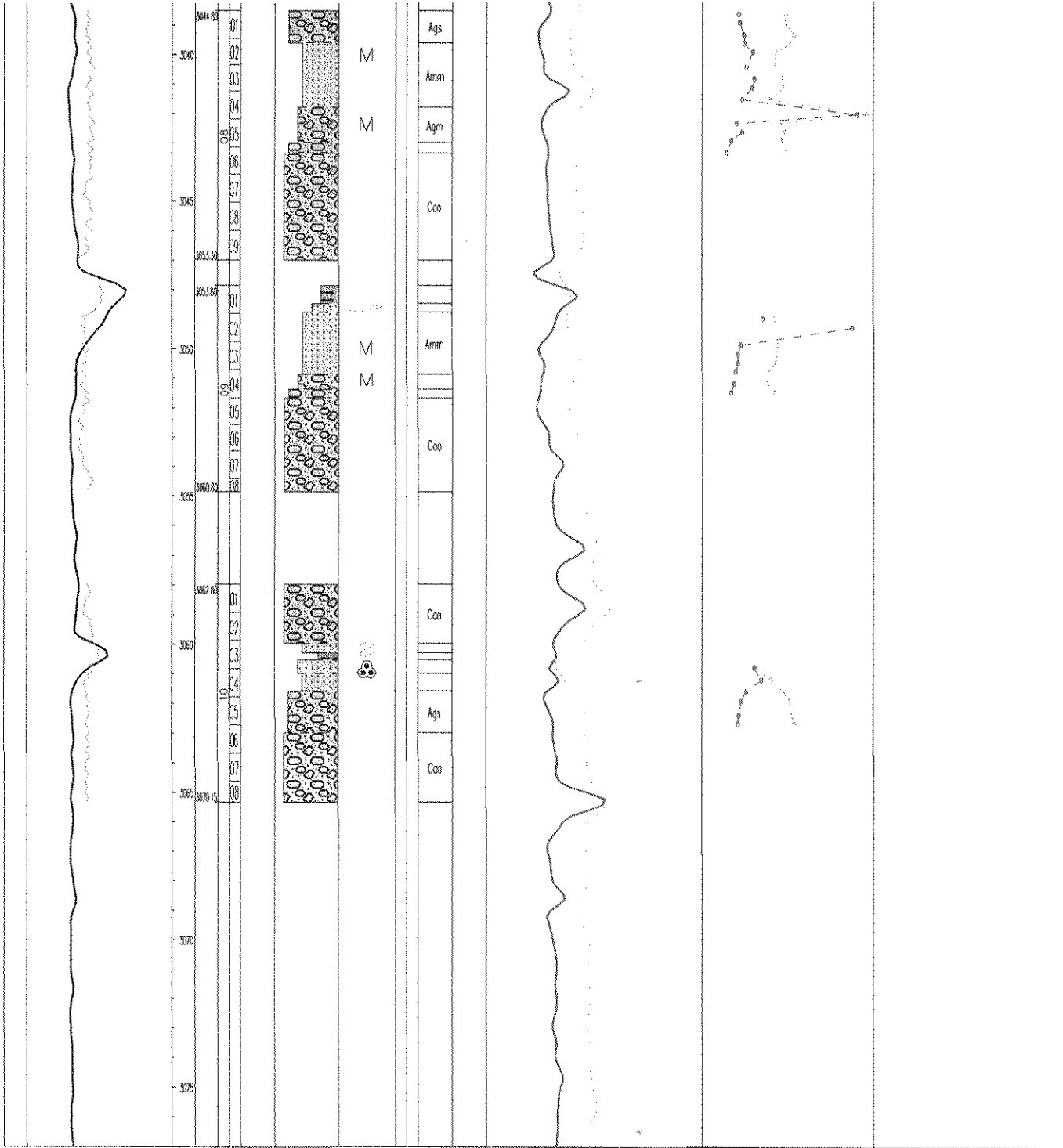


LEGENDA:

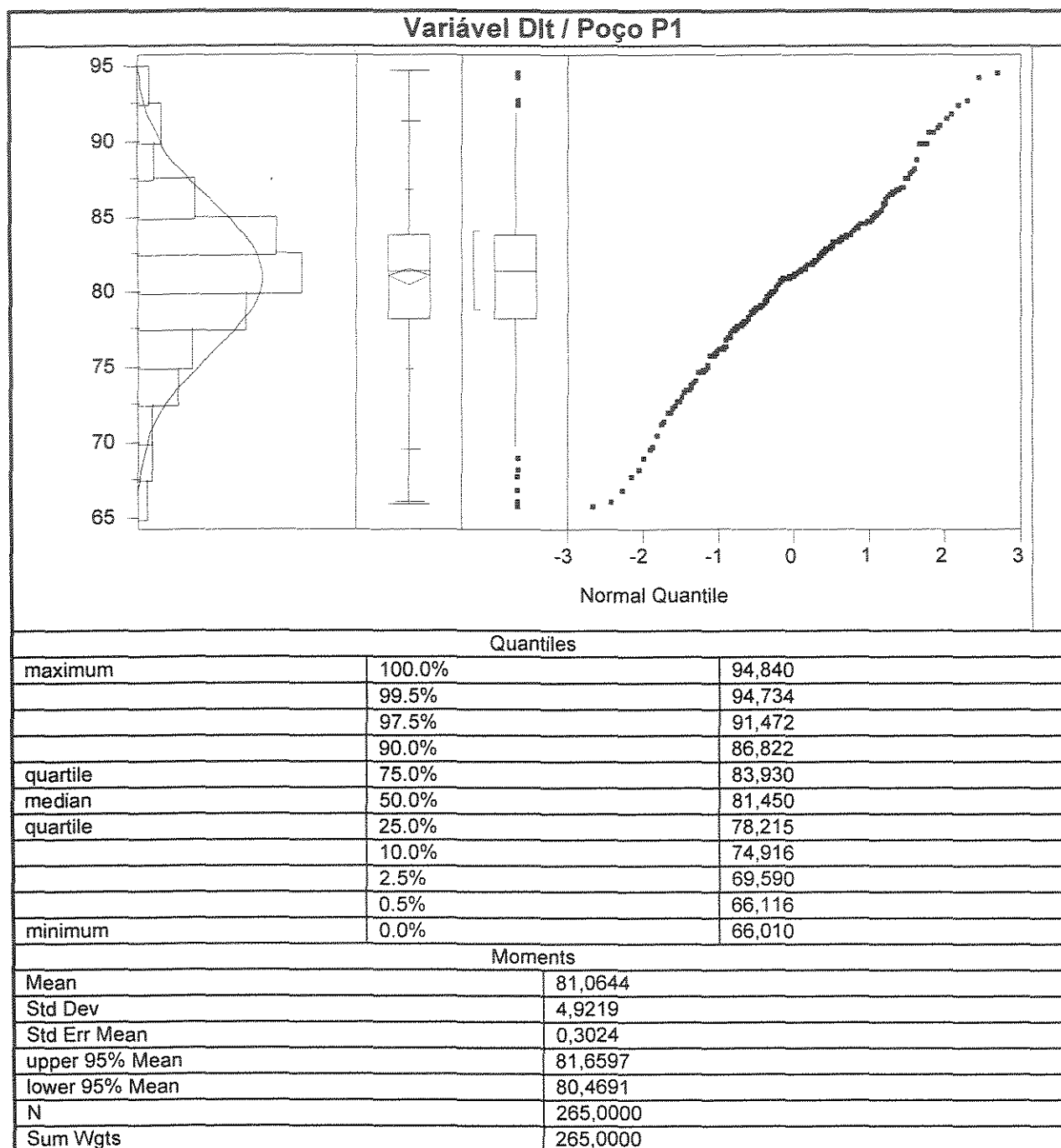
Bioturbacao intensa
 Bioturbacao moderada
 Escorregimento
 Estrat. lam. plano-paral.
 Microestrat. cruz. cavalg. indist.
 Intraclastos
 Mocico

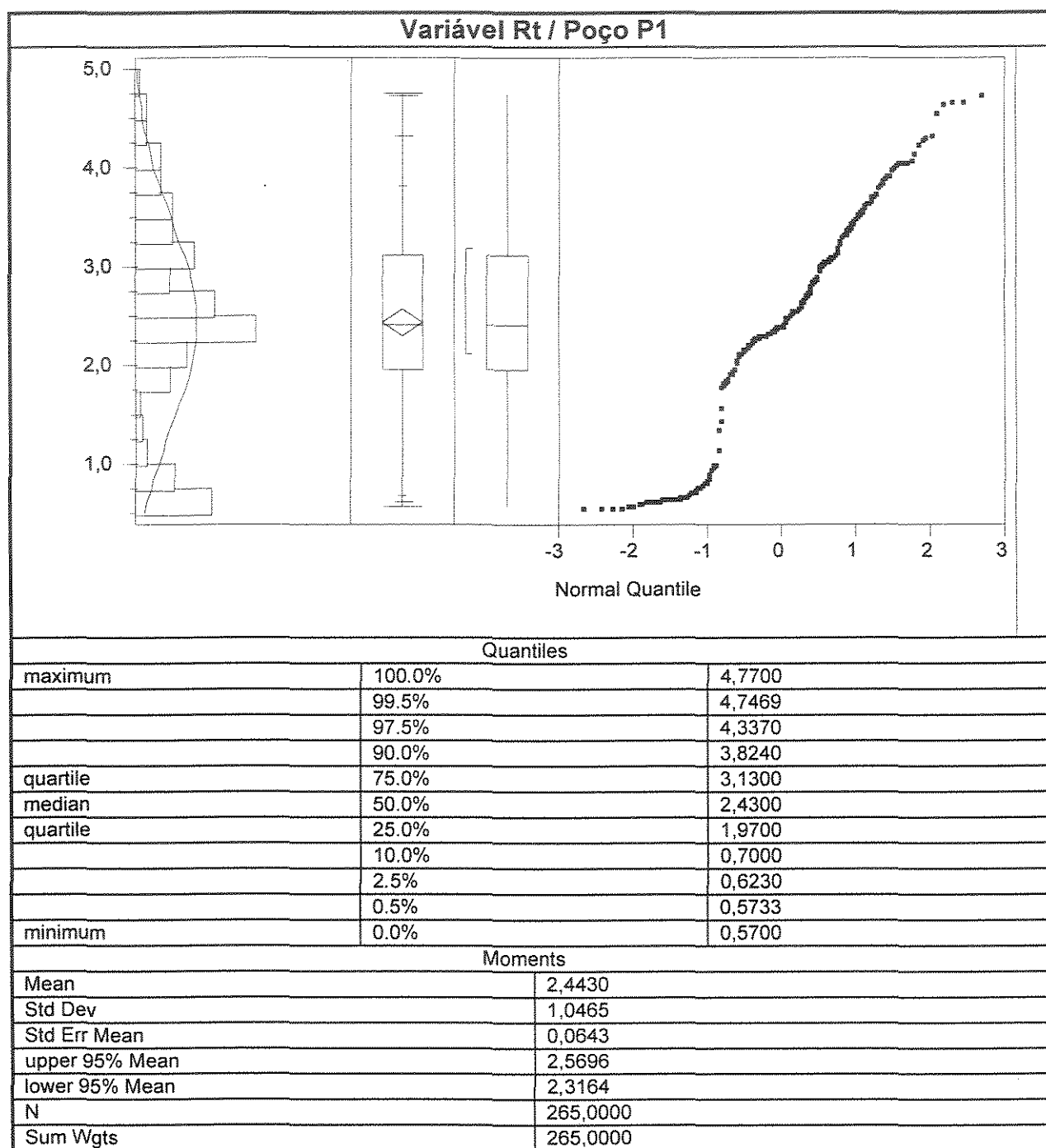
L-LÊXIMA
 D-DEB
 M-MES
 C-CALCINETRIA
 P-PLUG
 F-FOTO DE DETALHE
 PA-PALEONTOLOGIA

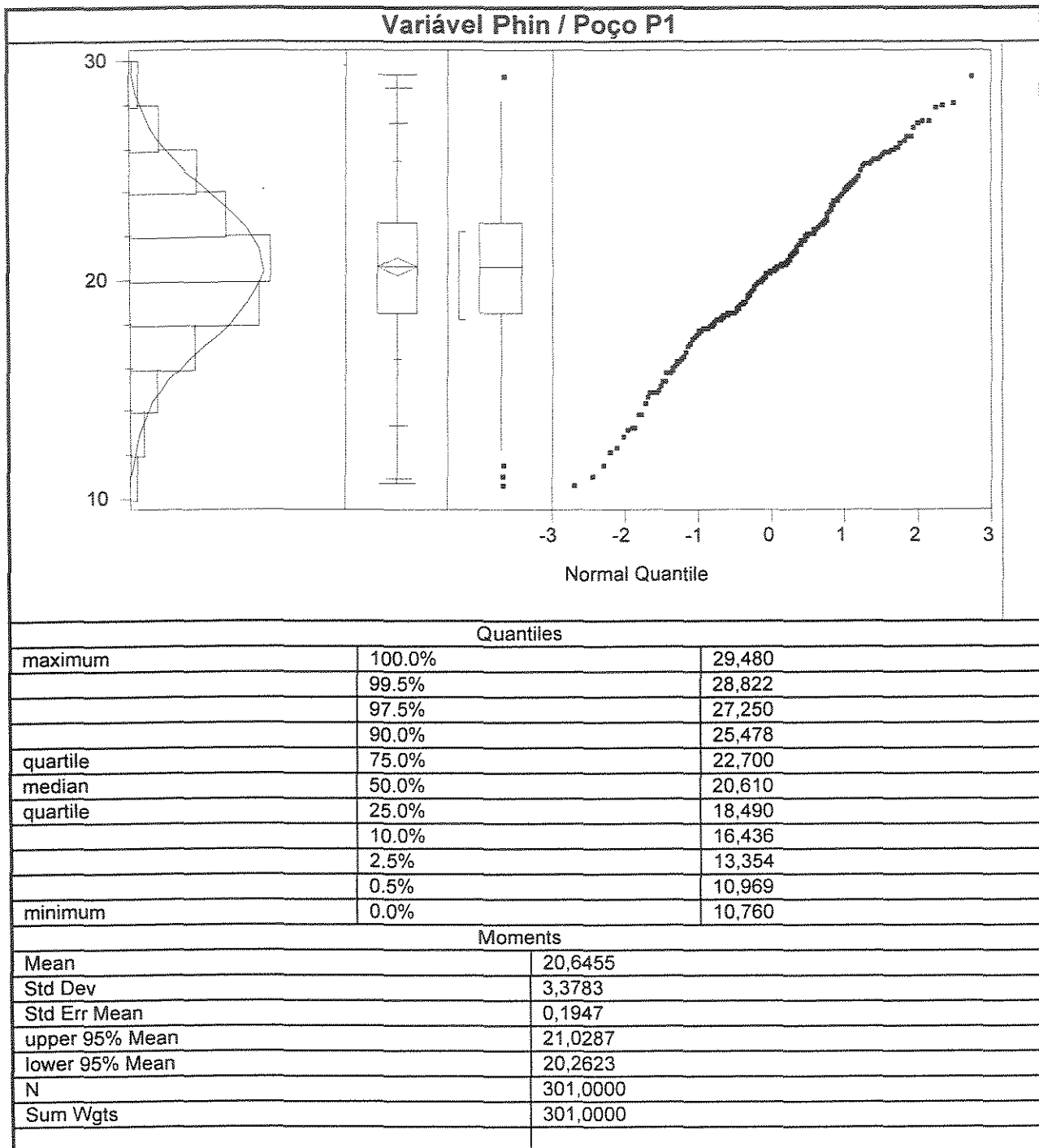




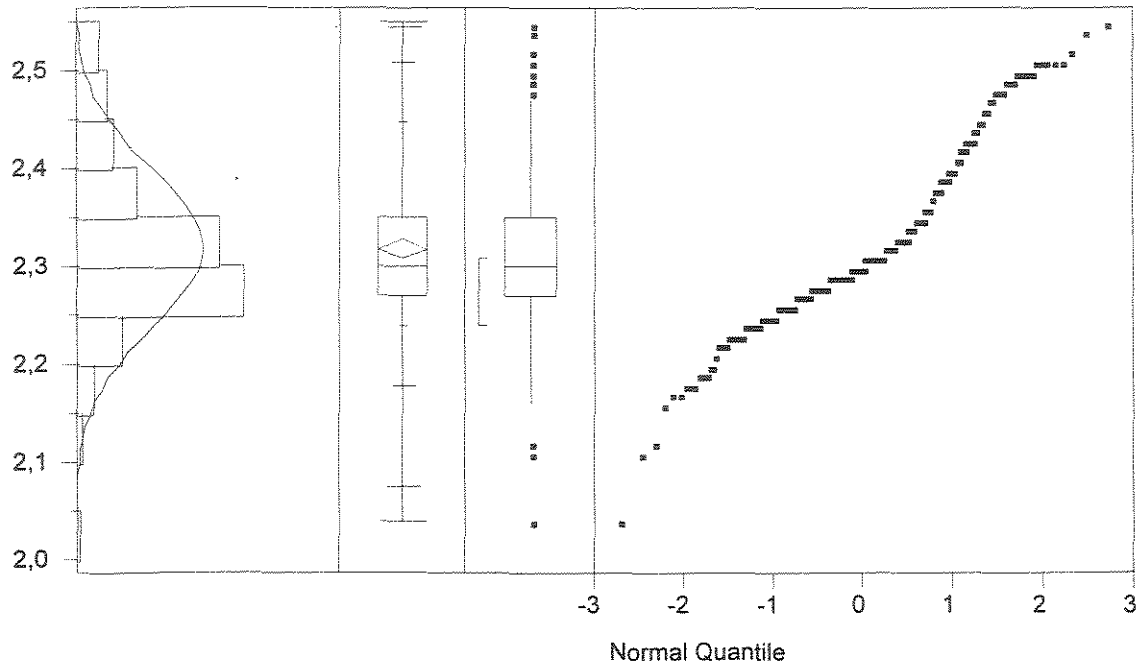
Apêndice 4 - Gráficos e estatísticas descritivas das variáveis de perfis.







Variável Rhoc / Poço P1

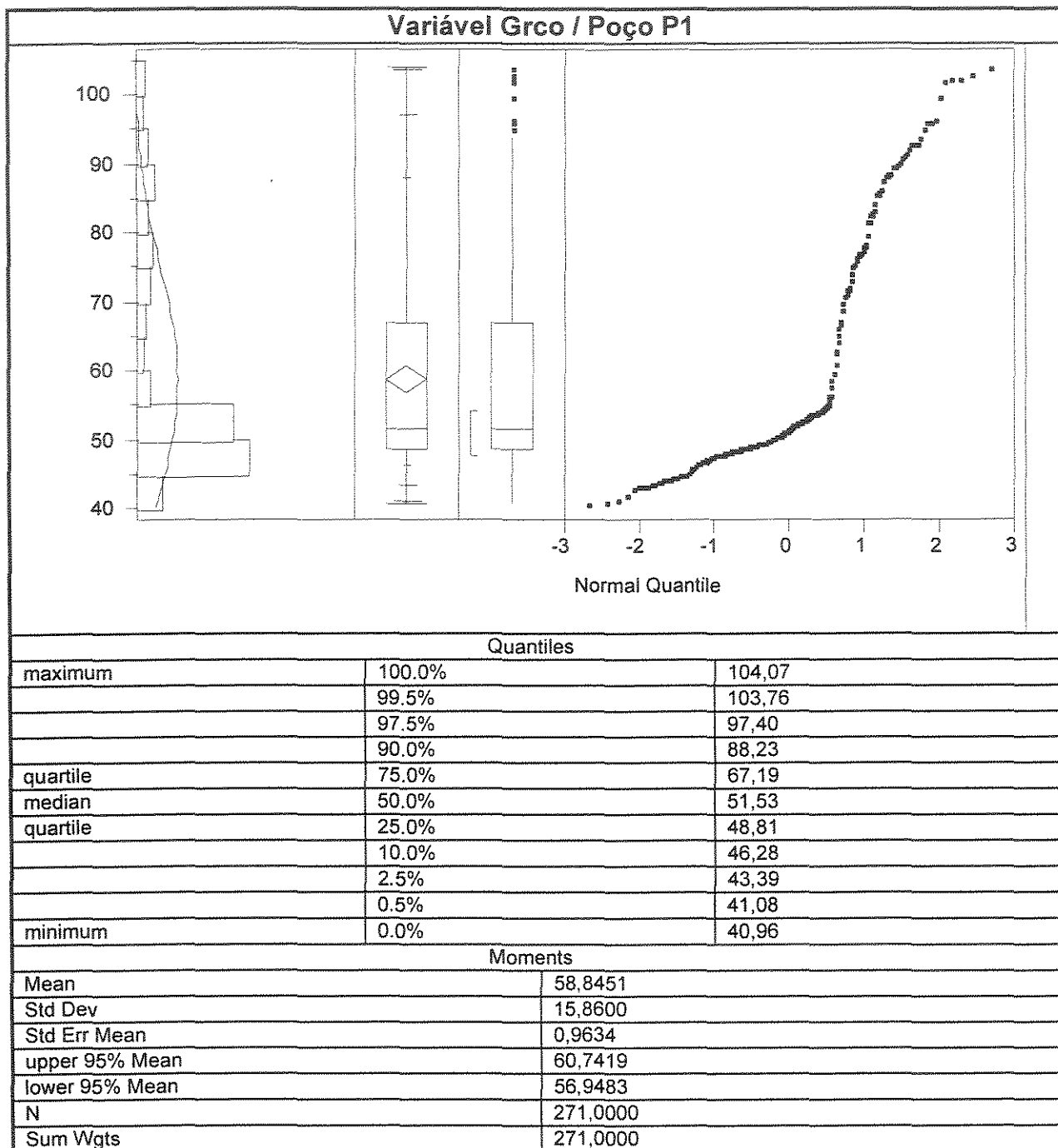


Quantiles

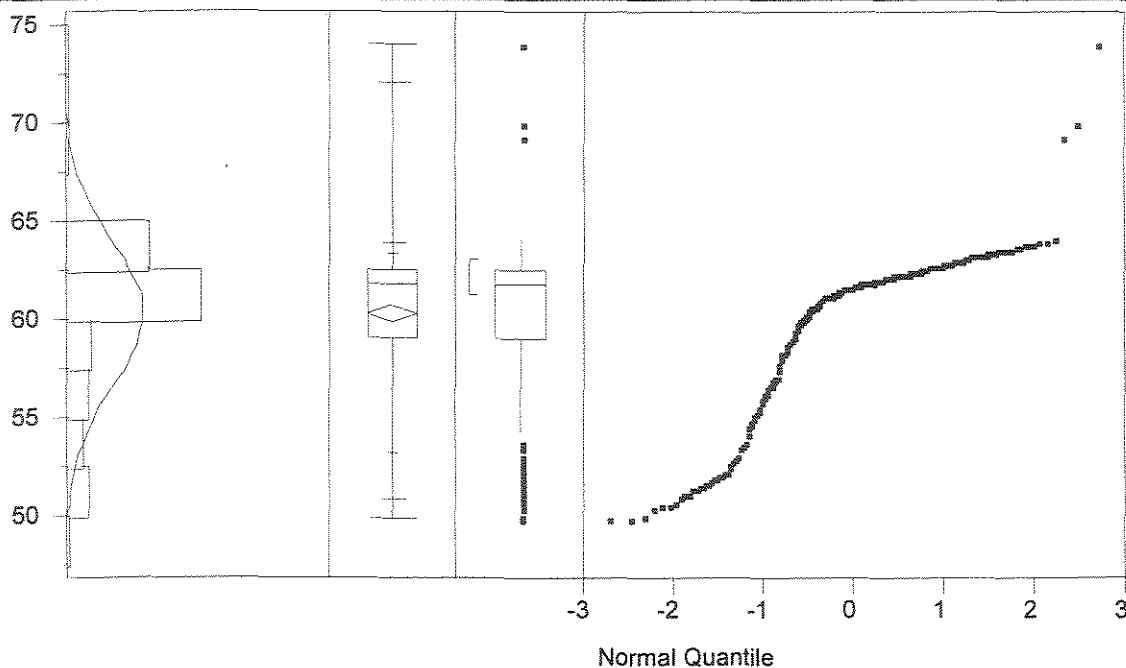
maximum	100.0%	2,5500
	99.5%	2,5449
	97.5%	2,5100
	90.0%	2,4480
quartile	75.0%	2,3500
median	50.0%	2,3000
quartile	25.0%	2,2700
	10.0%	2,2400
	2.5%	2,1800
	0.5%	2,0757
minimum	0.0%	2,0400

Moments

Mean	2,3191
Std Dev	0,0810
Std Err Mean	0,0047
upper 95% Mean	2,3283
lower 95% Mean	2,3099
N	301,0000
Sum Wgts	301,0000



Variável Nplt / Poço P1



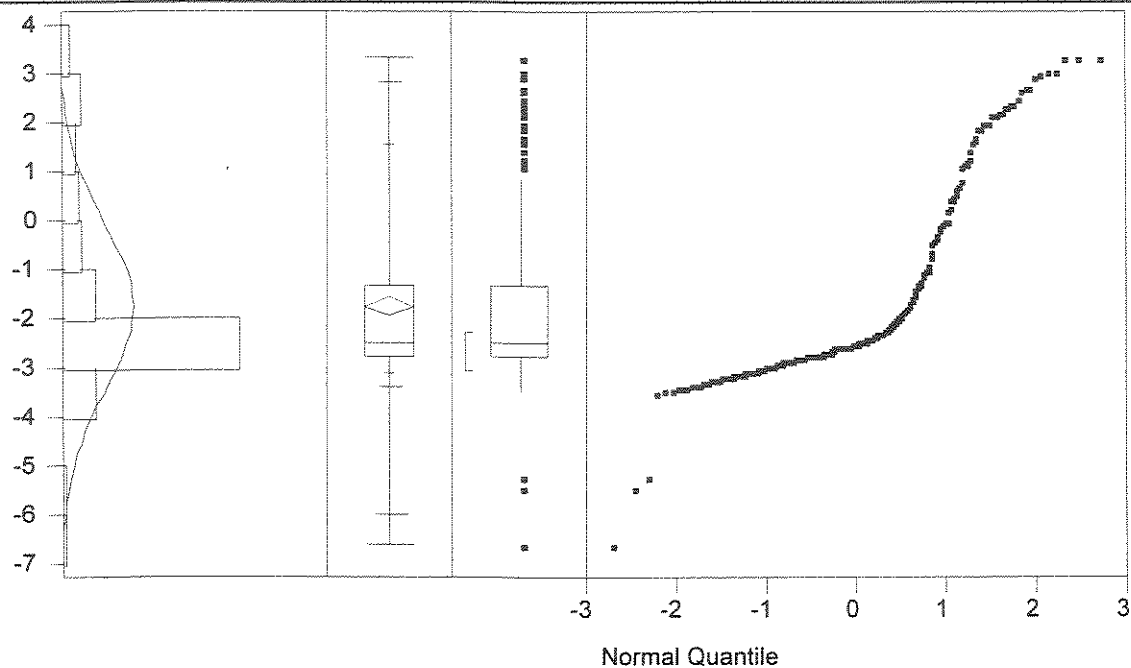
Quantiles

maximum	100.0%	74,160
	99.5%	72,095
	97.5%	63,984
	90.0%	63,340
quartile	75.0%	62,540
median	50.0%	61,810
quartile	25.0%	59,080
	10.0%	53,272
	2.5%	50,917
	0.5%	49,951
minimum	0.0%	49,920

Moments

Mean	60,3384
Std Dev	3,8074
Std Err Mean	0,2195
upper 95% Mean	60,7703
lower 95% Mean	59,9065
N	301,0000
Sum Wgts	301,0000

Variável Drdn / Poço P1

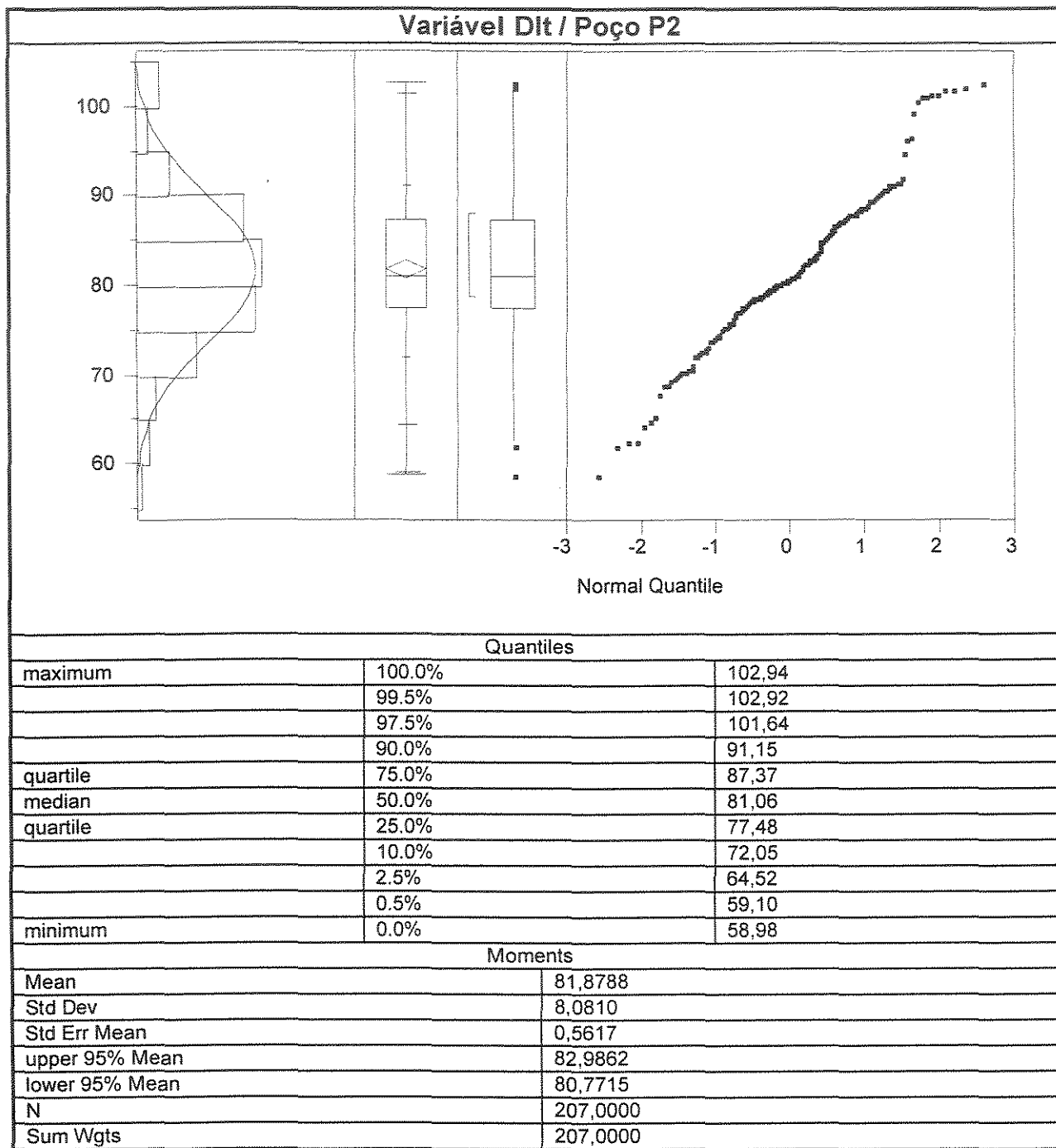


Quantiles

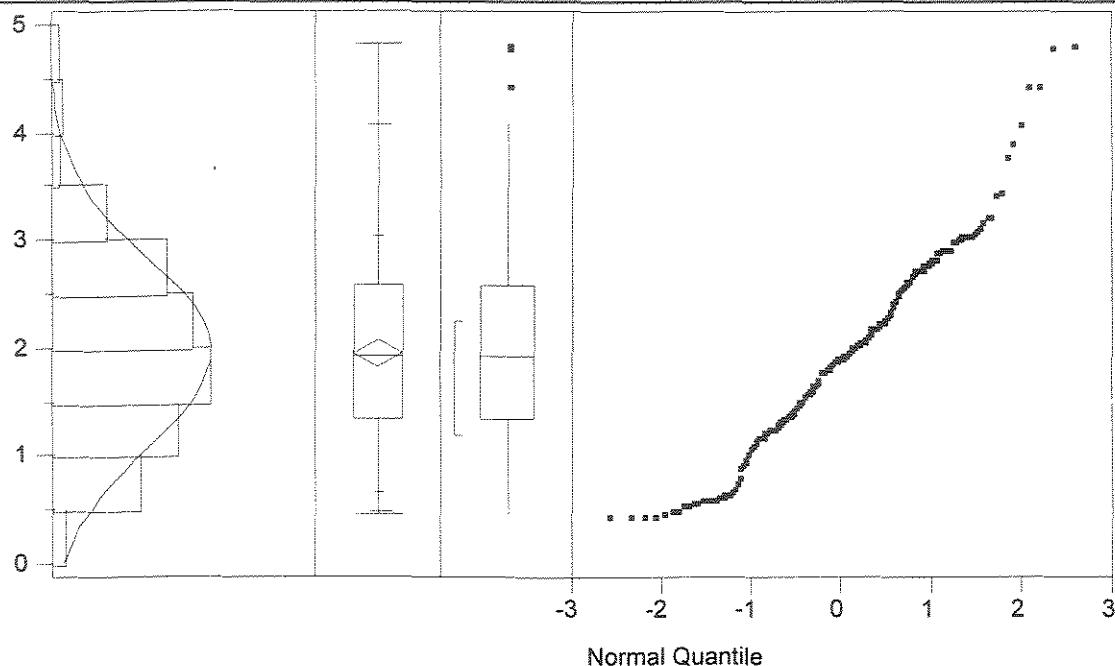
maximum	100.0%	3,3600
	99.5%	3,3600
	97.5%	2,8500
	90.0%	1,5960
quartile	75.0%	-1,2800
median	50.0%	-2,4600
quartile	25.0%	-2,7600
	10.0%	-3,0600
	2.5%	-3,3600
	0.5%	-5,9782
minimum	0.0%	-6,5800

Moments

Mean	-1,7357
Std Dev	1,7534
Std Err Mean	0,1011
upper 95% Mean	-1,5368
lower 95% Mean	-1,9346
N	301,0000
Sum Wgts	301,0000



Variável Rt / Poço P2



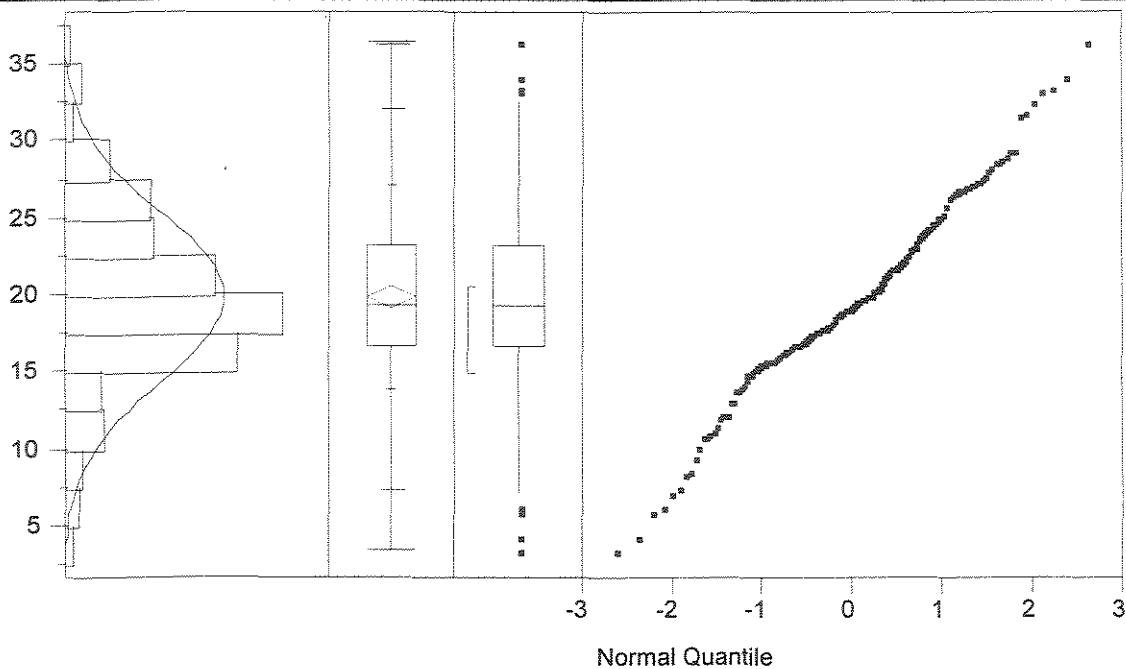
Quantiles

maximum	100.0%	4,8300
	99.5%	4,8296
	97.5%	4,0760
	90.0%	3,0420
quartile	75.0%	2,5800
median	50.0%	1,9400
quartile	25.0%	1,3400
	10.0%	0,6680
	2.5%	0,5040
	0.5%	0,4600
minimum	0.0%	0,4600

Moments

Mean	1,9624
Std Dev	0,8763
Std Err Mean	0,0609
upper 95% Mean	2,0825
lower 95% Mean	1,8423
N	207,0000
Sum Wgts	207,0000

Variável Phin / Poço P2

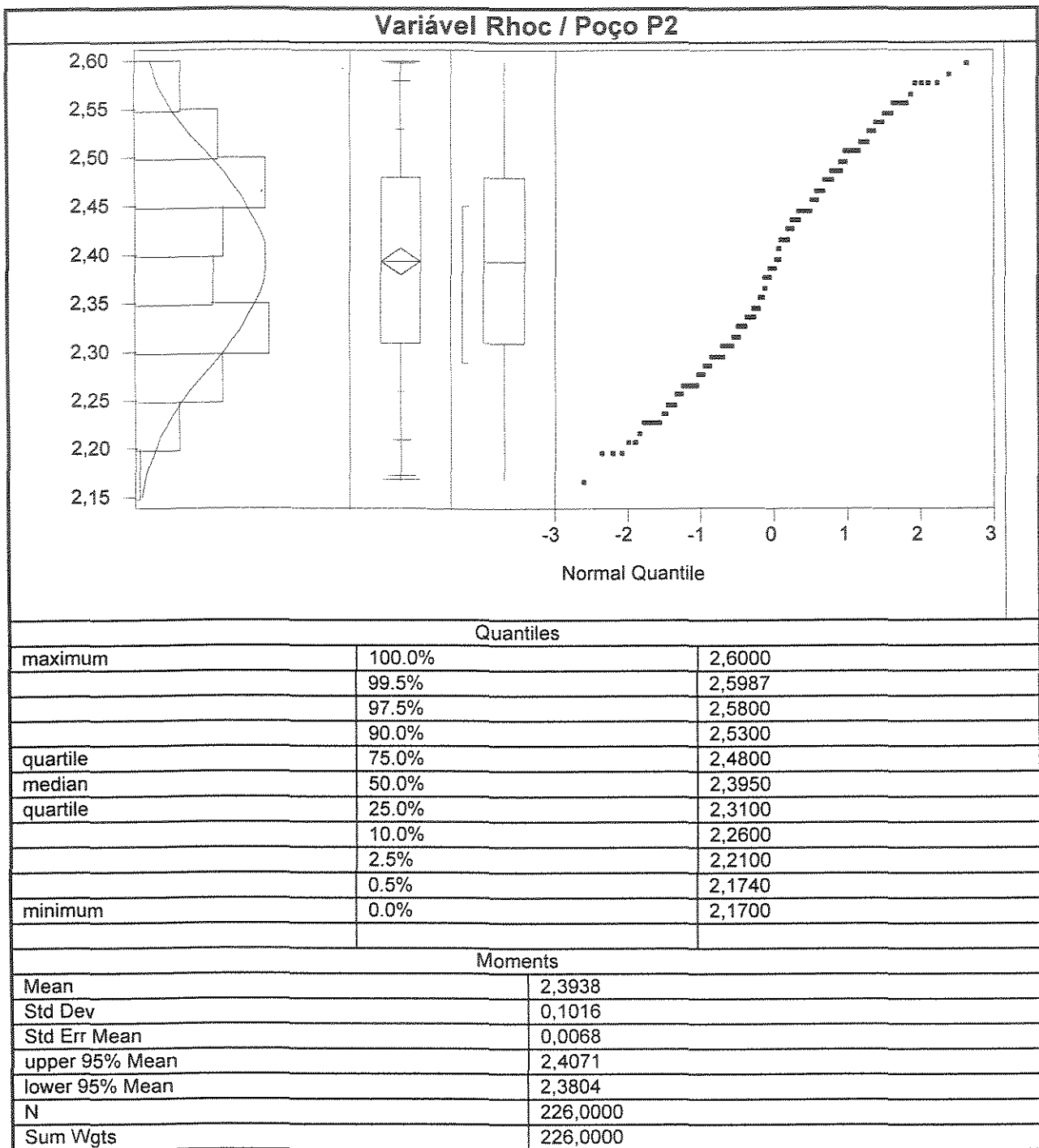


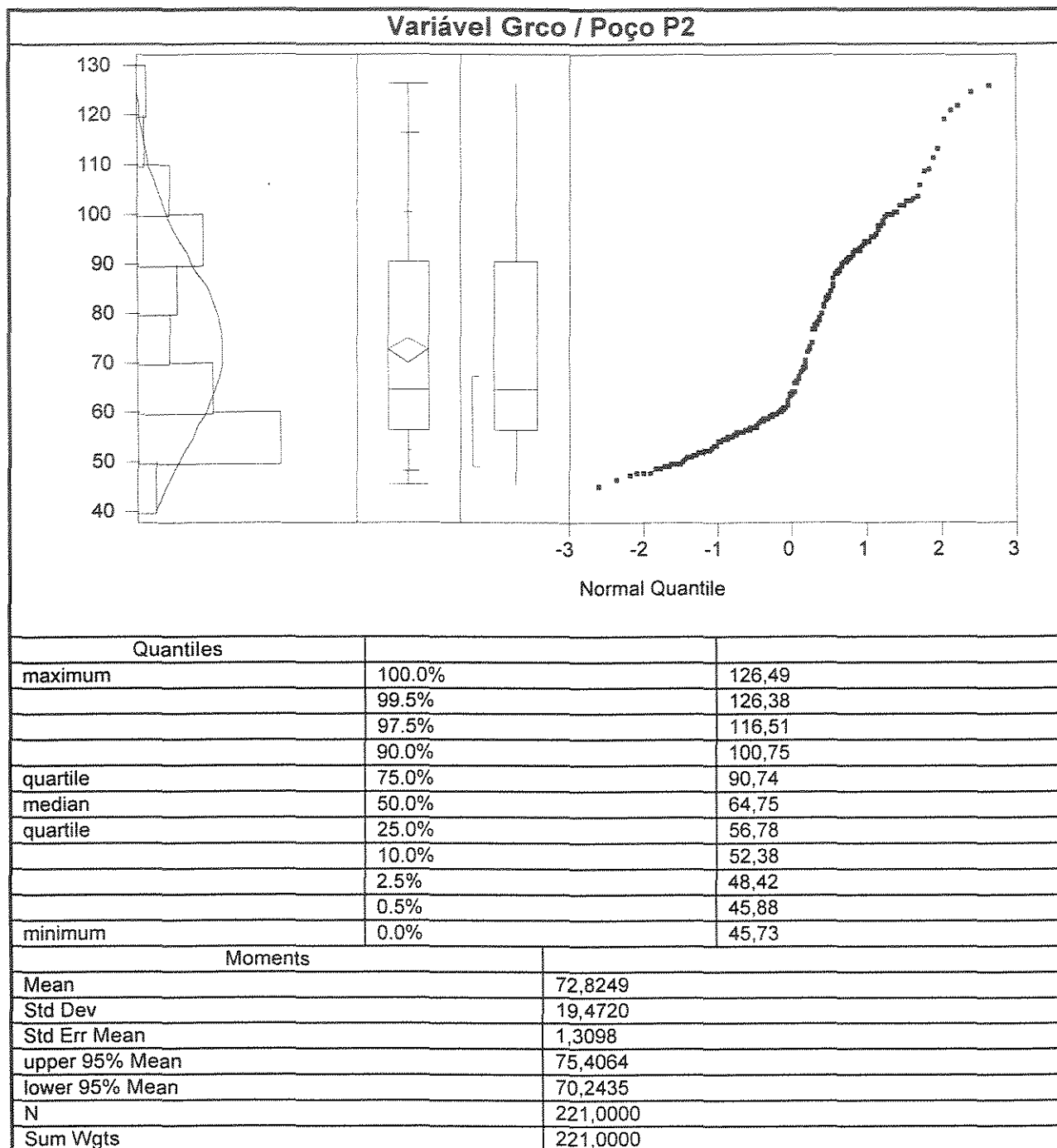
Quantiles

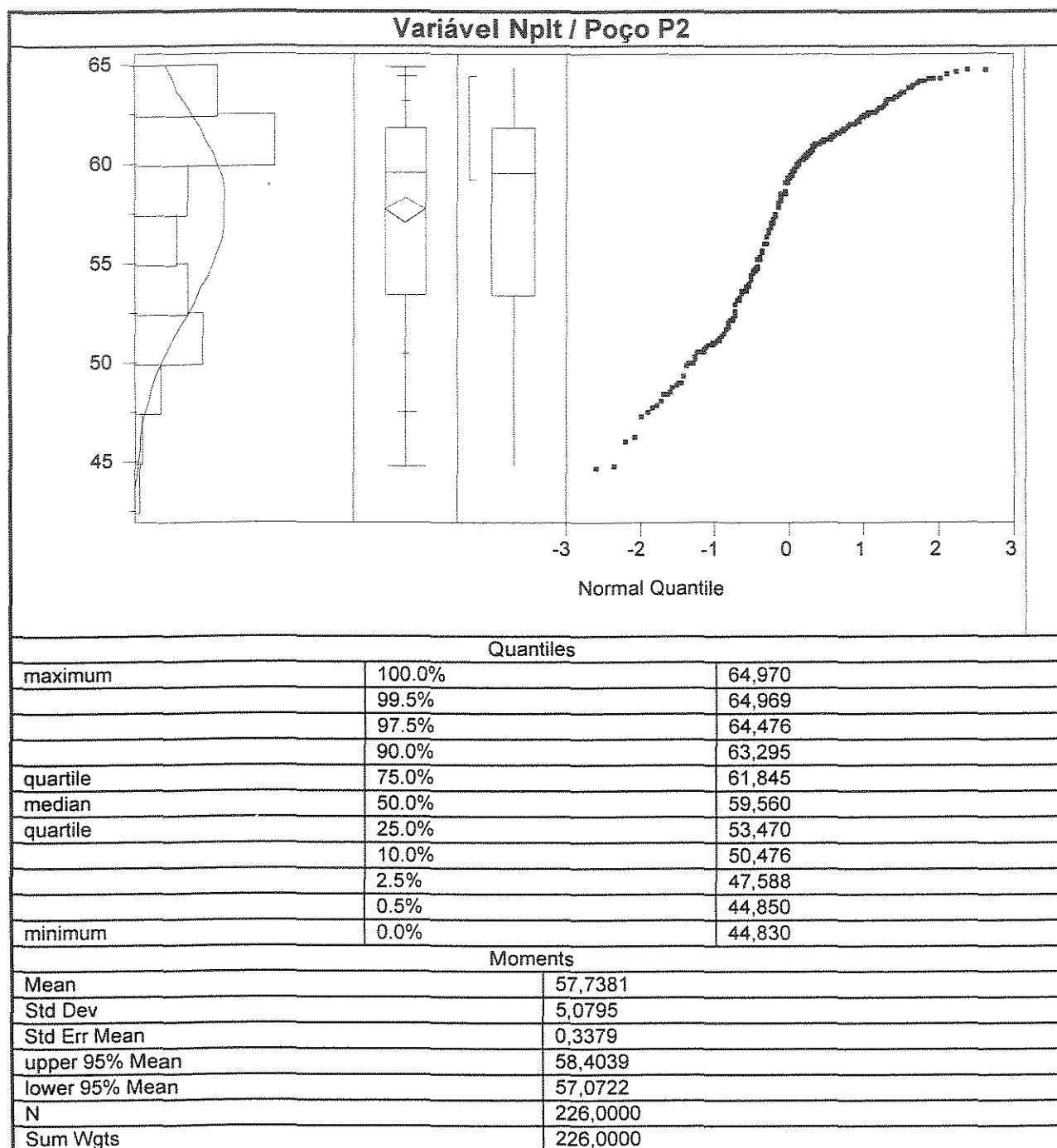
maximum	100.0%	36,600
	99.5%	36,269
	97.5%	32,194
	90.0%	27,146
quartile	75.0%	23,273
median	50.0%	19,350
quartile	25.0%	16,760
	10.0%	13,925
	2.5%	7,402
	0.5%	3,548
minimum	0.0%	3,420

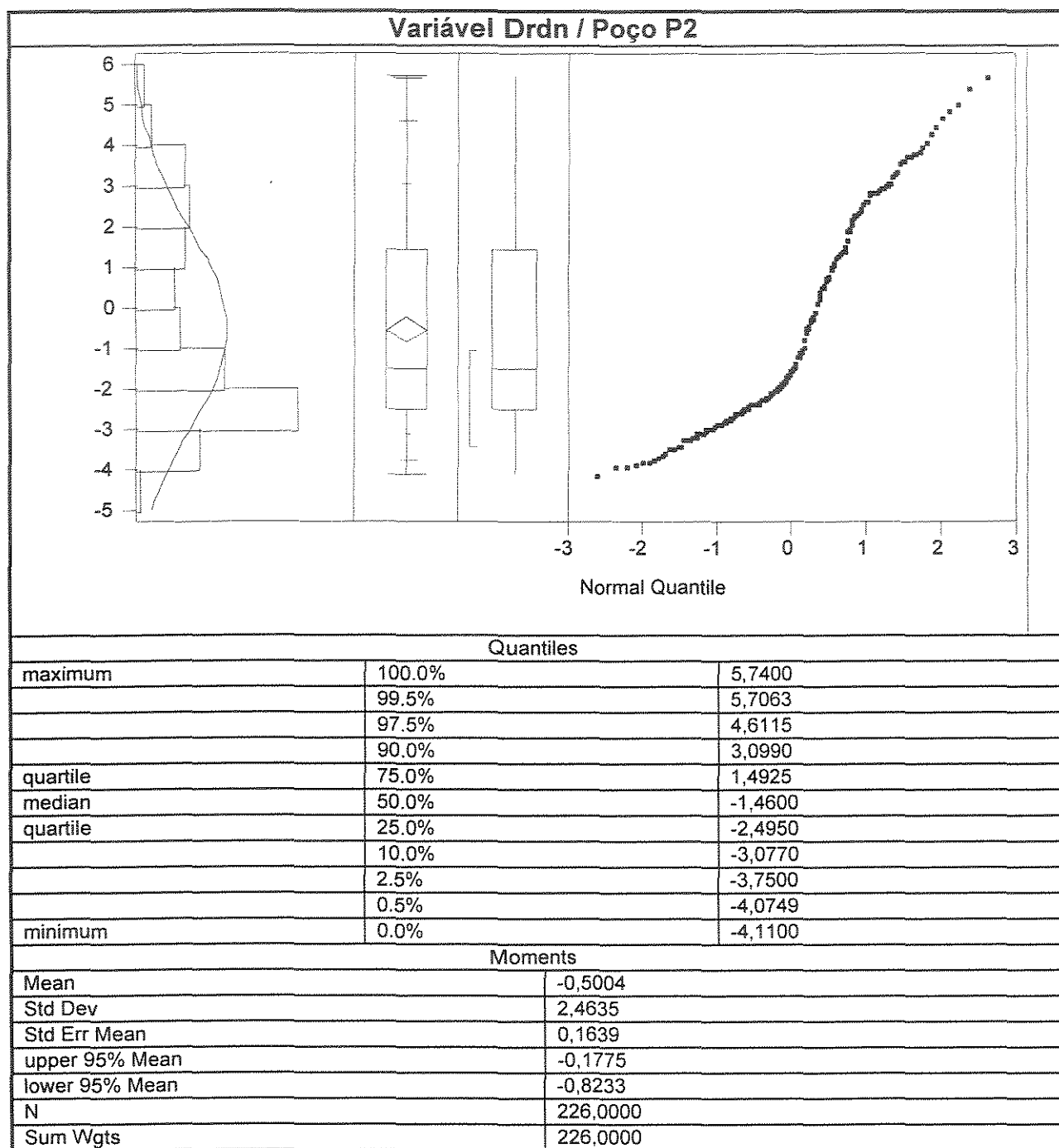
Moments

Mean	19,8724
Std Dev	5,4978
Std Err Mean	0,3657
upper 95% Mean	20,5931
lower 95% Mean	19,1518
N	226,0000
Sum Wgts	226,0000









Referências Bibliográficas

- Barros, A. J. D. - 1990 - *Modelagem Estatística em Estudos Epidemiológicos. O Modelo Logístico*. Dissertação de Mestrado. Universidade Estadual de Campinas. 93 p.
- Becker, M. R. et alli - 1989 - *Caracterização do Reservatório Principal do Campo de Pargo, Bacia de Campos, Brasil*. Relatório Técnico Cenpes-953. 121 p.
- Bruhn, C. H. L. & Moraes, M. A. S.- 1988 - Turbiditos Brasileiros : caracterização geométrica e faciológica. In : CONGRESSO BRASILEIRO DE GEOLOGIA, 35, Belém, 1988. *Anais*. Belém, SBG, v.2, p.224-238.
- Bruhn, C. H. L. - 1993 - *High-Resolution Stratigraphy, Reservoir Geometry, And Facies Characterization of Cretaceous and Tertiary Turbidites from Brazilian Passive Margin Basins*. Tese de Doutorado. McMaster University. 433 p.
- Bucheb, J. A. - 1991 - *Aplicação de Tratamento Estatístico Multivariante em Dados de Perfis de Poços da Bacia de SE - AL*. Tese de Mestrado. Universidade Federal do Pará. 138p.
- Bucheb, J. A. - 1992 - Aplicação de Análise de Componentes Principais em Dados de Perfis. *Boletim de Geociências da Petrobrás*. Rio de Janeiro, 6(1/2):5-16.
- Castro, W. B. de M. - 1991 - Um Algoritmo para Reconhecimento Semi-Automático de Eletrofácies em Perfis de Poços. In: SEMINÁRIO TEC. OPER. GEOL., 3, Cabo Frio, 1991. *Anais...* Rio de Janeiro, Petrobras, v.1, p. 183-190.
- Constanza, M. C. & Afifi, A. A. - 1979 - Comparison of Stopping Rules in Forward Stepwise Discriminant Analysis. *Journal of the American Statistical Association*, 74, 777-785
- Efron, B. - 1975 - The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis. *Journal of the American Statistical Association*, 70, 892-898.

- Elek, I. - 1988 - Some Application of Principal Component Analysis : Well-to-Well Correlation, Zonation. *Geobyte*, 3(2):25-31.
- Ferreira, A. B. de H. - 1986 - *Novo Dicionário da Língua Portuguesa*. Aurélio Buarque de Holanda Ferreira e J.E.M.M., Eds.. Rio de Janeiro, Nova Fronteira, 1499 p.
- Franco Filho, N. P. - 1993 - *Avaliação Quantitativa de Perfis. Reservatório Cretácico do Campo de Pargo*. Relatório Interno Petrobrás. 41 p.
- Greder, H. N. et alli - 1995 - Derivation of a Core Permeability Log and Extrapolation to Uncored Intervals. In : *SPWLA 36th Annual Logging Symposium*. 6 p.
- Hook, J. R. et alli - 1994 - Facies and Permeability Prediction from Wireline Logs and Core - A North Sea Case Study. In: *SPWLA 35th Annual Logging Symposium, Section AAA*. 22 p.
- Hosmer, D. W. & Lemeshow, S. - 1989 - *Applied Logistic Regression*. John Wiley & Sons. New York. 307 p.
- Jensen, J. L. - 1986 - *A statistical Study of Reservoir Permeability Distributions*. PhD Dissertation. The University of Texas at Austin. 243 p.
- Lemos, W. P. et alli - 1989 - *Estudo dos Problemas de Baixa Resistividade de Perfis Elétricos dos Campos de Carapeba e Pargo*. Relatório técnico Cenpes 997. Petrobras, RJ. 52 p.
- Lin, J. L. & Salish, H.A. - 1994 - Determination from Well Logs of Porosity and Permeability in a Heterogeneous Reservoir. *Society of Petroleum Engineers, SPE* 28792. 11 p.
- McCullagh, P. & Nelder, J.A. - 1983 - *Generalized Linear Models*. Chapman Hall, London. 261 p.
- Meisner, J. & Demirmen, F. - 1981 - The Creaming Method : A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces. *The Journal of the Royal Statistical Society, Series A* (general), v.144, 1, p. 1-31.

- Pregibon, D. - 1981 - Logistic Regression Diagnostics. *Annals of Statistics*, 9, 705-724.
- Sarzenski, D. J. & Toledo, J. B. de - 1990 - *Correlação Rocha-Perfil : Conceitos e Aplicação em Reservatórios Heterogêneos*. Relatório Técnico Petrobrás. 394 p.
- Serra, O. & Abbott, H. T. - 1980 - The Contribution of Logging Data to Sedimentology and Stratigraphy. In: *TECH. CONF. EXH. SPE 9270*, 55, Dallas. 19 p.
- Silva, R. R. da - 1983 - *Contribuição à Avaliação Probabilística de Potencial Petrolífero*. Tese de Mestrado. Universidade Federal do Rio de Janeiro. 70 p.
- Souza Jr., O. G. de - 1991 - Análise de Dados Multivariados, uma Eficiente Ferramenta para Descrição e Caracterização de Reservatórios. In: SEMINÁRIO TEC. OPER. GEOL., 3, Cabo Frio, 1991. *Anais...* Rio de Janeiro, Petrobras, v.1, p. 121-130
- Steagall, D. E. et alli - 1995 - Estimativa de Permeabilidade Através de Perfis no Reservatório Carbonático Quissam~do Campo de Bonito, Bacia de Campos - RJ. In: VI SIMPÓSIO DE QUANTIFICAÇÃO EM GEOCIÊNCIAS. *Boletim de Resumos Expandidos*, Unesp, Rio claro (SP), 1995. P. 185-188.
- Woronow, A. - 1987 - Cluster Analysis. *Geobyte*, 2(8):34-37.