



PONTIFÍCIA UNIVERSIDADE CATÓLICA  
DO RIO DE JANEIRO



Renata Chaomey Wo

## **Aproximador de Função para Análise Econômica de Projetos no E&P Utilizando Técnicas de Regressão**

Monografia apresentada ao Departamento de Engenharia Elétrica da  
PUC/Rio como parte dos requisitos para a obtenção do título de  
Especialização em Business Intelligence.

Orientador:  
Marley Maria B. Rebuzzi Vellasco, Ph.D

Rio de janeiro  
26 de dezembro de 2011



## **AGRADECIMENTOS**

Agradeço a Deus por guiar-me em todos os momentos de minha vida.

Agradeço a minha família pelo incentivo constante e pela compreensão no decorrer deste período.

Ao meu marido, pelo carinho e companheirismo durante a construção desta monografia e no decorrer do curso.

Aos meus amigos que me apoiaram e me incentivaram nos momentos mais difíceis.

E, em especial, agradeço a empresa OX por fomentar a especialização de novos profissionais na área de Inteligência no Negócio.

## RESUMO

Quando se trata em avaliar as oportunidades de negócio da empresa, os clientes em geral se deparam com a pergunta: Quais projetos a companhia deveria investir?

A resposta para essa questão envolve uma decisão econômica de investimento, onde é necessário avaliar a rentabilidade intrínseca do projeto, ou seja, identificar quais projetos valem mais que o capital necessário para a sua implementação.

No caso da empresa de petróleo OX, essa decisão econômica de investimento engloba uma análise econômica de projetos de produção de petróleo da área de negócios do E&P (Exploração e Produção). Estas análises são realizadas de forma padronizada e parametrizada, de acordo com as premissas econômicas da empresa e o regime fiscal brasileiro.

Atualmente, para essa análise econômica é utilizada uma solução que iremos chamar de PI (Programa de Indicadores) que realiza uma análise para cada campo petrolífero.

No entanto, a empresa OX possui diversas Unidades de Operações (UO) e cada uma delas é composta por vários campos. Portanto, para realizar uma análise econômica para todas as UO's da empresa, é necessário efetuar um grande número de chamadas ao PI, demandando um alto custo computacional.

Assim, quando o cliente deseja fazer um estudo para verificar como se comporta a rentabilidade dos seus projetos, realizando variações sobre os parâmetros de entrada, a partir de um cenário específico, ele demanda um tempo significativo para poder analisar as suas simulações.

Esta monografia apresenta um estudo da técnica de regressão linear múltipla para a criação de um Aproximador de Função capaz de prover, de forma rápida, uma simulação de análise econômica baseada em um cenário pré-existente. Posteriormente, os resultados obtidos pelo modelo serão comparados com os gerados pelo PI.

## **ABSTRACT**

When it comes to evaluate the company's business opportunities, customers often face the question: Which projects should the company invest?

The answer to this question involves an economic investment, which is necessary to evaluate the intrinsic profitability of the project and to identify which projects are worth more than the capital needed for its implementation.

In the case of OX Oil Company, the decision of economic investment includes an economic analysis of projects in Exploration and Production area. The analyses are performed in a standardized and parameterized form, according to the economic premises of the company and Brazilian fiscal system.

Today, the economic analysis is executed through the solution named PI (Program Indicators) that performs an analysis for each oil field. However, the company has several Operations Units and each is composed of several fields.

So, performing an economic analysis for all company Operation Units, it's necessary to make a lot of calls to PI Solution. It will require a high computational cost.

Thus, a significant time to analyze their simulations will be demanded when the client wants to study how behaves the profitability of their projects, by performing variations on the input parameters.

This work presents a study of multiple linear regression technique to create an approximate function that can provide, quickly, a simulation of economic analysis based on a pre-existing scenario. Subsequently, the results obtained by the model will be compared with those generated by the PI.

**LISTA DE SIGLAS**

- PE - Participação Especial
- PI - Programa de Indicadores
- E&P - Exploração e Produção
- OX - Nome fictício para a empresa de petrolífera
- SR - Sistema de Reservas
- UO - Unidade de Operação

## LISTA DE FIGURAS

Figura 1: Objetivo do Trabalho.....	2
Figura 2: Processo de Estimativa de Reservas.....	5
Figura 3: Cálculo do Ano de Corte.....	6
Figura 4: Análise Econômica.....	7
Figura 5: Consolidação e Processamento dos Dados de Reservas.....	8
Figura 6: Fluxo de Caixa Nominal Anual.....	9
Figura 7: Componentes do Fluxo de Caixa.....	10
Figura 8: Classificação dos Investimentos.....	13
Figura 9: Custos Operacionais.....	15
Figura 10: Tributos Diretos.....	16
Figura 11: Participação Especial.....	18
Figura 12: Modelo.....	20
Figura 13: Correlação Linear Positiva Perfeita entre X e Y.....	21
Figura 14: Correlação Linear Nula entre X e Y.....	21
Figura 15: Correlação Linear Negativa Perfeita entre X e Y.....	22
Figura 16: Correlação Não Linear entre X e Y.....	22
Figura 17: Correlação.....	22
Figura 18: Regressão Linear Simples.....	24
Figura 19: Arquitetura.....	28
Figura 20: Base Histórica: Simulador.....	29
Figura 21: Gráfico Coeficiente de Correlação de Pearson.....	30
Figura 22: Importância da Variável.....	31
Figura 23: Resultados dos Modelos de Regressão.....	32
Figura 24: Gráfico de Interações do Modelo de Regressão <i>Stepwise</i> .....	33
Figura 25: Estimativa dos Parâmetros do Modelo de Regressão <i>Stepwise</i> .....	33
Figura 26: Taxa de Erro do Modelo.....	35
Figura 27: Fluxo de Caixa Nominal Realizado x Estimado.....	36

**LISTA DE EQUAÇÕES**

Equação 1: Receita .....	10
Equação 2: Dispêndio Total .....	12
Equação 3: Investimentos.....	13
Equação 4: Normalização Linear .....	20
Equação 5: Coeficiente de Pearson.....	22
Equação 6: Regressão Linear Simples .....	24
Equação 7: Regressão Linear Múltipla .....	25

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1.	MOTIVAÇÃO .....	1
1.2.	OBJETIVOS DO TRABALHO .....	2
1.3.	DESCRIÇÃO DO TRABALHO .....	3
1.4.	ORGANIZAÇÃO DA MONOGRAFIA .....	3
<b>2.</b>	<b>ANÁLISE ECONÔMICA</b> .....	<b>5</b>
2.1.	FLUXO DE CAIXA NOMINAL .....	9
<b>3.</b>	<b>MODELO</b> .....	<b>20</b>
3.1.	NORMALIZAÇÃO .....	20
3.2.	SELEÇÃO DE VARIÁVEIS: CORRELAÇÃO .....	21
3.3.	MODELAGEM .....	23
<b>4.</b>	<b>ESTUDO DE CASO: EMPRESA OX</b> .....	<b>28</b>
4.1.	ESCOPO .....	28
4.2.	DADOS HISTÓRICOS .....	28
4.3.	PRÉ-PROCESSAMENTO .....	29
4.4.	MODELAGEM .....	31
4.5.	RESULTADOS .....	34
<b>5.</b>	<b>CONCLUSÕES e TRABALHOS FUTUROS</b> .....	<b>37</b>
<b>6.</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>38</b>



# 1. INTRODUÇÃO

## 1.1. MOTIVAÇÃO

A empresa OX é responsável por grande parte da exploração e produção de derivados do petróleo no mundo. Atualmente, sua área de Exploração e Produção (E&P) tem como objetivo realizar a estimativa de reservas, através da geração de informações técnico-econômicas de acordo com os critérios estabelecidos por órgãos reguladores (CAPPELLI et.al., 2007).

A estimativa de reservas por empresas de petróleo é feita com base na análise econômica de áreas descobertas e de áreas exploratórias. Para que o volume esperado seja classificado como reserva é necessário que sejam realizados estudos que identifiquem a viabilidade técnica e econômica do projeto, sob as condições econômicas e recursos tecnológicos vigentes na época da avaliação. Para as empresas do setor de petróleo, as reservas são cruciais, pois são indicativos de previsão de produção e da receita potencial que será gerada no futuro (PEREIRA, 2004).

Os procedimentos de estimativa de reservas são feitos com base no cumprimento dos contratos de E&P vigentes em cada país e suas respectivas legislações. Esses procedimentos utilizam as melhores práticas da indústria de petróleo e é aplicado em todas as Unidades da empresa.

O processo de análise econômica utilizado para calcular o fluxo de caixa de cada campo petrolífero (ou projeto), e seu respectivo o ano de abandono, realiza a projeção das entradas e das saídas dos recursos financeiros de um campo em um determinado período de tempo e avalia até quando é economicamente viável continuar produzindo (RAVAGNANI, 2008).

Para realizar a Análise Econômica, os analistas do E&P utilizam uma ferramenta conhecida como PI (Programa de Indicadores). Esta ferramenta permite a avaliação dos projetos de investimentos realizados em cada campo, sendo alimentada com todas as variáveis necessárias, como curva de produção, investimentos, além de toda a complexidade tributária deste tipo de investimento (MENDES et.al, 2004).

No entanto, embora o PI seja uma ferramenta essencial para a análise econômica dos projetos, sua arquitetura atual é limitada às análises realizadas por campo. Dessa forma, exigindo um grande número de execuções tal quanto é o número de campos existentes na empresa.

Dessa forma, se o analista deseja projetar variações em algumas curvas de entrada como produção, investimento, custo e preço, é preciso realizar diversas chamadas ao PI,

uma para cada campo e para cada cenário. Sabendo que a empresa OX possui aproximadamente 1.280 campos (ANP, 2010), esse processamento demanda um tempo e um custo computacionais significativos.

Para auxiliar nos estudos de diferentes cenários, é proposto neste trabalho um Aproximador de Função que assegure que em cada projeção sejam gerados indicadores semelhantes aos retornados pelo PI. Isto será realizado modelando as variáveis significativas para o problema, considerando seus limites, forma de variação e as inter-relações correspondentes.

Este Aproximador de Função será construído utilizando-se técnicas de regressão linear múltipla, devido a sua facilidade na interpretação dos resultados gerados.

## 1.2. OBJETIVOS DO TRABALHO

Podemos definir para este trabalho, os seguintes objetivos (Figura 1):

- Construir um aproximador de função que gere um resultado semelhante ao gerado pelo PI, supondo que as variações serão realizadas a partir de um determinado Exercício de Análise Econômica. Para isto, será utilizado um modelo construído através de regressão linear.

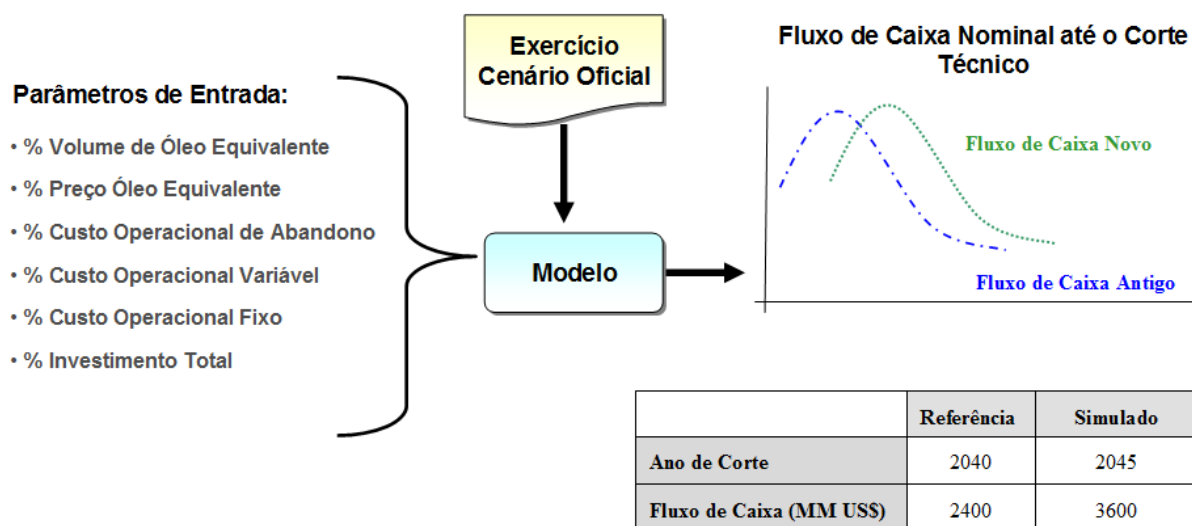


Figura 1: Objetivo do Trabalho

- Validação do modelo gerado comparando os resultados obtidos pelo Aproximador de Função com os reais gerados pelo PI.
- Minimização do tempo e custo computacional para projetar as diferentes variações nas curvas de entrada em cada campo petrolífero.

### **1.3. DESCRIÇÃO DO TRABALHO**

Este trabalho foi realizado em 6 etapas:

- Definição do Problema;
- Coleta dos dados e análise descritiva dos mesmos;
- Estudo da técnica de regressão linear;
- Modelagem do Aproximador de Função utilizando a técnica apresentada;
- Comparação do modelo com os resultados já existentes.

Inicialmente foi realizado um estudo sobre o problema em questão, principalmente relacionada às etapas do processo de Análise Econômica. Verificando quais as variáveis envolvidas e como são calculados os indicadores através do PI.

A fim de trabalhar com uma massa significativa de dados para obter um bom treinamento do modelo, foram simulados novos dados através de um simulador criado sobre o PI. Uma vez gerados esses dados, foi realizada uma análise descritiva dos mesmos verificando as correlações existentes entre os atributos, a necessidade de possíveis pré-processamentos e a identificação de variáveis relevantes.

Para aplicação da modelagem, foi realizado um estudo prévio dos conceitos relacionados à regressão. Uma vez consolidados estes conceitos, foi modelada uma equação alterando algumas técnicas de seleção do melhor modelo, assim como os seus respectivos parâmetros de configuração.

Depois de construído o modelo, foram realizados diversos experimentos, comparando os resultados gerados pelo modelo com os providos pelo PI.

### **1.4. ORGANIZAÇÃO DA MONOGRAFIA**

Este trabalho encontra-se estruturado da seguinte forma.

O capítulo 2 apresenta o problema relacionado para avaliação de projetos em campos petrolíferos, identificando as variáveis envolvidas e os resultados gerados.

Para um melhor entendimento do trabalho, é realizada, no capítulo 3, uma fundamentação dos conceitos que orientam a construção do aproximador de função baseado em técnicas de regressão.

No capítulo 4, é apresentado um estudo de caso, descrevendo desde o processo de pré-processamento dos dados e seleção de variáveis até a escolha do modelo propriamente dito. Neste mesmo capítulo, os resultados são apresentados comparando com os dados reais, a fim de demonstrar os ganhos em se utilizar a solução proposta.

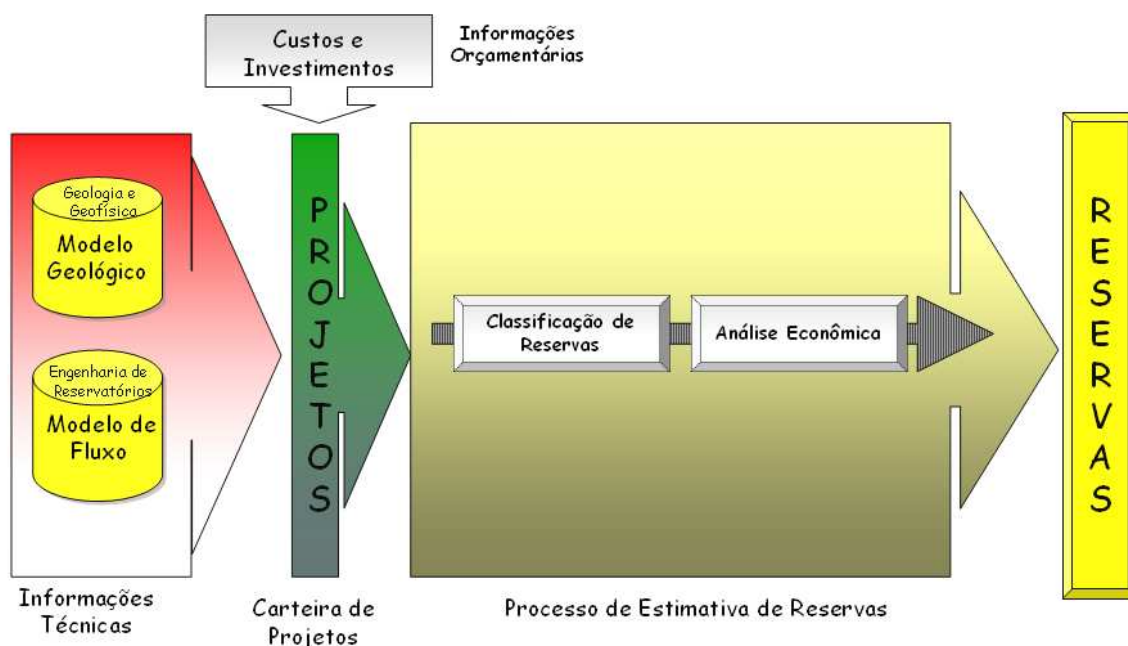
Finalmente, no capítulo 5 são destacadas as conclusões e contribuições oferecidas por esta solução, assim como identificados os possíveis trabalhos futuros.

## 2. ANÁLISE ECONÔMICA

As reservas<sup>1</sup> de petróleo são os ativos mais importantes de qualquer companhia que se proponha a atuar de forma competitiva no segmento de Exploração e Produção (E&P). A partir destas informações de reservas constrói-se boa parte da percepção de valor de mercado de uma companhia.

Como empresa de capital aberto, a OX divulga, anualmente, informações sobre suas atividades e seu desempenho financeiro. Dessa forma, a empresa busca que as informações publicadas sejam suficientes para que o público em geral e, em especial, seus acionistas atuais e potenciais e os analistas de crédito e investimentos possam fazer um julgamento apropriado do posicionamento da companhia na indústria petrolífera.

Para o processo de estimativa de reservas, é feito o que chamamos de análise econômica. A análise econômica é uma atividade cujo objetivo é calcular o fluxo de caixa dos campos (ou projetos) e, também o ano de abandono dos campos. A partir de curvas como produção, investimento, custo operacional e preços, além das premissas econômicas, são gerados o fluxo de caixa e os indicadores econômicos do elemento analisado (campo ou projeto), conforme pode ser visto na Figura 2.



**Figura 2:** Processo de Estimativa de Reservas

<sup>1</sup> Reservas – são os volumes de petróleo e gás natural que, com base na análise dos dados de geo-engenharia, se estima recuperar comercialmente de reservatórios descobertos e avaliados, cuja estimativa considere as condições econômicas vigentes, os métodos operacionais usualmente viáveis e os regulamentos instituídos pelas legislações petrolíferas e tributárias dos países onde a empresa opera (SOUZA, 2006).

Realizar uma análise econômica de um campo ou projeto equivale a estudar a viabilidade econômico-financeira deste objeto, estimando e analisando as perspectivas de desempenho financeiro do produto resultante. Pode-se dizer que através desta análise é avaliado o ano de corte de um campo, ou seja, até quando é economicamente viável continuar produzindo.

Para esclarecer o que seria o ano de corte, vamos supor que um determinado campo possui os seguintes anos de produção para um determinado exercício<sup>2</sup> (Figura 3):

Data de Produção	Produção (MMboe)	Preço (US\$)	Receita (MM US\$)	Dispendio (MM US\$)	Fluxo de Caixa Nominal (MMUS\$)
2010	0,69	40,00	27,73	10,00	17,73
2011	1,39	40,00	55,45	20,00	35,45
2012	1,79	40,00	71,67	30,00	41,67
2013	2,08	40,00	83,18	10,00	73,18
2014	2,30	40,00	92,10	20,00	72,10
2015	2,48	40,00	99,40	30,00	69,40
2016	2,64	40,00	105,56	10,00	95,56
2017	2,77	40,00	110,90	20,00	90,90
2018	2,89	40,00	115,61	30,00	85,61
2019	3,00	40,00	119,83	10,00	109,83
2020	3,09	40,00	123,64	20,00	103,64
2021	3,00	40,00	119,83	30,00	89,83
2022	2,89	40,00	115,61	10,00	105,61
2023	2,77	40,00	110,90	20,00	90,90
2024	2,64	40,00	105,56	30,00	75,56
2025	2,48	40,00	99,40	10,00	89,40
2026	2,30	40,00	92,10	20,00	72,10
2027	2,08	40,00	83,18	30,00	53,18
2028	1,79	40,00	71,67	80,00	-8,33
2029	1,39	40,00	55,45	60,00	-4,55
2030	0,69	40,00	27,73	40,00	-12,27

Figura 3: Cálculo do Ano de Corte

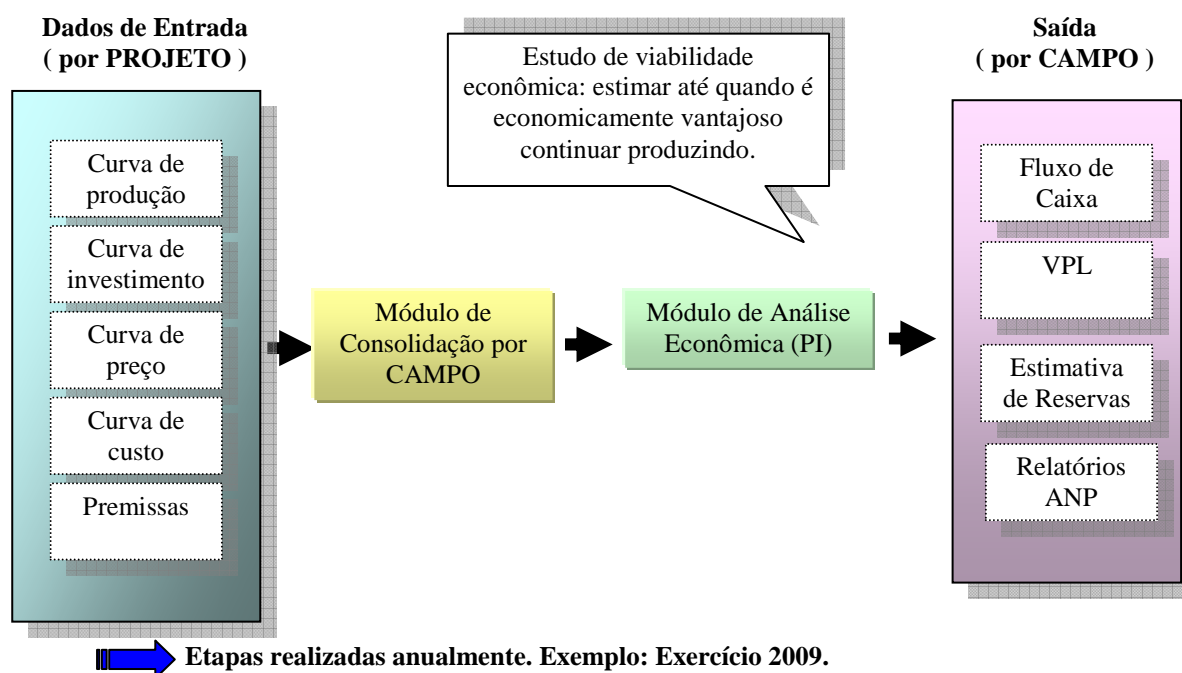
Analisando esta tabela de baixo para cima, podemos dizer, de forma simplificada, que o ano de corte seria o primeiro ano em que o fluxo de caixa passaria de negativo para positivo. Neste exemplo, vemos que este campo ainda possui 20 anos de produção até o seu abandono técnico<sup>3</sup>. No entanto, apenas é viável economicamente produzir até o ano 2027.

Uma vez definido o ano de corte, é realizada a etapa de composição das reservas para estimar quanto de reserva a empresa possui. Em outras palavras, é prevista a reserva da empresa desde o ano 2010 até o ano de 2027.

<sup>2</sup> Exercício: nomenclatura usada para identificar o ciclo de reserva anual em que é executada a análise econômica para estimar as reservas da empresa.

<sup>3</sup> Abandono técnico: corresponde ao último ano de produção de um determinado campo, ou seja, quando não há mais óleo para extrair.

O primeiro passo para a realização da análise econômica é a montagem do fluxo de caixa, isto é, a definição do fluxo de entradas e saídas de recursos financeiros durante o ciclo de vida planejado para o campo ou projeto (Figura 4).



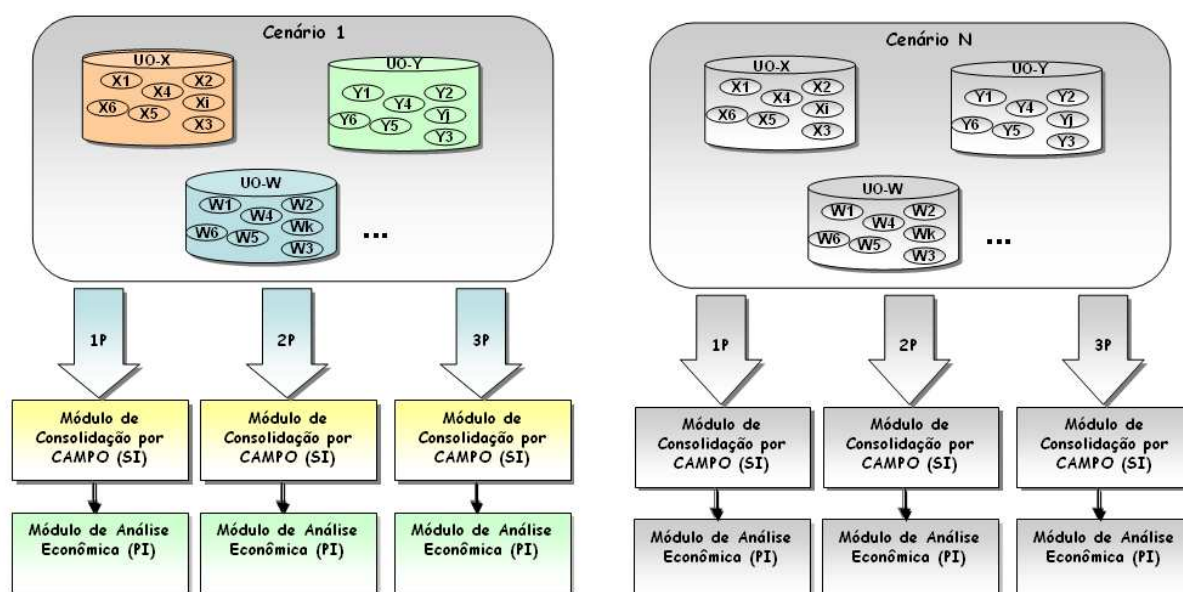
**Figura 4:** Análise Econômica

Um dos itens que está incluído nas saídas da empresa é a Participação Especial. O cálculo da Participação Especial é complexo, pois considera diversas regras que afetam e muito no fluxo de caixa final.

Atualmente, para realizar tanto o cálculo da Participação Especial como o do Fluxo de Caixa são utilizadas duas ferramentas conhecidas como SR (Sistema de Reservas) e o PI. A primeira é principalmente responsável na parte de consolidação das informações, como por exemplo, consolidar curvas de produção de diversos projetos em apenas uma curva para um campo, onde esta curva será utilizada como entrada para o cálculo do fluxo de caixa deste campo.

Desta forma, vamos supor que temos na empresa  $n$  UO's, onde cada UO possui  $x$  campos e que cada campo possui associado  $w$  projetos<sup>4</sup>. Assim, será necessário realizar uma consolidação de suas curvas de produção, custo, investimento, dentre outras para cada campo e, conseqüentemente, cada UO (Figura 5).

<sup>4</sup> De forma genérica, a exploração de um campo de petróleo localizado sobre uma determinada área de concessão pode ser realizada através da implantação seqüencial de  $n$  projetos de E&P  $\{P(1), P(2), \dots, P(n)\}$ .



**Custo computacional e tempo de resposta elevados!**

**Figura 5:** Consolidação e Processamento dos Dados de Reservas

Agora suponha que frequentemente o especialista desejasse realizar essas execuções, alterando as curvas de entrada, variando-as apenas percentualmente para poder assim ter uma noção de quão próximo ou longe a empresa está das metas definidas para aquele ano. Ou muitas vezes, para tomar medidas preventivas a fim de aumentar as reservas da empresa.

Como por exemplo, analisar como ficariam as reservas se diminuísse em vinte por cento os custos operacionais. Isso seria uma forma antecipar os resultados do final do ano, tendo uma ferramenta para calcular a previsibilidade dos volumes de reserva e realizar o planejamento de ações para cumprimento das metas. Além de permitir que diversas simulações sejam executadas para uma análise de sensibilidade<sup>5</sup> com os itens que compõem o cálculo do fluxo de caixa e, portanto, a estimativa de reservas.

No entanto, nos dias de hoje, realizar essas simulações em diversos cenários, acarreta um alto custo e tempo computacional, já que é não apenas para um campo, mas também para diversas UO's.

Neste capítulo, serão descritos os conceitos que compõem o problema apresentado neste trabalho. Foi utilizado o site da ANP como referencia para este capítulo.

<sup>5</sup> A Análise de Sensibilidade avalia o impacto que produz a variação isolada de uma variável do projeto mantendo-se as demais variáveis constantes. Esta análise possibilita verificar a magnitude da influência de uma determinada variável do projeto sobre os indicadores econômicos, provendo uma informação adicional à análise do projeto.



## 2.1. FLUXO DE CAIXA NOMINAL

Por mais complexos e diferentes que sejam os projetos a serem analisados, todos eles podem ser representados por um fluxo de caixa, ou seja, entradas e saídas de capital, divididas em várias categorias e dispostos em um horizonte temporal.

Devido às limitações práticas de se precisar os exatos momentos em que as despesas e receitas ocorrerão, a escala de tempo é dividida em anos e se estabelece uma mesma posição para os vetores monetários anuais, início, meio ou fim de período (Figura 6).

	Ano 0	Ano 1	Ano 2	Ano 3	Ano 4
Fluxo de caixa	-1.000,00	500,00	400,00	300,00	100,00

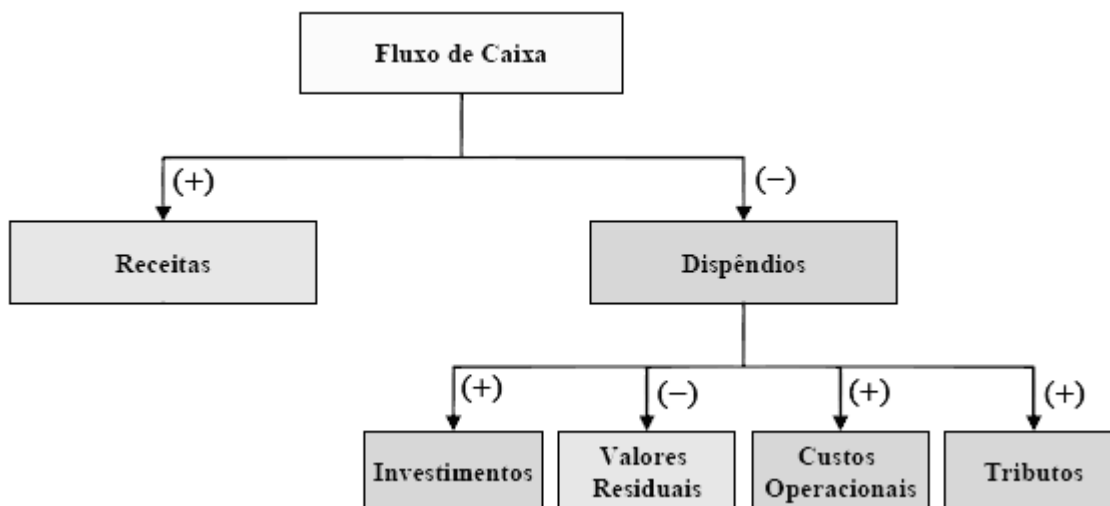
**Figura 6:** Fluxo de Caixa Nominal Anual

O fluxo de caixa nominal é um instrumento muito utilizado na tomada de decisões gerenciais de uma empresa. Atua principalmente na análise de desempenho financeiro e na previsão orçamentária através da coleta e a organização dos dados (MARTINOVICH, 1996).

O cálculo do fluxo de caixa deve considerar o impacto dos investimentos, custos operacionais, tributos e da receita. O resultado é o fluxo de caixa nominal, que apresenta o investimento, os custos e os ganhos adicionais previstos em relação às premissas supostas ao longo de sua vida econômica.

Lembrando que estas análises são realizadas de forma padronizada e parametrizada, de acordo com as premissas econômicas da empresa OX e o regime fiscal do país.

Os fluxos de caixa de um projeto de E&P referem-se às entradas e saídas de capital a serem auferidas com sua execução. Tipicamente, estes fluxos de capital podem ser classificados em algumas categorias ou componentes. Na Figura 7, serão detalhados os componentes formadores do fluxo de caixa.



**Figura 7:** Componentes do Fluxo de Caixa

### 2.1.1. Receitas

De acordo com a natureza do projeto, a receita pode ser vista como:

- Direta, quando gerada através da venda de um produto ou serviço no mercado;
- Indireta, quando resulta de uma redução de custo, ou seja, quando gera uma economia para a Companhia como um todo, ou quando resulta de perdas evitadas, não relacionadas a questões contratuais.

Em geral, podemos identificar a receita como as entradas de capital auferidas pela execução do projeto e defini-las como:

$$\text{Receita} = \text{Volume Produzido} * \text{Preço}$$

**Equação 1:** Receita

No entanto, existem alguns tipos diferentes, dependendo se o volume avaliado é o tributado ou não e se a receita é líquida ou não. Assim, teremos os seguintes tipos de receita:

- Receita Líquida: a receita líquida da produção é o resultado da diferença entre a receita bruta da produção do campo e as deduções autorizadas.

- **Receita Bruta:** A receita bruta da produção de um campo no período base é determinada pela soma das receitas de óleo e gás, dados seus volumes de produção fiscalizados e preços de referência.

As receitas tributadas são aquelas que levam em consideração o preço tributado.

### **2.1.2. Curvas de produção de petróleo e de oferta de gás natural**

As previsões de produção de óleo e de gás natural são definidas com base em estudos geológicos, de reservatórios e de plantas de produção.

No caso do gás natural, são abatidos o consumo próprio e a reinjeção e é adicionada a compra de gás importado, obtém-se então a oferta de gás para processamento nas Unidades de Processamento de Gás Natural (UPGNs), para uso industrial, residencial ou veicular e para utilização nas termelétricas.

No caso do petróleo, parte deve vir a abastecer o parque de refino no país e parte pode vir a ser exportada. O volume máximo para exportação é indicado pelo Marketing e Comercialização.

### **2.1.3. Preços de petróleo**

A estimativa de preços para petróleos importados baseia-se em abordagens estruturais e de tendências do mercado internacional.

O preço do petróleo é definido para o cenário de referência da Companhia e considerando os critérios de robustez. O Abastecimento, Marketing e Comercialização definem os preços de importação dos demais petróleos e os preços de exportação dos petróleos nacionais, coerentes com o cenário de robustez. A moeda base desses preços é o dólar americano de valor constante no tempo.

### **2.1.4. Preços do gás natural**

A lógica de formação de preços de gás natural deve respeitar os compromissos da empresa OX de rentabilidade para os acionistas, responsabilidade social e desenvolvimento sustentável.

Órgãos governamentais, por sua vez, desejam que os mercados, sejam eles competitivos ou regulados, tenham preços que sinalizem corretamente as decisões de produção e consumo, de forma a maximizar o bem-estar da sociedade.

Admitindo que o governo seja bem sucedido na implementação de suas políticas, os preços que se formarão nos mercados resultarão de um equilíbrio de longo prazo entre

oferta e demanda. A curva de oferta é obtida através da ordenação crescente dos custos marginais das diversas fontes de suprimento, incluindo a produção local do E&P e importação, seja via gasoduto ou gás natural liquefeito (GNL). A curva de demanda é estimada a partir do cálculo do valor máximo que o consumidor está disposto a pagar pelo gás, em substituição a outros energéticos (GLP, diesel, gasolina, óleo combustível, entre outros), nos diversos segmentos de mercado (industrial, residencial, automotivo e termelétrico).

### 2.1.5. Dispêndios

A categoria dos dispêndios modela as saídas de capital auferidas pela execução do projeto. Os dispêndios totais são calculados da seguinte maneira:

$$\begin{aligned}
 \textit{Dispêndio Total} &= \textit{Investimentos} \\
 &\quad - \textit{Valores Residuais} \\
 &\quad + \textit{Custos Operacionais} \\
 &\quad + \textit{Tributos Governamentais}
 \end{aligned}$$

**Equação 2:** Dispêndio Total

Em uma visão geral, os Valores Residuais são o remanescente dos Investimentos ao final da vida econômica do projeto. Por esta razão os Investimentos e Valores Residuais podem ser vistos como um único componente dos dispêndios.

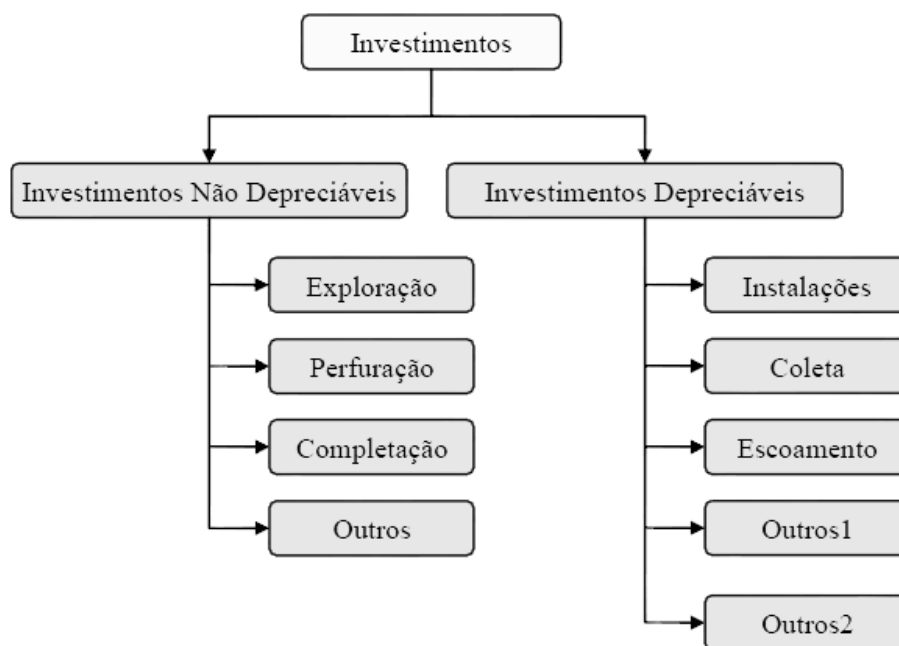
### 2.1.6. Investimentos

O componente dos investimentos, também conhecido como CapEx (*Capital Expenditures*), corresponde aos dispêndios realizados a fim de implementar e colocar o projeto em operação e encerrá-lo. Existem diversos tipos de investimentos, tais como: estudos de viabilidade, projetos de engenharia, obras civis, equipamentos e materiais, serviços de instalação e montagem, serviços de perfuração e de completação de poços, despesas financeiras durante a execução da obra, despesas pré-operacionais, arrasamento de poços, abandono de campos e acondicionamento ambiental (RAVAGNANI, 2008). Tipicamente, os investimentos são classificados como Investimentos Não Depreciáveis e Investimentos Depreciáveis:

<i>Investimentos</i> =	<i>Investimentos Não Depreciáveis</i> + <i>Investimentos Depreciáveis</i>
------------------------	--

**Equação 3:** Investimentos

Os Investimentos Depreciáveis, como terrenos, equipamentos e instalações, geram quase sempre Valores Residuais após seu uso no projeto. Já os Investimentos Não Depreciáveis, como estudos de viabilidade, serviço de instalações e montagem, dificilmente geram Valores Residuais. Ambos os tipos de investimentos, ainda podem ser classificados nos seguintes grupos (Figura 8):



**Figura 8:** Classificação dos Investimentos

### 2.1.7. Depreciação

A depreciação mostra a perda progressiva do valor, legalmente contabilizável, dos ativos imobilizados. O valor da depreciação é lançado no fluxo de caixa para abater o lucro tributável e seu cálculo segue a legislação da contabilidade societária. No fluxo de caixa do projeto, a depreciação não deve ser lançada como um custo, pois não representa um

dispêndio de caixa, e sim um benefício fiscal usado para cálculo do imposto de renda. Esse benefício pode ser obtido a partir da entrada em operação.

O tempo de depreciação é determinado pelas autoridades fiscais, geralmente em função da vida útil média dos diversos tipos de equipamentos. Recomenda-se que a depreciação para fins fiscais seja igual à depreciação contábil registrada em Livro. Em uma unidade industrial, entretanto, a depreciação costuma ser calculada em bloco, pela vida útil das instalações e equipamentos, e não dos seus componentes isolados.

Existem vários métodos de cálculo da depreciação: linha reta, soma dos dígitos dos anos, balanço declinante. No Brasil, a legislação vigente geralmente só admite o método de linha reta, que consiste em um valor anual constante resultante da divisão do valor do ativo fixo pelo tempo de depreciação (ANP, 2011).

### **2.1.8. Valores Residuais**

Caso alguns itens de investimento em determinado projeto tenham utilidade além da vida econômica do projeto (por exemplo, um equipamento, uma instalação), devem-se valorar esses itens de modo coerente com o seu custo de oportunidade, e considerar essas estimativas como receita tributável adicional ao final do projeto.

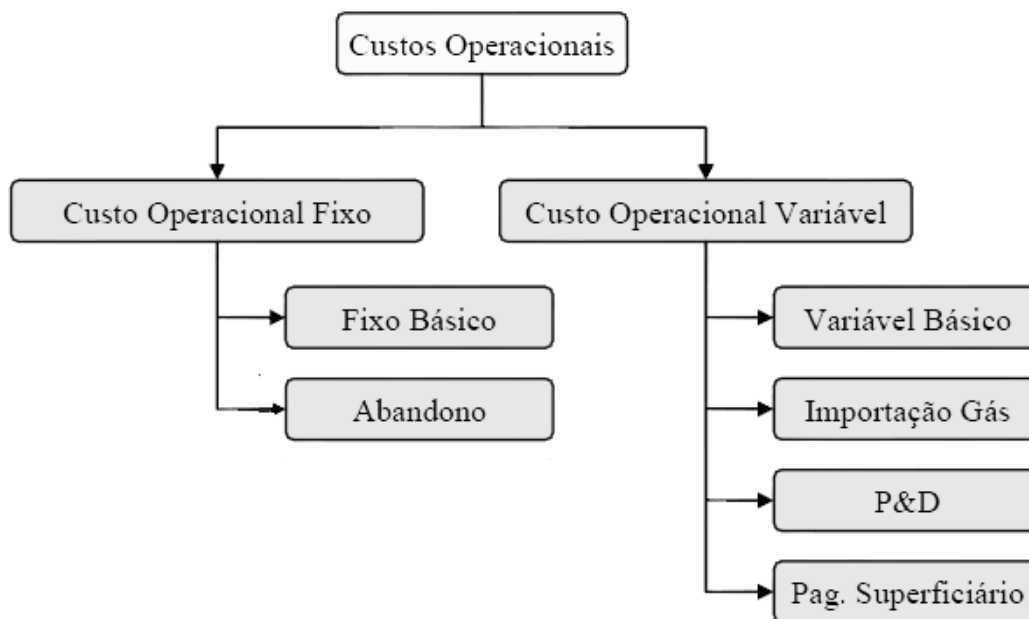
No caso em que não se disponha do valor de mercado do bem no estado em que se encontra ao fim da vida econômica do projeto, pode-se adotar como valor residual uma parcela do bem em estado novo.

Caso a vida útil do bem seja menor que a vida econômica do projeto, devem-se prever novos investimentos para a reposição desse bem ou a sua reforma, visando o prolongamento da vida útil do bem.

Quando um projeto apresentar vida econômica igual ou superior ao tempo de depreciação dos bens envolvidos não haverá necessidades de ajustes no valor residual.

### **2.1.9. Custos Operacionais**

O componente dos custos operacionais, também conhecidos como OpEx, corresponde aos dispêndios decorrentes durante a vida útil do projeto. Os custos são agrupados em duas categorias: Custos Operacionais Fixos e Custos Operacionais Variáveis (Figura 9).



**Figura 9:** Custos Operacionais

Os custos operacionais fixos compreendem as despesas independentes do nível de produção, tais como: gastos com seguros de instalações e aluguéis. Já os custos operacionais variáveis têm correlação direta com o nível de produção, a exemplo de gastos com produtos químicos, transporte de óleo, entre outros.

### 2.1.10. Tributos – Participações Governamentais

A consideração de todos os dados e informações referentes aos tributos que incidirão sobre o investimento, as receitas e os custos operacionais de um projeto é fundamental na Análise Empresarial, pois a determinação da carga tributária a ser incorrida é imprescindível para o cálculo correto dos indicadores econômicos.

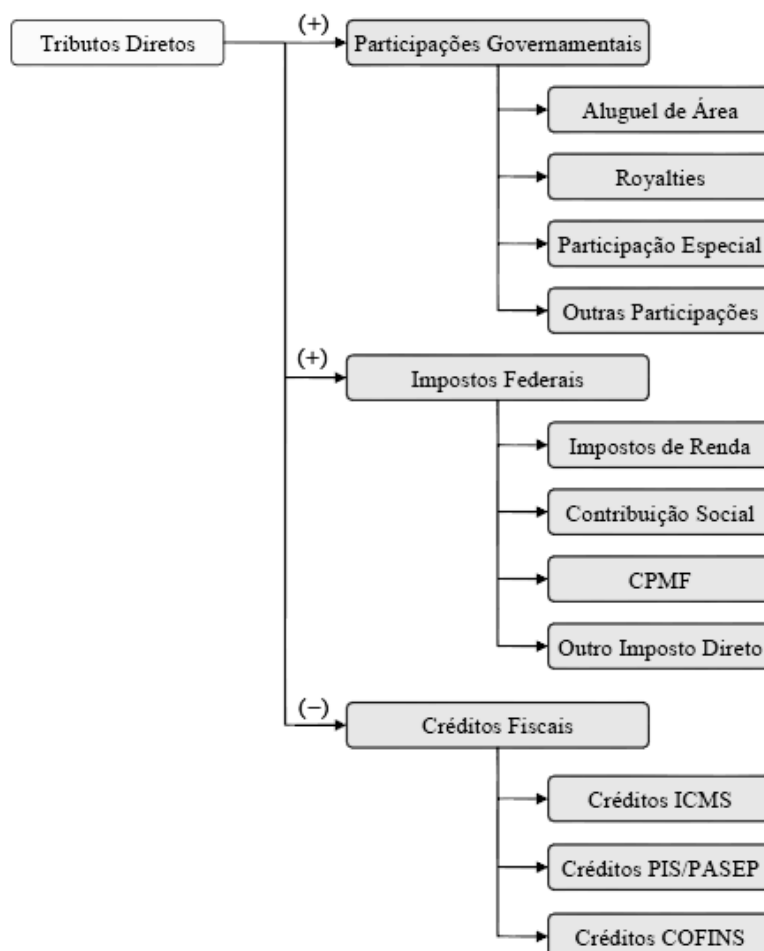
O conhecimento da legislação tributária e a avaliação do seu impacto sobre os indicadores do projeto inserem-se no contexto mais amplo de Planejamento Tributário da Companhia.

A questão tributária em análise de projetos é muito complexa por várias razões. Em primeiro lugar, a legislação é extremamente dinâmica, implicando, muitas vezes, mudanças não só nas alíquotas dos tributos como também na sua base de incidência.

Em segundo lugar, por vezes, a legislação não é clara, nem didática, implicando na necessidade de análise por um especialista da área, como, por exemplo, o caso de incentivos fiscais.

As participações governamentais são pagamentos a serem efetuados pelos concessionários das atividades de exploração e produção de petróleo e gás natural, compreendendo: bônus de assinatura; *royalties*; participação especial e ocupação ou retenção de área. O cálculo dessas participações é feito com base em preços de referência.

O componente dos tributos governamentais refere-se aos Impostos Diretos incidentes na produção (Figura 10).



**Figura 10:** Tributos Diretos

### 2.1.11. Bônus de assinatura

Montante ofertado para obter a concessão e pago no ato da assinatura em parcela única pelo licitante vencedor, não podendo ser inferior ao valor mínimo fixado pela Agência Nacional do Petróleo – ANP – no edital de licitação (ANP, 2011).

O bônus é passível de dedução do IRRF (Imposto de Renda Retido na Fonte) a ser pago de acordo com as unidades produzidas na área em cada ano em relação ao total de produção estimado para a área durante todo o período da concessão.



### **2.1.12. Royalties**

Os Royalties podem ser definidos como uma forma de compensação financeira devida pelos concessionários de exploração e produção de petróleo e de gás natural paga mensalmente e calculada pela multiplicação do equivalente a um percentual do volume total de produção de petróleo e gás natural, incluídos o consumo próprio e a queima de gás, excluída a reinjeção, pelos respectivos preços de referência:

- Preço de referência do petróleo - média ponderada dos preços de venda, desde que maior ou igual ao mínimo estabelecido pela ANP.
- Preço de referência do gás natural - média ponderada dos preços de venda estabelecidos em contrato, excluindo tributos sobre a venda e deduzindo as tarifas de transporte até os compradores.

### **2.1.13. Pagamento pela ocupação ou retenção de área**

Apurado a cada ano civil e pago em janeiro do ano seguinte, prevê valores por km<sup>2</sup> que são fixados no edital e no contrato de concessão, sendo aplicáveis nas fases de exploração, desenvolvimento e produção. O valor é baseado nas características geológicas e na localização da bacia sedimentar.

### **2.1.14. Participação especial**

A Participação Especial (PE) foi introduzida pela primeira vez na legislação brasileira como Compensação Financeira Extraordinária estabelecida pela Lei 9.478/97 (Lei do Petróleo), Decreto 2.705/98, e regulamentada pelas Portarias ANP 10/99 e 102/99. Trata-se de um pagamento adicional aos Royalties para os campos com grandes volumes de produção ou de grande rentabilidade (ANP, 1999).

A compensação financeira extraordinária nos casos de grandes volumes de produção ou de grande rentabilidade pela Agência Nacional do Petróleo (ANP) é calculada com base em alíquotas progressivas sobre a receita líquida de produção trimestral: receita bruta menos royalties, investimentos na exploração e demais investimentos não depreciáveis, custos operacionais, depreciação e tributos diretamente relacionados à operação do campo (ANP, 2011).

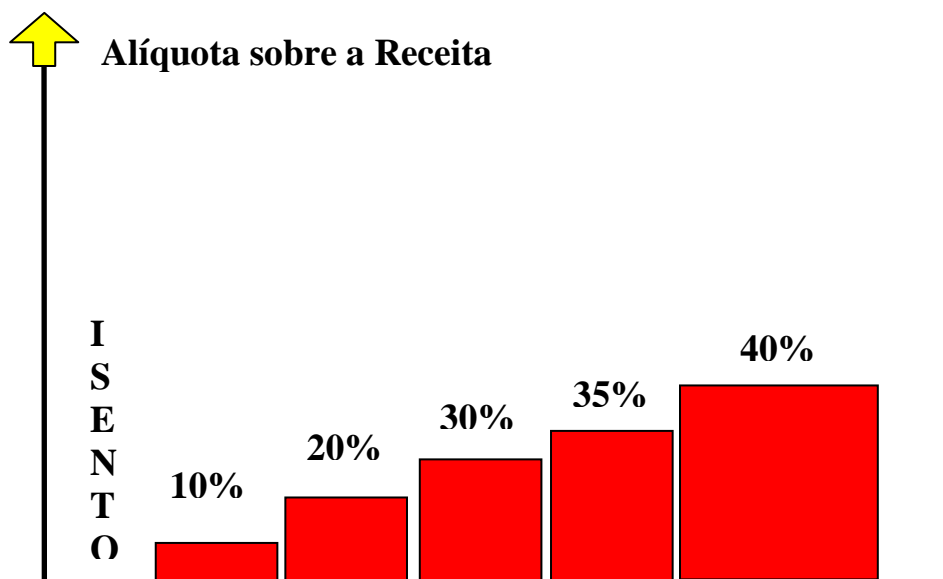
A determinação da PE é feita mediante aplicação de alíquotas progressivas, de 10%, 20%, 30%, 35% e 40%, em função do volume trimestral de produção, a partir de um volume limite de isenção. A alíquota efetiva é calculada em função de três itens:

- Ano de produção - 4 cenários: (i) 1º ano, (ii) 2º ano, (iii) 3º ano, e (iv) 4º ano e seguintes;

- Local onde ocorre a lavra - 3 cenários: (i) terra, (ii) mar com lâmina d'água até 400 m, (iii) mar com lâmina d'água superior a 400 m; e

- Volume de produção, a alíquota efetiva é crescente com a produção.

Conforme ilustra a Figura 11.



Vazão diária no 4º ano, em diante (mil boepd)	A1	A2	A3	A4	A5	Terra, rio, lago...
	B1	B2	B3	B4	B5	Plat.Cont.Ida < 400m
	C1	C2	C3	C4	C5	Plat.Cont.Ida > 400m
Vazão diária no 3º ano (mil boepd)	D1	D2	D3	D4	D5	Terra, rio, lago...
	E1	E2	E3	E4	E5	Plat.Cont.Ida < 400m
	F1	F2	F3	F4	F5	Plat.Cont.Ida > 400m
Vazão diária no 2º ano (mil boepd)	G1	G2	G3	G4	G5	Terra, rio, lago...
	H1	H2	H3	H4	H5	Plat.Cont.Ida < 400m
	I1	I2	I3	I4	I5	Plat.Cont.Ida > 400m
Vazão diária no 1º ano (mil boepd)	J1	J2	J3	J4	J5	Terra, rio, lago...
	K1	K2	K3	K4	K5	Plat.Cont.Ida < 400m
	L1	L2	L3	L4	L5	Plat.Cont.Ida > 400m

**Figura 11:** Participação Especial

Como pode ser observado na Portaria ANP Nº 10, DE 13/01/99 (ANP, 1999), realizar o cálculo da Participação Especial para cada campo anualmente é uma tarefa muitas vezes complexas considerando que a legislação tributaria é extremamente dinâmica e que a estimativa adequada da carga tributaria incidente em um dado projeto requer uma boa compreensão de vários aspectos tributários (fatos geradores dos tributos, base de calculo, forma de calculo e isenções possíveis). Sendo, portanto, um dos componentes mais difíceis de estimar durante a determinação do fluxo de caixa nominal.

### 3. MODELO

Este trabalho é composto principalmente de 3 módulos (Figura 12) no processo de construção do modelo.

Inicia-se com a parte de pré-processamento dos dados que engloba a normalização e seleção de variáveis através de métodos de correlação. E ao final, é utilizada uma regressão linear múltipla para modelar o problema em questão.

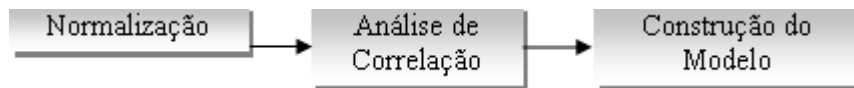


Figura 12: Modelo

#### 3.1. NORMALIZAÇÃO

Em casos em que as escalas das variáveis numéricas são muito diferentes, utilizar a normalização é uma solução bastante adequada para realizar uma equalização dos domínios, de forma que as variáveis com valores de maior ordem de magnitude que outras não dominem ou distorçam os pesos de um modelo.

Existem vários tipos de normalizações, tais como:

- Normalização linear
- Normalização por desvio padrão
- Normalização pela soma dos elementos
- Normalização pelo valor máximo dos elementos
- Normalização por escala decimal

Devido à simplicidade na sua aplicação e a distribuição das variáveis do nosso problema, optou-se neste trabalho em aplicar a normalização linear que pode ser definida pela equação (DEVORE, 2006):

$$x' = (x - x_{min}) / (x_{max} - x_{min})$$

Equação 4: Normalização Linear

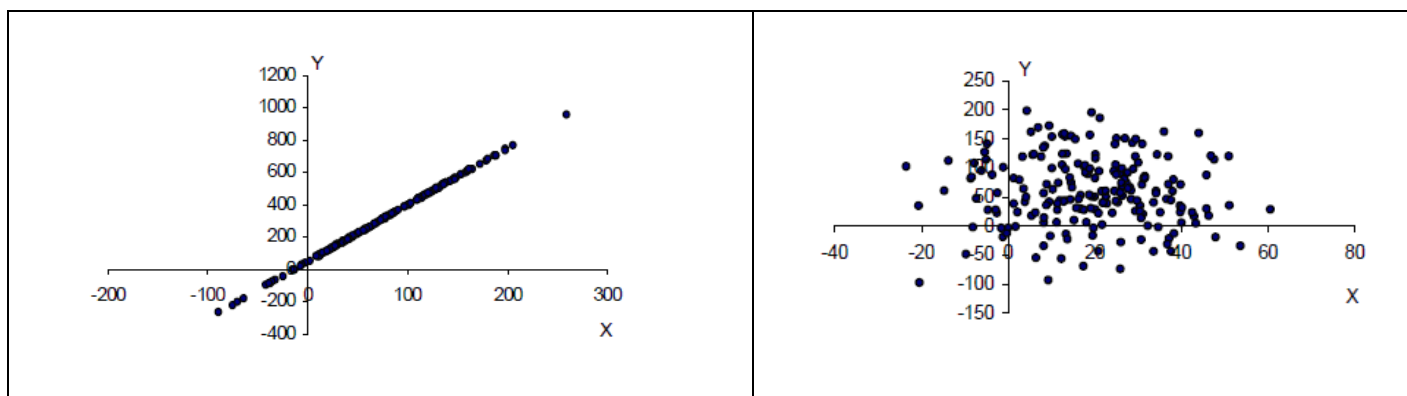
### 3.2. SELEÇÃO DE VARIÁVEIS: CORRELAÇÃO

Os métodos de seleção de variáveis buscam encontrar as variáveis minimamente correlacionadas que contenham informações relacionadas ao parâmetro de interesse. Assim, a modelagem pode ser realizada com base nessas variáveis a fim de construir modelos mais simples, robustos, eficientes e fáceis de interpretar.

Diferentes formas de correlação podem existir entre as variáveis. O caso mais simples e mais conhecido é a correlação simples, envolvendo duas variáveis, X e Y. A relação entre duas variáveis será linear quando o valor de uma pode ser obtido aproximadamente por meio da equação da reta. Assim, é possível ajustar uma reta da forma  $Y = \alpha + \beta X$  aos dados. Neste caso, a correlação é linear simples.

Entretanto, quando não for possível o ajuste da equação anterior, não significa que não existe correlação entre elas, poderá haver correlação não linear entre as mesmas.

Uma forma simples de verificar o tipo de correlação existente entre duas variáveis é através do gráfico chamado diagrama de dispersão. Trata-se de um gráfico onde são representados os pares  $(X_i, Y_i)$ ,  $i = 1, 2, \dots, n$ , onde  $n$  = número total de observações. Os gráficos a seguir representam alguns exemplos de “diagrama de dispersão” entre as variáveis X e Y.



**Figura 13:** Correlação Linear Positiva Perfeita entre X e Y

**Figura 14:** Correlação Linear Nula entre X e Y

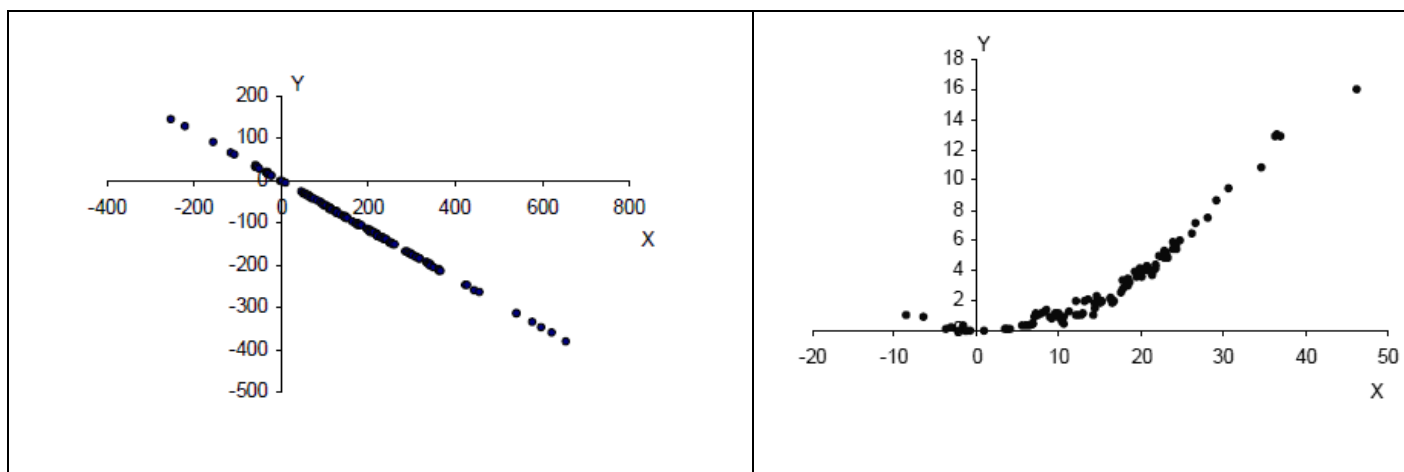


Figura 15: Correlação Linear Negativa Perfeita entre X e Y

Figura 16: Correlação Não Linear entre X e Y

No uso estatístico geral, a correlação linear indica a magnitude e a direção da relação linear entre duas variáveis aleatórias. Existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados. O mais conhecido é o coeficiente de correlação de Pearson ( $r$ ), o qual é obtido dividindo a covariância de duas variáveis pelo produto de seus desvios padrão (DEVORE, 2006).

$$r = \frac{\text{Cov}(X, Y)}{s_X \cdot s_Y}$$

onde

$\text{Cov}(X, Y)$  = covariância entre as variáveis X e Y  
 $s_X$  = desvio padrão da variável X  
 $s_Y$  = desvio padrão da variável Y

Equação 5: Coeficiente de Pearson

Os valores do coeficiente  $r$  variam entre -1 e 1. Quanto mais extremo é o valor (positivo ou negativo), maior é o grau de associação linear entre duas variáveis (Figura 17).

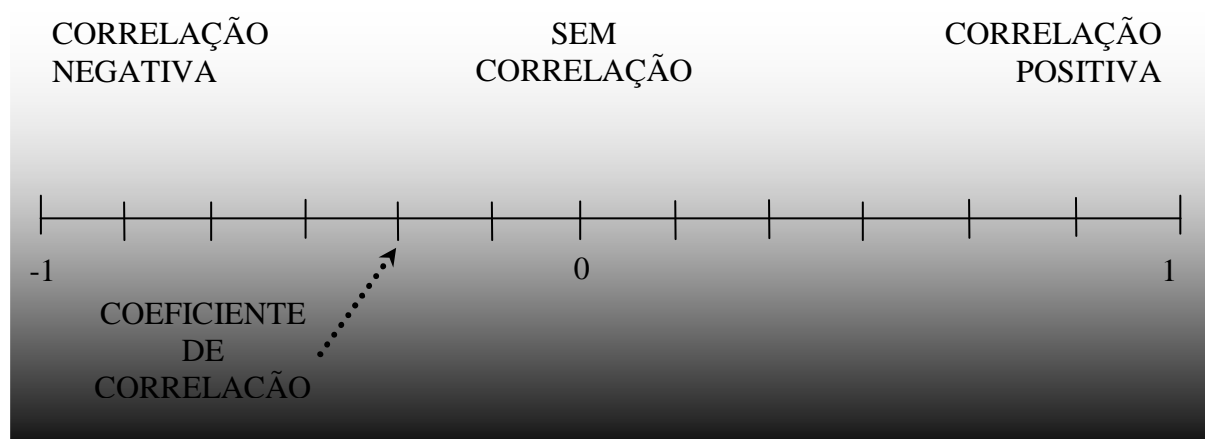


Figura 17: Correlação

A Figura 17 pode ser interpretada da seguinte forma:

- Se  $r = 1$ , então existe uma correlação total e ela é positiva, ou seja, à medida que aumentam os valores de  $X$ , aumentam proporcionalmente os valores de  $Y$ .
- Se  $r = -1$ , então há também uma correlação total, mas neste caso, negativa, isto é, à medida que aumentam os valores de  $X$ , os valores de  $Y$  diminuem.
- Se  $r = 0$ , então não existe correlação.

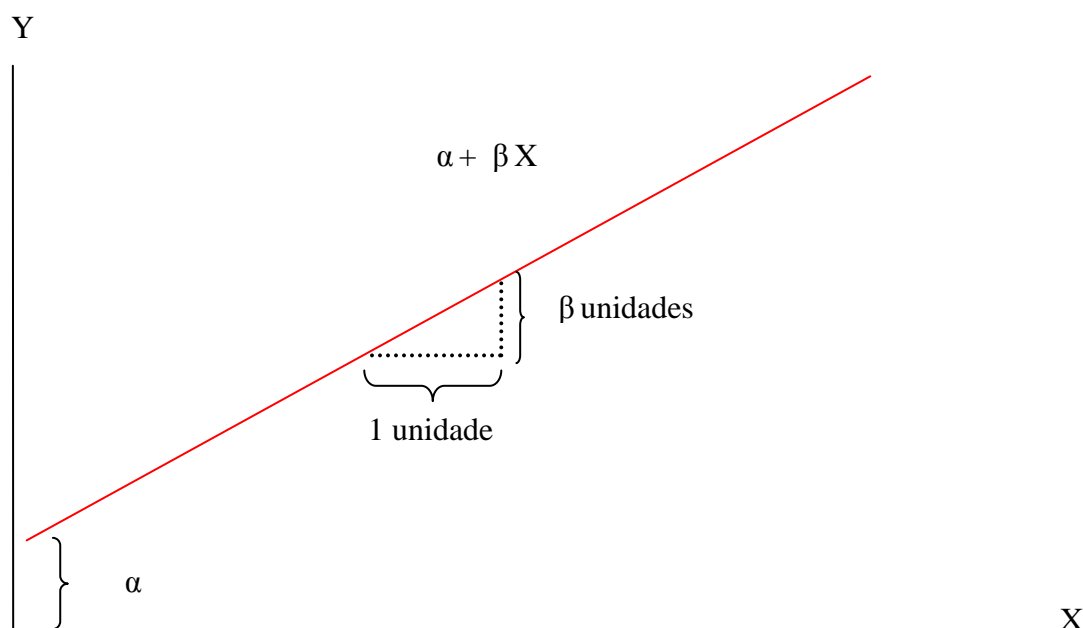
Pode ocorrer, ainda, a situação onde se tem dois conjuntos de variáveis, um composto por uma variável ( $Y$ ) e o outro com as variáveis ( $X_1, X_2, \dots, X_p$ ), e se deseja analisar a correlação entre a variável  $Y$  e a variável  $X_i$ ,  $i = 1, 2, \dots, p$ . Neste caso a correlação é chamada de múltipla e pode ser expressa pelo hiperplano  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ , admitindo relação linear entre  $Y$  e  $X_1, X_2, \dots, X_p$ .

### 3.3. MODELAGEM

#### 3.3.1. Regressão Linear Simples

Os coeficientes de correlação indicam somente a existência ou não de algum tipo de relacionamento entre variáveis. Para investigar a forma desse relacionamento, o método mais apropriado é a análise de regressão, onde a relação pode ser expressa sob forma matemática, por meio de uma equação que interligue as variáveis.

Em termos gerais um modelo de regressão linear simples tem como objetivo avaliar a significância da variável preditora  $X$  (ou explicativa) na explicação da variabilidade ou do comportamento da variável resposta  $Y$  (ou alvo) (DEVORE, 2006) (Figura 18).



**Figura 18:** Regressão Linear Simples

O relacionamento entre a variável resposta Y e a preditora X pode ser caracterizado pela equação (DEVORE, 2006):

$$Y = \alpha + \beta X + \varepsilon$$

**Equação 6:** Regressão Linear Simples

Onde através do método dos mínimos quadrados<sup>6</sup>, podem ser estimar:

- $\alpha$ : parâmetro conhecido como intercepto que corresponde ao valor da variável resposta quando a variável preditora for zero
- $\beta$ : parâmetro de angulação que corresponde a magnitude da mudança na variável resposta dado 1 unidade de mudança da variável preditora.
- $\varepsilon$ : o termo que representa o erro.

Segundo Washington, Karlaftis e Mannering (2003), existem algumas premissas que precisam ser verificadas. Quando uma destas premissas não é satisfeita, ações corretivas podem ser adotadas e, em alguns casos, outros modelos podem ser adotados. As premissas são:

<sup>6</sup> Método dos mínimos quadrados: consiste em minimizar os erros entre o dado real e a estimativa da linha de regressão. Este método pode ser visto em mais detalhes em Devore, 2006.



- A resposta deve ser contínua, isto é, não se pode utilizar regressão linear para prever uma variável ordinal.
- A relação entre as variáveis deve ser linear, no entanto, o autor enfatiza que, em alguns casos, através de transformações matemáticas as variáveis não lineares podem ser modificadas para lineares.
- As observações devem ser independentes e a amostragem randômica para que se possa fazer inferência sobre a população de interesse.

No entanto, como o problema exposto neste trabalho é composto por mais de uma variável explicativa, será necessário utilizar um regresso múltipla, como será descrito na seção 3.3.2.

### 3.3.2. Regressão Linear Múltipla

A regressão linear múltipla é a extensão da regressão linear simples, onde existe uma única variável dependente e mais de uma variável independente.

A finalidade das variáveis independentes adicionais é melhorar a capacidade de predição em confronto com a regressão linear simples. O objetivo é obter o mais alto relacionamento explanatório com o mínimo de variáveis independentes, sobretudo em virtude do custo na obtenção de dados para muitas variáveis e também pela necessidade de observações adicionais para compensar a perda de graus de liberdade decorrente da introdução de mais variáveis independentes.

Segundo Snedecor e Cochran (1989), a regressão linear múltipla pode ser descrita como:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

**Equação 7:** Regressão Linear Múltipla

Onde:

- $\alpha$  = intercepto do eixo y e representa o valor de Y quando todos os valores de X são zero.
- $\beta_i$  = coeficiente angular da i-ésima variável, ou seja, é o coeficiente de regressão para cada X e estabelece a quantia que Y varia com cada mudança unitária de  $X_i$ . É calculado utilizando o método dos mínimos quadráticos.
- $k$  = número de variáveis independentes.
- $\varepsilon$ : o termo que representa o erro.

Enquanto uma regressão simples de duas variáveis resulta na equação de uma reta, um problema de três variáveis implica num plano, e um problema de  $k$  variáveis implica em um hiperplano.

Podemos inferir que assim como na regressão linear simples, a resposta da regressão deve ser contínua, ou seja, não se pode utilizar regressão linear para prever uma variável ordinal.

Além disso, as observações devem ser independentes e a amostragem randômica para que se possa utilizar a regressão para fazer inferência sobre a população de interesse.

Não obstante seu grande potencial, o método regressão linear múltipla apresenta alguns problemas que limitam sua aplicação. Um deles é que o número de amostras deve ser igual ou superior ao número de variáveis. Uma vez que o modelo consiste na resolução de um sistema de equações lineares simultâneas, a essa condição necessita ser satisfeita. Caso contrário, o sistema torna-se indeterminado.

Outro problema importante em muitas aplicações da análise de regressão envolve selecionar o conjunto de variáveis preditoras a ser usado no modelo, a fim de diminuir a complexidade do modelo e, conseqüentemente, o tempo computacional demandado.

Algumas vezes, a experiência prévia ou considerações teóricas podem auxiliar o analista a selecionar um conjunto apropriado de variáveis mais informativas e não redundantes para o modelo. Para alcançar esse objetivo, várias técnicas de seleção de variáveis têm sido propostas na literatura, como a que vimos na seção 3.2.

Além disso, é possível também selecionar as variáveis que irão compor uma regressão linear múltipla através dos métodos: *stepwise*, *forward* e *backward*.

### **3.3.3. Regressão *Stepwise*, *Forward* e *Backward***

O método *forward* é um método iterativo que começa com uma variável e, progressivamente, adiciona mais variáveis ao modelo de regressão até que um critério de parada seja satisfeito. A primeira variável a ser inserida no modelo é aquela com a maior correlação com a variável resposta. Posteriormente, a cada interação são avaliadas as variáveis preditivas pelo teste  $F^7$ , sendo que são incluídas, sequencialmente, as variáveis com o maior valor de  $F$ , sendo este superior ao  $F$ -crítico definido pelo especialista. Uma vez a variável inserida no modelo, ela não é mais retirada (SAS, 2010).

No método *backward*, inicia-se com a construção de um modelo de regressão com todas as variáveis disponíveis e, subseqüentemente, variáveis são retiradas e o efeito dessa

---

<sup>7</sup> O teste  $F$  é utilizado para testar se existem informações úteis sobre a variável dependente em qualquer uma das variáveis preditoras (Devore, 2006).

eliminação é avaliado, da mesma forma que no método *forward*. As variáveis com valores de F menores que F-crítico são descartadas do modelo. E uma vez que a variável é removida, ela não retorna mais para o modelo (SAS, 2010)

O método de regressão *stepwise* é um procedimento padrão para seleção de variáveis que combina dois outros métodos, a seleção *forward* e a eliminação *backward*.

O procedimento constrói iterativamente uma seqüência de modelos de regressão pela adição ou remoção de variáveis em cada etapa. Ele começa formando um modelo com uma variável preditora que tenha a mais alta correlação com a variável de resposta, como no método *forward* e incorpora um mecanismo de eliminação de variáveis igual ao método de *backward*. Assim, apesar de uma variável ter entrado no modelo, não necessariamente ela irá permanecer no mesmo (SAS, 2010).

## 4. ESTUDO DE CASO: EMPRESA OX

Neste capítulo serão descritos o processo de geração dos dados analisados, o escopo escolhido, além das técnicas de regressão linear e seleção de variáveis utilizadas neste trabalho.

Para estas etapas (Figura 19), foram utilizadas as ferramentas Enterprise Guide e Enterprise Miner da empresa SAS.



Figura 19: Arquitetura

### 4.1. ESCOPO

Como existe uma gama de Unidades Operacionais e, conseqüentemente, de campos envolvidos, foi definido um escopo reduzido, mas que fosse suficiente para apresentar uma proposta ao problema exposto com as técnicas descritas no capítulo 3.

Desta forma, foi utilizado um campo terrestre que nomearemos de C1 cujas informações foram de um determinado ciclo econômico, grupo de reserva<sup>8</sup>, rodadas pré Imposto, dados antes do Corte (até o fim da produção) e 100% da empresa OX. Lembrando que todas as simulações serão a partir de uma mesma carteira de projetos de um exercício pré-estabelecido.

### 4.2. DADOS HISTÓRICOS

Conforme descrito no capítulo 3, todo o ano a empresa OX realiza o que chamamos de Análise Econômica na aplicação PI. Esta análise resulta em um conjunto de indicadores anuais para cada ano futuro de produção de um determinado campo.

Assim, para construir uma base que fosse suficiente para modelarmos o problema, foi construído um simulador sobre as curvas produção, investimento, preços e custos operacionais do campo C1, usadas como entrada para a aplicação PI (Figura 20).

---

<sup>8</sup> Grupo de reserva: originalmente existem 3 grupos de reserva, conhecidos como: 1P = Provadas, 2P = Provadas + Prováveis e 3P = Provadas + Prováveis + Possíveis.

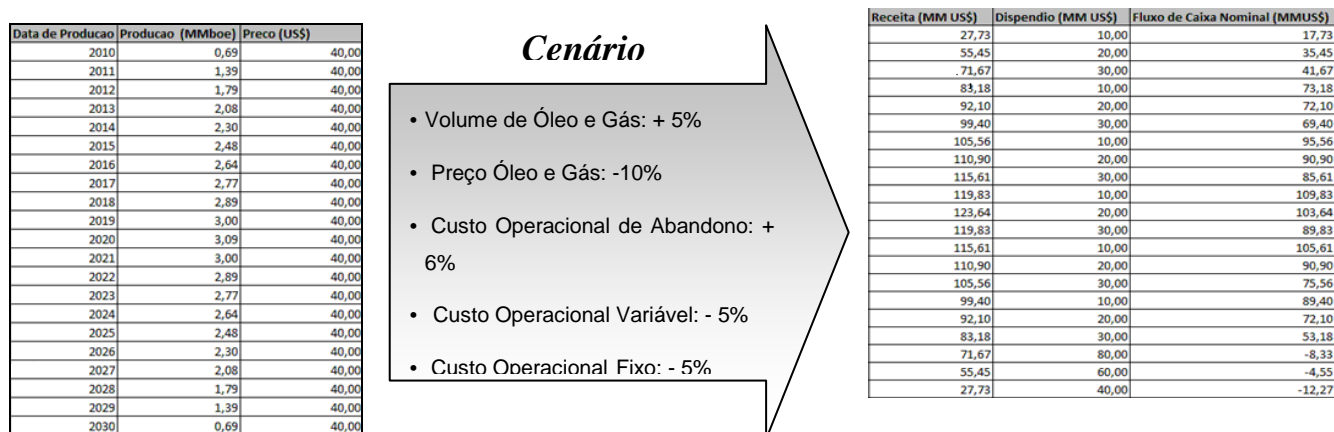


Figura 20: Base Histórica: Simulador

Dessa forma, a função do simulador era alterar percentualmente essas curvas de entrada e executar a aplicação PI, a fim de gerar novos indicadores econômicos para modelar o problema em questão.

Para essa configuração percentual sobre as curvas de entrada damos o nome de Cenário. Então, por exemplo, se aplicamos as variações percentuais abaixo, podemos dizer que essa configuração representa 1 Cenário:

- Volume de Óleo e Gás: + 5%
- Preço Óleo e Gás: -10%
- Custo Operacional de Abandono: + 6%
- Custo Operacional Variável: - 5%
- Custo Operacional Fixo: - 5%
- Investimento Total: + 10%

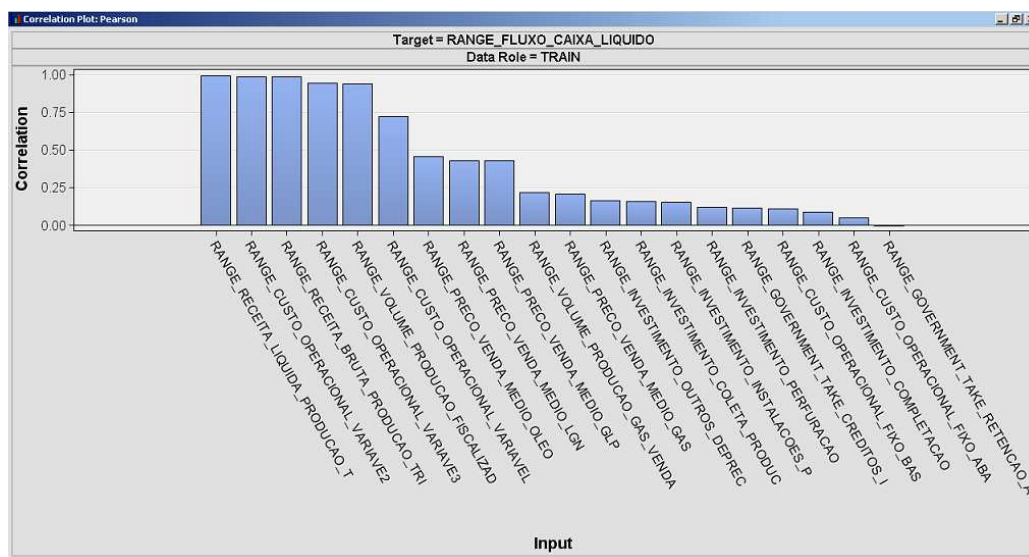
No final das simulações, foi obtido um total de 6.181 cenários diferentes que resultaram em uma base histórica com 259.619 observações com 130 indicadores cada.

### 4.3. PRÉ-PROCESSAMENTO

A etapa de pré-processamento foi realizada em 4 etapas:

- Análise descritiva das variáveis: onde foram identificados dados estatísticos como a quantidade de valores faltantes (*missing values*), média, mínimo e máximo. Analisando a base histórica, foi verificado que não existiam valores faltantes em nenhuma das variáveis.

- **Normalização:** durante a análise descritiva foi verificado que a escala das variáveis explicativas eram muito distintas entre elas. Dessa forma, para que todas as variáveis tenham a mesma influencia no modelo, independente da escala utilizada, foi realizada uma normalização linear entre 0 e 1 para equalizar o domínio das variáveis explicativas.
- **Análise de correlação:** para este estudo foi aplicado o coeficiente de correlação de Pearson, descrito na seção 3.2, para verificar o nível de importância da variável analisada individualmente em relação a variável dependente. Nesta etapa, foram selecionados os itens de maior correlação e significância<sup>9</sup> em relação a variável alvo. Além disso, foram removidas as variáveis cujo domínio se restringia ao valor zero (Figura 21 e Figura 22).



**Figura 21:** Gráfico Coeficiente de Correlação de Pearson

<sup>9</sup> Selecionadas as variáveis com a maior correlação com o Fluxo de Caixa Nominal (Rho) e maior significância (p-valor menor que 0.05, ou seja, confiabilidade de 95%).

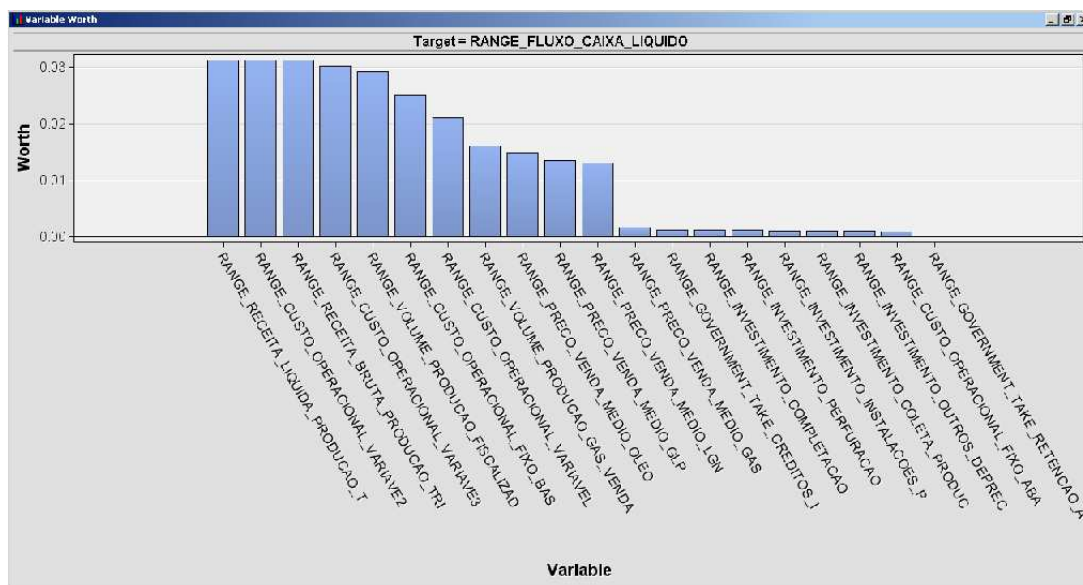


Figura 22: Importância da Variável

- Particionamento dos dados: os dados simulados foram divididos em 3 bases – treinamento (40%), validação (30%) e teste (30%). Essa estratégia foi realizada com o objetivo de evitar que o modelo seja criado especificamente para os dados de treinamento, perdendo assim uma certa generalização. Dessa forma, a estratégia é criar um modelo usando os dados de treinamento, mas escolher aquele que tem a menor erro sobre os dados de validação. E posteriormente, uma vez escolhido o modelo, são aplicados os dados de teste para que possamos verificar a efetiva performance do modelo.

#### 4.4. MODELAGEM

Uma vez realizada a etapa de pré-processamento descrita na seção 4.3, iniciamos a fase de escolher o modelo que melhor presente o cálculo do Fluxo de Caixa Nominal.

Para este problema, escolhemos aplicar o modelo de regressão linear múltipla, visando uma melhor interpretação dos resultados uma vez que podemos explicitamente verificar como cada variável afeta o modelo. Consequentemente, auxiliando o especialista na interpretação do modelo. Ao contrário de um modelo de rede neural que, apesar de ser um aproximador universal, existe uma complexidade na interpretação dos seus resultados, por ser o que muitos chamam de caixa preta.

Além disso, a fim de construir um modelo mais simplificado, escolheu-se testar 3 técnicas: *forward*, *backward* e *stepwise* para selecionar o melhor modelo seleção das

variáveis explicativas que irão compor o modelo. Em todos os casos, foram utilizados como critério de seleção do melhor modelo:

- erro de validação (*validation error*);
- nível de significância da variável para entrar no modelo: 0.05
- nível de significância da variável para permanecer no modelo: 0.05

Seguem os resultados (Figura 23):

Modelo	Nº parâmetros	R <sup>2</sup>	R <sup>2</sup> ajustado	MSE	RMSE
Original	20	0.9999	0.9999	0.00000313	0.001769140
Stepwise	15	0.9999	0.9999	0.00000313	0.001768853
Backward	15	0.9999	0.9999	0.00000313	0.001769131
Forward	17	0.9999	0.9999	0.00000313	0.001769104

**Figura 23:** Resultados dos Modelos de Regressão

Analisando os resultados obtidos entre os modelos construídos, foi escolhido o modelo de regressão *stepwise*, por possuir a menor quantidade de parâmetros em conjunto com os menores erros indicados (R<sup>2</sup>, R<sup>2</sup> ajustado, MSE<sup>10</sup> e RMSE<sup>11</sup>) para determinar o melhor modelo linear, segundo BEAL, 2011.

Foram construídos outros modelos com diferentes níveis de significância tanto para entrar como para permanecer no modelo. No entanto, todos tiveram erros superiores aos apresentados na tabela anterior.

Em relação à otimização do modelo, esta foi feita através das bases de treinamento e validação. Usado para a escolha do modelo menos complexo e com a melhor performance sobre os dados de validação.

A seguir, é apresentado o gráfico do modelo construído com o *stepwise*. Este gráfico (Figura 24) mostra o número de interações até encontrar o melhor modelo de acordo com o erro de validação. Foram 19 *steps* executados, sendo que o 17º *step* foi o escolhido.

<sup>10</sup> MSE: Mean Square Error.

<sup>11</sup> RMSE: Root Mean Squared Error.



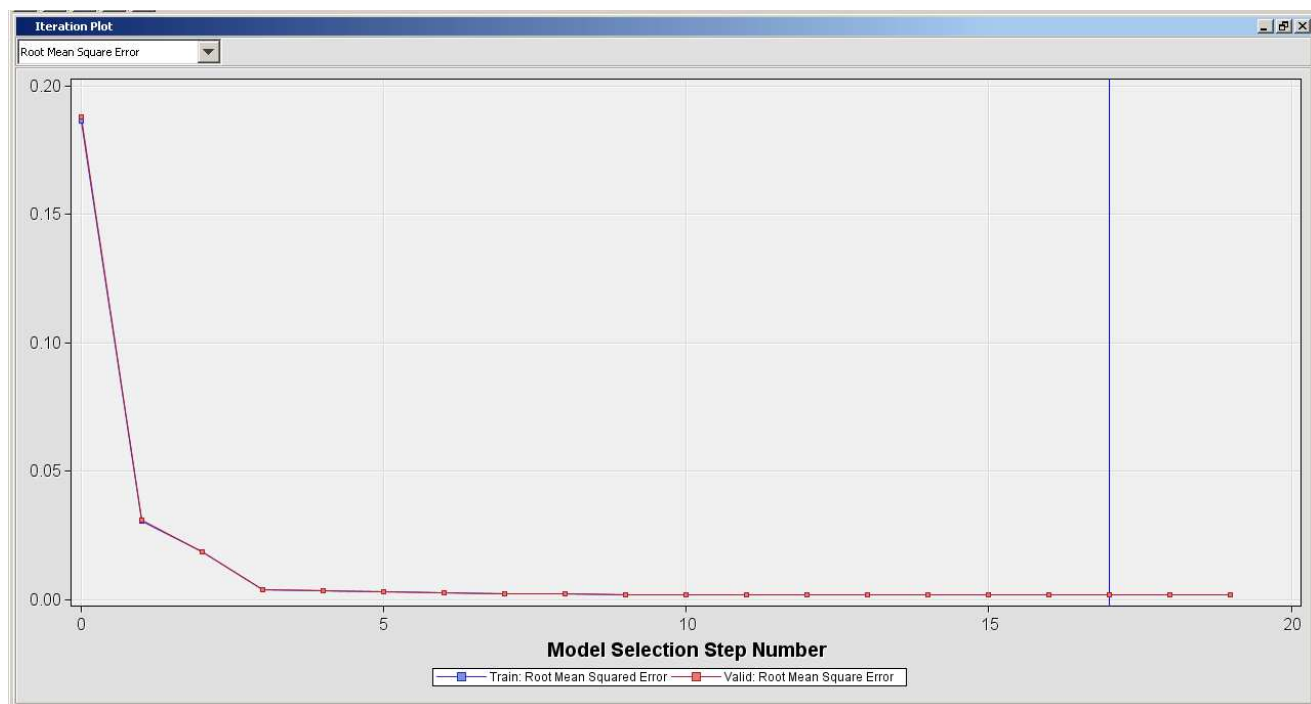


Figura 24: Gráfico de Interações do Modelo de Regressão *Stepwise*

Na Figura 25, está apresentado as estimativas, em termos absolutos, relacionadas com cada parâmetros que compõem a equação final do modelo.

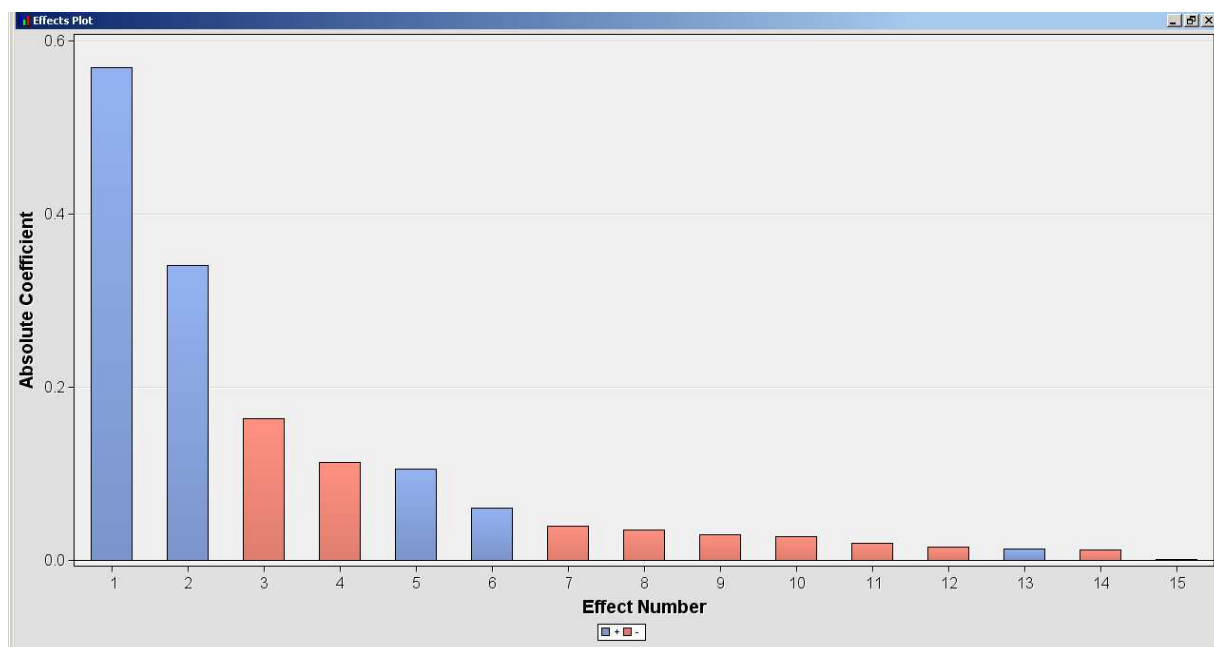


Figura 25: Estimativa dos Parâmetros do Modelo de Regressão *Stepwise*

## 4.5. RESULTADOS

Nas seções anteriores foi apresentado todo o processo de construção do modelo de regressão linear múltipla, desde a construção de uma base de dados histórica até a escolha do melhor modelo propriamente dito.

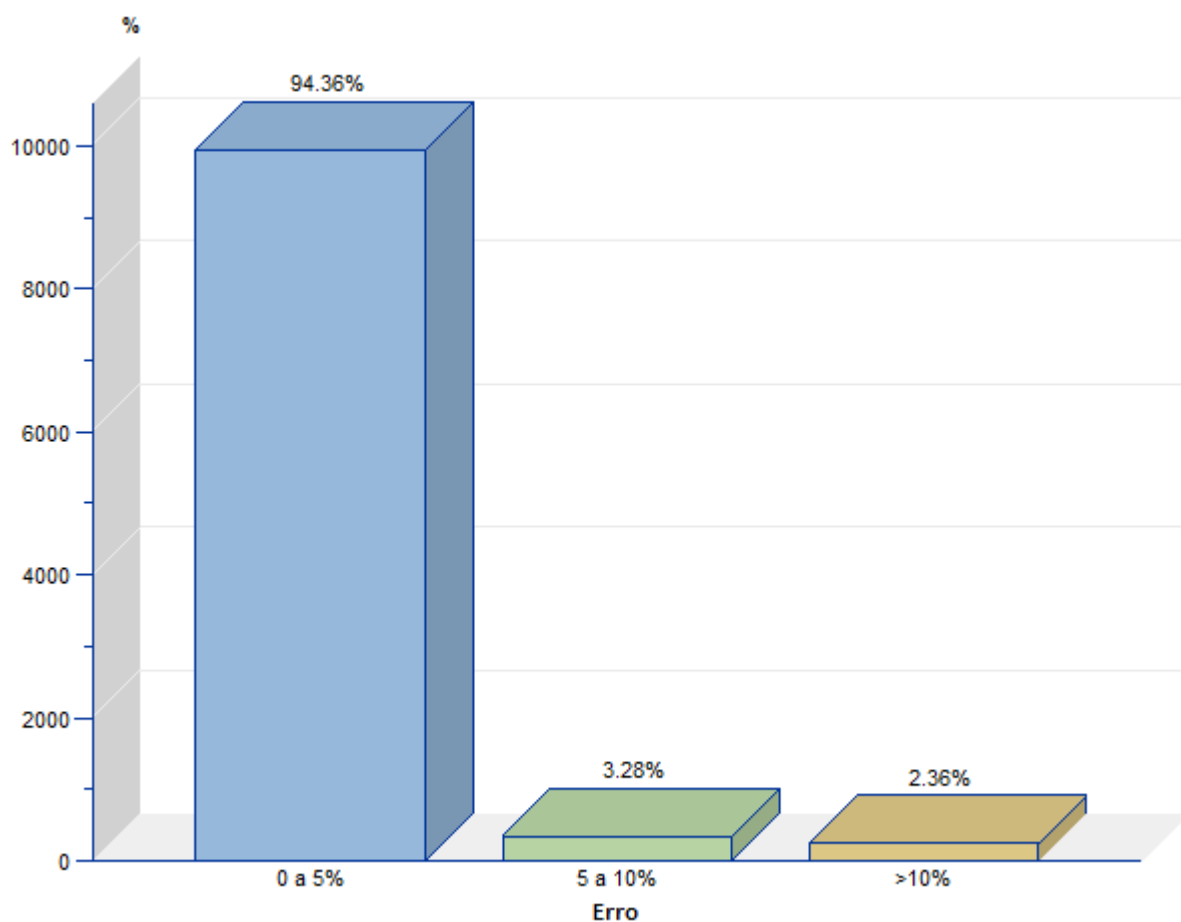
Na seção, serão apresentados os resultados encontrados a partir da utilização do modelo sobre uma base de escoragem (*score*) composta por 251 cenários diferentes, totalizando 10.543 observações.

É importante lembrar que essa base está dentro de um mesmo domínio que as bases de treinamento, validação e testes utilizadas na construção do modelo. Caso contrário, não faria sentido realizar uma previsão em um modelo totalmente diferente do domínio em questão.

Além disso, quando os novos dados foram aplicados sobre o modelo construído, eles também passaram pela mesma normalização ocorrida na criação do modelo. Da mesma forma que o valor estimado de saída passou por uma desnormalização para poder, posteriormente, ser comparado com o resultado real.

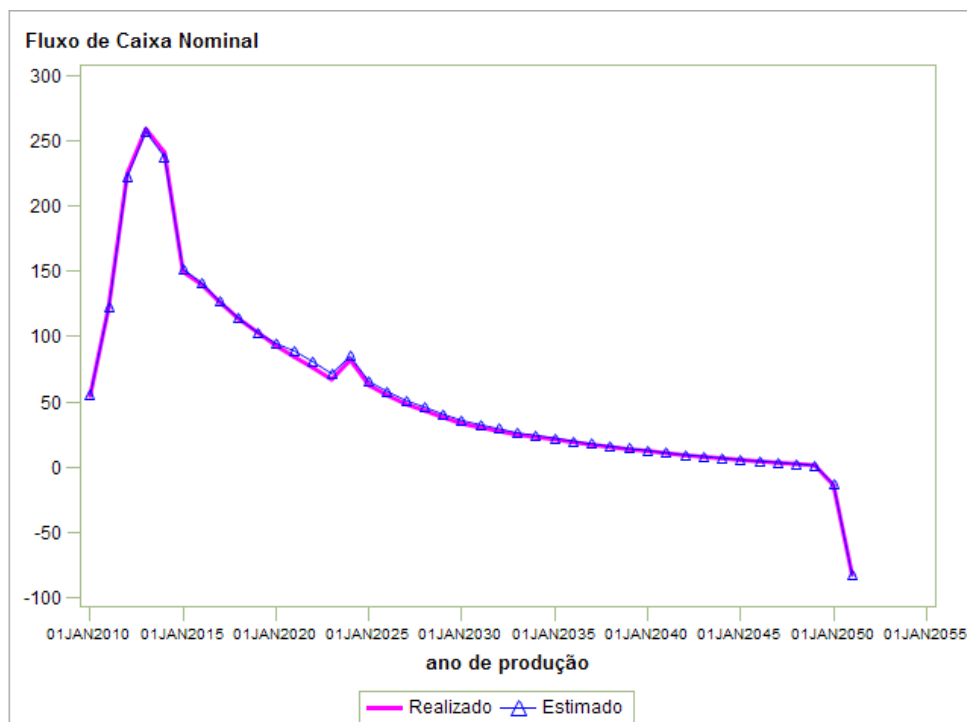
Assim, a partir dos resultados do modelo, foram calculadas as taxas de erro entre o Fluxo de Caixa Nominal Realizado menos o Estimado. Posteriormente, para facilitar a interpretação dos resultados, os erros foram divididos em 4 categorias como mostra a Figura 26:

- Igual a zero;
- Entre 0 e 5%;
- Entre 5 a 10%;
- Maior que 10%.



**Figura 26:** Taxa de Erro do Modelo

Nesta figura é possível observar que 94,36% das observações escoradas tiveram uma taxa de erro menor que 5% e apenas 2.36% tiveram erros superiores a 10%, em um tempo computacional significativamente pequeno. Além disso, na Figura 27 é apresentado um cenário comparando o Fluxo de Caixa Nominal Realizado versus o Estimado.



**Figura 27:** Fluxo de Caixa Nominal Realizado x Estimado

De acordo com a Figura 27, pode-se verificar que o modelo de regressão linear múltipla criado para solucionar o problema proposto consegue se aproximar dos resultados reais com uma taxa de erro significativamente pequena. Em outras palavras, ao verificar que existe um relacionamento linear entre as variáveis do problema, quantificando o poder que cada uma possui sobre a variável alvo e, finalmente, selecionar as que mais influenciam no resultado, pode-se gerar um aproximador de função para o problema de simulação da análise econômica.

## 5. CONCLUSÕES E TRABALHOS FUTUROS

Com o passar do tempo a concorrência entre empresas tem se tornado mais acirrada, exigindo métodos cada vez mais sofisticados na disputa pelos mercados. Saber investir corretamente os recursos disponíveis em projetos realmente rentáveis é a chave do sucesso das organizações.

Um instrumento importante na avaliação da viabilidade econômica de um projeto ou campo é o cálculo do fluxo de caixa nominal que é realizado durante o processo de análise econômica. Através deste processo são estimadas, anualmente, as reservas da empresa que, posteriormente, são divulgadas para os órgãos regulamentadores e o mercado em geral.

Entretanto, poder realizar esse cálculo de forma simplificada e menos custosa computacionalmente é uma importante oportunidade de conhecer melhor as suas receitas e dispêndios e quanto isso afetará o lucro final da empresa.

Desta forma, neste trabalho foi proposto um modelo de regressão linear múltipla para criar um aproximador de função do programa PI que é utilizado na empresa no processo do cálculo do fluxo de caixa nominal.

Foram apresentados conceitos desde a criação de uma base histórica, o seu pré-processamento e a fase de construção do modelo propriamente dito.

Na fase de construção do modelo, foram comparados 3 métodos: *stepwise*, *forward* e *backward*, onde o primeiro teve o melhor performance.

Como resultado, foi obtida uma equação, com apenas 15 variáveis das 130 originais, em que foi possível identificar a influência de cada uma no fluxo de caixa nominal estimado pelo modelo.

Ao final, foram apresentadas as taxas de erro estimadas a partir da base de escoragem utilizada para verificar o comportamento do modelo.

Este trabalho foi implementado utilizando um modelo de regressão, sendo este um ponto favorável em relação à facilidade de interpretar como cada variável afeta o modelo. Mas uma proposta para trabalhos futuros seria utilizar técnicas de redes neurais e programação genética para comparação.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

1. ANP. **Portaria ANP Nº 10, DE 13/01/99 (D. O. U. DE 14/01/99).** [http://www.anp.gov.br/brasil-rounds/round1/Docs/LDOC15\\_pt.pdf](http://www.anp.gov.br/brasil-rounds/round1/Docs/LDOC15_pt.pdf), 1999.
2. ANP. **Campos na fase de produção – etapa de produção.** ANP: Agência Nacional de Petróleo, Gás Natural e Biocombustíveis, 2010. <http://www.anp.gov.br/?dw=1084>
3. ANP. **Questões fiscais e tributárias.** <http://www.anp.gov.br/brasil-rounds/round2/Pdocs/Pleis/Pleis.htm#fiscais>. Acesso: 2011.
4. BEAL, D. J. **SAS Code to Select the Best Multiple Linear Regression Model for Multivariate Data Using Information Criteria.** Science Applications International Corporation, Oak Ridge, TN. Acesso: 2011.
5. BISHOP, C.M. **Neural networks for pattern recognition.** Oxford, Oxford University Press, 1995.
6. CAPPELLI, C. L.al. **Uma abordagem de construção de ontologia de domínio a partir do modelo de processos de negócio.** WOMSDE: II Workshop on Ontologies and Metamodeling in Software and Data Engineering, 2007
7. DEVORE, J.L. **Probabilidade e Estatística para Engenharia e Ciências.** Thomson. 6<sup>o</sup> Edição, 2006.
8. MARTINOVICH, M. **Como gerenciar o capital de giro. Agenda do Empresário.** São Paulo, nº 11, p. 1-6, 1996.
9. MENDES, A. et.al. **Análise de riscos em projetos de investimento.** Relatório Final do Grupo de Trabalho, 2004. Documentação interna da empresa OX.
10. PEREIRA, M. **Avaliação do impacto dos tributos na incorporação de reservas nas empresas do setor de petróleo.** Dissertação. COPPE/UFRJ. 2004.
11. RAVAGNANI, A. **Avaliação Econômica de Projetos de E&P.** Unicamp: UNISIM. Ano 3, Volume 2, 23<sup>o</sup> Edição, 2008.
12. SONDAK, N.E; SONDAK, V.K. **Neural networks and artificial intelligence.** ACM Press: New York, ACM SIGCSE Bulletin, Preceedings of Twentieth SIGCSE Technical Symposium on Computer Science Education, v.21, p.241 – 245, 1989.

13. RAVAGNANI, A.T. **Avaliação econômica de projetos de E&P**. UNISIM, Ano 3, Vol. 2, 2008.
14. SAS. **Curso SAS Enterprise Guide: ANOVA, Regression, and Logistic Regression Course Notes**, 2010.
15. SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Methods**. IOWA State University Press. Ed.: 8. 1989.
16. SOUZA, F. R. **Impacto do preço do petróleo na política energética mundial**. Dissertação. COPPE/UFRJ, 2006.
17. WASHINGTON; KARLAFTIS & MANNERING. **Statistical and econometric methods for transportation data analysis**, Chapman & Hall/CRC, 2003.