



MODELO PARAMÉTRICO PARA A ESTIMATIVA DO CUSTO DA CONSTRUÇÃO DE SONDAS DE PERFURAÇÃO NO MAR

Iago Emanuel Barbosa da Costa Veiga

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação de Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Produção.

Orientadores: Virgílio José Martins Ferreira Filho
Ricardo de Melo e Silva Accioly

Rio de Janeiro
Setembro de 2012

MODELO PARAMÉTRICO PARA A ESTIMATIVA DO CUSTO DA
CONSTRUÇÃO DE SONDAS DE PERFURAÇÃO NO MAR

Iago Emanuel Barbosa da Costa Veiga

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Examinada por:

Prof. Virgílio José Martins Ferreira Filho, D.Sc.

Prof. Ricardo de Melo e Silva Accioly, D.Sc.

Prof. Luiz Fernando Loureiro Legey, Ph.D.

Prof. Edilson Fernandes de Arruda, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO DE 2012

Veiga, Iago Emanuel Barbosa da Costa

Modelo Paramétrico Para a Estimativa do Custo da
Construção de Sondas de Perfuração no Mar/ Iago Emanuel
Barbosa da Costa Veiga. – Rio de Janeiro: UFRJ/COPPE, 2012.

XIV, 135 p.: il.; 29,7 cm.

Orientadores: Virgílio José Martins Ferreira Filho

Ricardo de Melo e Silva Accioly

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de
Engenharia de Produção, 2012.

Referências Bibliográficas: p. 107-110.

1. Sondas de perfuração no mar. 2. Estimativas de custos. 3.
Modelos paramétricos. I. Filho, Virgílio José Martins, et al., II.
Universidade Federal do Rio de Janeiro, COPPE, Programa de
Engenharia de Produção. III. Título.

Para meus pais, Ana Luiza e José Antônio,
pelo apoio e confiança.

AGRADECIMENTOS

Ao Professor Virgílio José por sua orientação, paciência, confiança e contribuição para o amadurecimento da minha formação acadêmica e, especialmente, na realização deste trabalho.

Ao Professor Ricardo Accioly pelo apoio técnico e teórico sem os quais não seria possível a realização desse trabalho.

A UFRJ/COPPE pelo auxílio concedido, sem os quais esse trabalho não poderia ter sido realizado.

Aos meus pais, Ana Luiza e José Antônio, por seu amor, dedicação e exemplo de vida que contribuíram, de forma inestimável, para a minha formação profissional e como ser humano.

A minha irmã Jordana pela ajuda e motivação.

A amiga Juliana por todo apoio, paciência e compreensão.

A todos os amigos e familiares que de alguma forma ou de outra me estimularam ou me ajudaram.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MODELO PARAMÉTRICO PARA A ESTIMATIVA DO CUSTO DA CONSTRUÇÃO DE SONDAS DE PERFURAÇÃO NO MAR

Iago Emanuel Barbosa da Costa Veiga

Setembro/2012

Orientadores: Virgílio José Martins Ferreira Filho
Ricardo de Melo e Silva Accioly.

Programa: Engenharia de Produção

Esta dissertação aborda a questão de como formular uma estimativa de custos para sondas marítimas utilizando modelos paramétricos, através de métodos estatísticos multivariados. Estimativas de custos paramétricas usam características relevantes do objeto de estudo para aferir assim a influência dessas na composição do preço e custo final. A base de dados utilizada como entrada do modelo foi uma amostra com as características técnicas de 100 sondas. Somente dois tipos de sondas foram escolhidos para o estudo: sondas semissubmersíveis e navios sonda. Os modelos multivariados escolhidos foram a regressão linear multivariada e a análise fatorial. Na primeira, as características de maior correlação com o custo e as que, pela teoria, são mais significativas foram usadas como variáveis independentes. Com relação à análise fatorial, esta buscou consolidar as informações contidas em todas as características fornecidas na construção de fatores, que foram utilizados como variáveis independentes em uma nova regressão, gerando assim um novo modelo. Os resultados foram apresentados assim como propostas de futuros estudos.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

PARAMETRIC COST ESTIMATION FOR THE CONSTRUCTION OF
OFFSHORE DRILLING RIGS

Iago Emanuel Barbosa da Costa Veiga

September/2012

Advisors: Virgílio José Martins Ferreira Filho

Ricardo de Melo e Silva Accioly.

Department: Industrial Engineering

This dissertation addresses the question of how to formulate a cost estimation for offshore rigs using parametric models, through multivariate statistic methods. Parametric cost estimates use relevant characteristics of the studied object to assess their influence in the composition of the final price and cost. The data base inputted for the model was a sample with the technical characteristics of 100 probes. Only two types of probes were chosen for the study: semi submersibles and probe ships. The chosen multivariate models were multivariate linear regression and factor analysis. In the first case, characteristics with higher correlation to the cost and the ones which, according to the theory, are more significant were used as independent variables. In regard to factor analysis, the aim was to consolidate the information contained in all the characteristics supplied in the buildup of factors, which were used as independent variables in a new regression, thereby generating a new model. The results were presented, as well as propositions for future studies.

SUMÁRIO

1 INTRODUÇÃO	1
1.1. JUSTIFICATIVA E MOTIVAÇÃO DO TEMA	2
1.2. OBJETIVOS	3
1.3. DELIMITAÇÃO DO ESTUDO	4
1.4. ESTRUTURA	4
2 REVISÃO BIBLIOGRÁFICA	5
2.1. ESTRUTURA DE CUSTOS	5
2.1.1. PROCESSO DE ESTIMATIVA DE CUSTOS	6
2.2. ANÁLISE MULTIVARIADA	14
2.2.1. CONCEITOS INICIAIS	14
2.2.2. TÉCNICAS MULTIVARIADAS	18
2.2.3. ANÁLISE FATORIAL	23
2.2.4. REGRESSÃO MÚLTIPLA	28
3 SONDAS MARÍTIMAS	37
3.1. O MERCADO MUNDIAL DE SONDAS MARÍTIMAS	38
3.2. AS ESPECIFICAÇÕES DE UMA SONDA DE PERFURAÇÃO	39
3.2.1. PRINCIPAIS COMPONENTES DE UMA SONDA DE PERFURAÇÃO MARINHA	45
4 METODOLOGIA	53
4.1. ESTUDOS PRELIMINARES	54
4.1.1. ANÁLISE DE DADOS COM APENAS UMA VARIÁVEL	55
4.1.2. ANÁLISE DE DADOS COM DUAS VARIÁVEIS	58
4.2. ANÁLISES MULTIVARIADAS PARA A ESTIMATIVA DE CUSTOS	59
5 RESULTADOS	61
5.1. ESTUDOS PRELIMINARES	61
5.1.1. ANÁLISE COM UMA VARIÁVEL	62

5.1.2. ANÁLISES COM DUAS VARIÁVEIS	67
5.2. ANÁLISES MULTIVARIADAS	84
5.2.1. REGRESSÃO LINEAR SIMPLES	84
5.2.2. REGRESSÃO MÚLTIPLA	86
5.2.3. ANÁLISE FATORIAL	92
5.2.4. VALIDAÇÃO DO MODELO E COMPARAÇÃO DOS RESULTADOS	102
6 CONCLUSÕES	105
7 REFERÊNCIA BIBLIOGRÁFICA	107
8 APÊNDICE	111
8.1. ANÁLISE COM VARIÁVEL	111
8.1.1. ANÁLISE DAS VARIÁVEIS INDEPENDENTES QUANTITATIVAS	111
8.1.2. ANÁLISE DAS VARIÁVEIS INDEPENDENTES QUALITATIVAS	119
8.2. ANÁLISE COM DUAS VARIÁVEIS	126
8.3. ANÁLISE DAS REGRESSÕES MÚLTIPLAS	127
GLOSSÁRIO	130

LISTA DE FIGURAS

Figura 1 - Etapas no processo de Estimativa de custos paramétrica	9
Figura 2 - Fluxo de atividades para o desenvolvimento para um modelo de estimativa de custos	12
Figura 3 - Esquema de escala de medidas	15
Figura 4 - Técnicas de análise multivariadas	18
Figura 5 - Análise gráfica dos resíduos	32
Figura 6 - Fluxograma do método de estimação <i>stepwise</i>	33
Figura 7 - Principais tipos de sondas <i>offshore</i> : fixas, <i>Jack-up</i> , semissubmersíveis e navio sonda.....	40
Figura 8 - Sonda do tipo <i>Jack-up (Cantilever)</i>	41
Figura 9 - NavioSonda.....	42
Figura 10 - Sonda semissubmersível	43
Figura 11 - Casco de um navio.....	45
Figura 12 - Esquema de uma sonda de perfuração.....	46
Figura 13 - Mastro de uma sonda marítima.....	47
Figura 14 - Sistema de movimentação de cargas	47
Figura 15 - <i>Top drive</i>	48
Figura 16 - Esquema do sistema de rotação e sistema de circulação	49
Figura 17 - Esquema gráfico de um BOP.....	50
Figura 18 – Fotografia de um BOP	50
Figura 19 - Funcionamento do posicionamento dinâmico	51
Figura 20 - Navio sonda de dupla atividade	52
Figura 21 - Histograma de Custo bruto	63
Figura 22 - Histograma de custo corrigido pela inflação	63
Figura 23 - Boxplot do custo corrigido	64
Figura 24 - Histograma do Custo corrigido e sem dados anômalos.....	66
Figura 25 - Boxplot do Custo corrigido sem dados anômalos	66
Figura 26 - Gráfico de dispersão entre os custos corrigidos e a profundidade da lâmina d'água	70

Figura 27 - Gráfico de dispersão entre os custos corrigidos e a profundidade perfuração	71
Figura 28 - Gráfico de dispersão entre os custos corrigidos e a carga máxima da sonda	71
Figura 29 - Boxplot da carga máxima da sonda pelo tipo de sonda.....	72
Figura 30 - Gráfico de dispersão entre os custos corrigidos, a carga máxima e o tipo de sonda.....	73
Figura 31 - Gráfico de dispersão entre os custos corrigidos e capacidade de carga variável da sonda	74
Figura 32 - Boxplot da capacidade de carga variável da sonda pelo tipo de sonda ...	75
Figura 33 - Gráfico de dispersão entre os custos corrigidos e capacidade de carga variável da sonda	75
Figura 34 - Gráfico de dispersão entre os custos corrigidos e capacidade do alojamento	76
Figura 35 - Gráfico de dispersão entre os custos corrigidos e a carga máxima suportada pela torre	77
Figura 36 - Relação entre o tipo de sonda e oriente	82
Figura 37 - Relação entre o tipo de sonda e a capacidade de EWT	82
Figura 38 - Gráfico de probabilidade normal: resíduos padronizados da regressão simples.....	85
Figura 39 - Gráfico de dispersão dos resíduos padronizados da regressão simples...	86
Figura 40 - Gráfico de probabilidade normal: resíduos padronizados da regressão múltipla.....	90
Figura 41 - Gráfico de dispersão dos resíduos padronizados para a regressão múltipla	91
Figura 42 - Gráfico de probabilidade normal: resíduos padronizados da regressão múltipla dos fatores	100
Figura 43 - Histograma dos resíduos da regressão após análise fatorial	101
Figura 44 - Gráfico de dispersão dos resíduos padronizados para a regressão múltipla dos fatores.....	102
Figura 45 - histograma da profundidade da lâmina d'água	113
Figura 46 - Boxplot da profundidade da lâmina d'água	114
Figura 47 - Histograma carga máxima suportada pela sonda.....	115
Figura 48 - Boxplot carga máxima suportada pela sonda	115

Figura 49 - Histograma da capacidade de carga variável da sonda.....	116
Figura 50 - Boxplot da capacidade de carga variável da sonda.....	117
Figura 51 - Histograma da profundidade de perfuração.....	118
Figura 52 - Boxplot da profundidade de perfuração.....	118
Figura 53 - Proporção entre os tipos de sondas	119
Figura 54 - Estado da construção das sondas	120
Figura 55 - Proporção entre sondas feitas de forma especulativa ou não.....	120
Figura 56 - Distribuição das sondas pelo local de construção.....	121
Figura 57 - Distribuição das sondas por possuir posicionamento dinâmico ou não.	122
Figura 58 - Sondas adaptados a um ambiente hostil	123
Figura 59 - Sondas com a presença de Top Drive	124
Figura 60 - Sondas com dupla atividade	125
Figura 61 - Gráfico de dispersão variáveis independentes quantitativas.....	126

LISTA DE TABELAS

Tabela 1 - Matriz de classificação de estimativas	8
Tabela 2 - Pontos positivos e negativos do método de estimativa de custos paramétricas	10
Tabela 3 - Pontos positivos e negativos do método de estimativa de custos por analogia	11
Tabela 4 - Pontos positivos e negativos do método de estimativa de custos <i>bottom-up</i>	11
Tabela 5 - Tipos de erros	17
Tabela 6 - Relação entre métodos de dependência.....	19
Tabela 7 - Aumento dos custos dos insumos e serviços de E&P entre 2000 e 2007 .	37
Tabela 8 - Relação de companhias e quantidade de sondas marítimas - 2008.....	38
Tabela 9 - Índices de inflação usados na deflação dos custos	56
Tabela 10 - Lista de variáveis independentes selecionadas.....	61
Tabela 11 - Estatísticas descritivas dos custos bruto e corrigido	62
Tabela 12 - Estatísticas descritivas dos custos corrigidos com e sem valores anômalos	65
Tabela 13 - Teste de normalidade Kolmogorov-Smirnov para a variável dependente	67
Tabela 14 - Lista de correlação entre Custo e variáveis quantitativas	68
Tabela 15 - Correlação entre as variáveis independentes quantitativas (I).....	68
Tabela 16 - Correlação entre as variáveis independentes quantitativas (II).....	69
Tabela 17 - Resumo da significância e da estatística F para a <i>one way</i> ANOVA.....	79
Tabela 18 - Análise da correlação entre a variável dependente e as variáveis independentes qualitativas.....	80
Tabela 19 - Correlação entre as variáveis independentes qualitativas	81
Tabela 20 - Sumário da regressão do custo corrigido pela variação de carga suportável	84
Tabela 21 - Análise da variância da regressão do custo corrigido pela variação de carga suportável	84
Tabela 22 - Coeficientes da regressão do custo corrigido pela variação de carga	

suportável	84
Tabela 23 - Variáveis entrantes e removidas pelo método <i>stepwise</i>	87
Tabela 24 - Sumário da regressão múltipla do custo corrigido pelo método <i>stepwise</i>	88
Tabela 25 - Análise da variância da regressão múltipla do custo corrigido pelo método <i>stepwise</i>	88
Tabela 26 - Coeficientes da regressão da regressão múltipla do custo corrigido pelo método <i>stepwise</i>	88
Tabela 27 - Estatísticas de colinearidade da regressão múltipla	89
Tabela 28 - Teste de KMO e de Bartlett's	92
Tabela 29 - Comunalidades	93
Tabela 30 - Autovalores iniciais.....	94
Tabela 31 - Total da variância explicada.....	95
Tabela 32 - Matriz componente.....	96
Tabela 33 - Matriz componente rotacionada	97
Tabela 34 - Matriz de correlação entre o custo corrigido e componentes fatoriais....	98
Tabela 35 - Sumário da regressão do custo corrigido pelos fatores	98
Tabela 36 - Análise da variância da regressão do custo corrigido pelos fatores	99
Tabela 37 - Coeficientes da regressão do custo corrigido pelos fatores.....	99
Tabela 38 - Estatísticas de colinearidade da regressão múltipla dos fatores	100
Tabela 39 - Valores de erro médio e erro máximo para as regressões múltiplas	103
Tabela 40 - Validação da regressão múltipla.....	103
Tabela 41 - Lista de variáveis independentes selecionadas.....	111
Tabela 42 - Estatísticas descritivas variáveis independentes quantitativas	112
Tabela 43 - Teste de normalidade Kolmogorov-Smirnov para variáveis independentes quantitativas.....	112
Tabela 44 - Resultado da previsão das regressões múltiplas.....	127

1 INTRODUÇÃO

Nas últimas décadas, o Brasil apresentou um grande crescimento na descoberta de novas reservas de petróleo, sendo sua grande maioria localizada no Oceano Atlântico, principalmente nas áreas conhecidas como bacia de Campos e de Santos. Atualmente, houve o interesse pela exploração do pré-sal brasileiro e a diminuição das reservas em terra mundiais.

No cenário mundial, com o aumento constante da demanda por derivados de petróleo gerando a elevação do patamar de preços, novas áreas de produção se tornaram economicamente viáveis. A exploração e a produção sob águas marinhas são mais caras que as em solo firme. Com esse aumento de preços essas áreas se tornaram atrativas. Como os equipamentos utilizados nesse tipo de projeto se encontram em número limitado e é necessário tempo de espera até que a oferta desses se materialize, existe um excesso de demanda elevando o seu custo de aluguel, assim como a espera por sua disponibilidade. Existe um aquecimento do mercado brasileiro e mundial para construção de sondas de perfuração no mar.

Nos Estudos de Viabilidade Técnica e Econômica (EVTE) dos projetos de investimento, é fundamental a estimação dos custos envolvidos nos projetos e sua análise. Quanto mais precisas forem as informações que irão compor os custos e contingências dos projetos, melhor a estimativa de sua viabilidade econômica, pois essa será realizada com base nestes custos. Um modelo de estimativa de custos de sondas, que dê suporte a esta análise é essencial para EVTE's dos projetos da área de Exploração e Produção de empresas petrolíferas. Hoje ainda existe uma grande variância entre as estimativas iniciais e o custo final do projeto, além do tempo que é decorrido entre a contratação para a construção de uma sonda até sua entrega, que pode chegar a muitos anos, dependendo do projeto.

O mercado petrolífero apresenta altos riscos e custos, os maiores na área de exploração e produção, aliados a uma incerteza quanto ao índice de sucesso na descoberta de petróleo por perfuração realizada, esse variando entre cerca de 10% a 40%. Segundo LIMA (2003) “a taxa média mundial de sucesso na perfuração de poços exploratórios é de aproximadamente 1 para 8, ou seja, para cada 8 poços perfurados

somente em um se encontram reservas de hidrocarbonetos, podendo estas ainda ser tanto comerciais como sub comerciais”.

LIMA (2003) também destaca que “a tecnologia desenvolvida pela Petrobrás vem permitindo alcançar taxas de sucesso da ordem de 40% para os poços pioneiros. Porém, nem mesmo as novas técnicas exploratórias eliminam, nem eliminarão os riscos de o poço estar seco ou ser subcomercial”. O momento correto para avançar nas campanhas exploratórias também deve ser considerado e estar sempre condicionado ao preço do petróleo, levando também em conta que essas são intensivas em investimentos e demandam certo tempo. Em média, no Brasil, são necessários cinco anos entre a aquisição da concessão da área até a declaração de comercialidade do campo, em caso de sucesso na fase exploratória.

1.1. JUSTIFICATIVA E MOTIVAÇÃO DO TEMA

Desde o ano de 2004, em função de sucessivas altas no preço do petróleo, várias empresas petrolíferas passaram por uma grande corrida exploratória em diversas partes do mundo. Em função disso, a demanda por equipamentos e materiais utilizados na fase de exploração aumentou vertiginosamente, em destaque observam-se as sondas marítimas como recurso crítico, pois:

- Existe uma demora relativamente grande entre a aquisição e o recebimento do produto, se comparada a outros equipamentos;
- Houve um aquecimento do mercado brasileiro e mundial de sondas de perfuração marítima, limitando assim a disponibilidade de sondas;
- Existe um processo de construção de sondas marítimas no Brasil e um modelo de previsão poderia ser outra forma de avaliação dos custos envolvidos; e,
- A sonda realiza as principais atividades de exploração petróleo no mar, além de possuir um grande conjunto de equipamentos.

A indústria de exploração petrolífera envolve altos riscos, incertezas e custos elevados, necessitando de investimentos iniciais de grande escala, pelo porte dos equipamentos e pela estrutura necessária para a execução da atividade, além de atuar muitas vezes em ambientes agressivos e que envolvem a possibilidade de contaminação ambiental. Isso, somado à baixa eficiência na perfuração de poços, gera a necessidade

de se controlar bem os investimentos, assim como a necessidade de capitalização da empresa, aumentando, dessa forma, a importância de uma boa estimativa dos custos.

Para exemplificar a importância da exploração, a indústria do petróleo, no Brasil, tem seus maiores reservatórios no mar e perfura poços a grandes profundidades, utilizando tecnologias caras e realizando investimentos da ordem de dezenas de milhões de dólares em cada uma dessas perfurações. No período entre 1998 e 2007, a Petrobras desenvolveu atividades exploratórias que compreenderam mais de 790 poços, desses, 459 poços foram perfurados no mar, onde a complexidade e os custos são maiores.

A descoberta de petróleo na camada pré-sal gerou um aumento na expectativa de perfuração de poços no mar para os próximos anos.

No meio acadêmico existem muitos livros que tratam de modelos paramétricos para estimativas de custos; outros muitos que tratam de análise multivariada de dados e da atividade de perfuração de poços e seus equipamentos. No entanto, poucos artigos acadêmicos tratam dos três temas de forma conjunta. A ocorrência do tema estimativa de custos de poços utilizando modelos paramétricos é bem mais comum, como visto em LEAMON (2006), mas trata-se de uma abordagem totalmente diferente da proposta nesse estudo, cujo objeto de estudo é a sonda e não a atividade de perfuração em si.

1.2.OBJETIVOS

O objetivo principal deste estudo é desenvolver um modelo paramétrico para estimativa de custos de construção de sondas de exploração no mar, utilizando modelos de análise multivariada de dados.

Além disso, o estudo busca encontrar as características mais relevantes para a estimativa de custo de uma sonda e o grau de importância de cada uma delas. Para se atingir esse objetivo, outros objetivos intermediários precisam também ser alcançados.

- Determinar as variáveis que fazem parte do rol de variáveis independentes;
- Fazer o tratamento adequado nas variáveis estudadas, incluindo a normalização dos custos para uma mesma data;
- Realizar a construção de fatores que condensem as informações de diversas variáveis;
- Comparar a regressão múltipla feita com variáveis e as feitas com os fatores criados e analisar os resultados.

1.3. DELIMITAÇÃO DO ESTUDO

A pesquisa aborda uma amostra de 100 sondas que foram contratadas entre fevereiro de 2001 e novembro de 2010, em todo o mundo e a fonte de dados foi o IHS *Petrodata Market Survey System*, sendo esses coletados em dezembro de 2010.

Cada uma das sondas apresenta cerca de 60 propriedades estudadas. Será feita uma análise estatística dos dados e o desenvolvimento de modelos paramétricos assim como a avaliação dos mesmos, usando como variáveis explicativas as principais características das sondas. No glossário é possível ver a descrição de todas elas.

1.4. ESTRUTURA

Esse estudo apresenta cinco capítulos após a introdução. O capítulo 2 apresenta os conceitos básicos sobre os temas abordados no trabalho. Serão feitos breves comentários sobre engenharia de custos e sobre análise multivariada de dados.

O capítulo 3 trata das características de sondas de exploração marítimas, além de uma contextualização sobre o seu mercado internacional. Nesses dois capítulos serão apresentadas referências bibliográficas com o objetivo de explicitar a relação do tema com outros trabalhos relacionados.

Já no capítulo 4, será apresentado o tema da pesquisa, as premissas básicas adotadas, seus objetivos específicos, a metodologia aplicada para a aferição do modelo, assim como os critérios de seleção escolhidos.

Em seguida, no capítulo 5, serão apresentados os resultados da experimentação e a sua análise, assim como a seleção do melhor modelo.

Por último, serão apresentadas as conclusões e as recomendações de temas para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

Os principais conceitos teóricos utilizados neste trabalho serão detalhados neste capítulo. Inicialmente, serão apresentados alguns tópicos sobre engenharia de custos e em seguida uma breve descrição de técnicas de análise multivariada de dados, com ênfase nas que foram aplicadas nesse estudo.

2.1. ESTRUTURA DE CUSTOS

Segundo o Instituto Brasileiro de Engenharia de Custos, a Engenharia de Custos “é o ramo da engenharia que estuda os métodos de projeção, apropriação e controle dos recursos monetários necessários à realização dos serviços que constituem uma obra ou projeto, de acordo com um plano de execução previamente estabelecido”.

A análise de custos talvez seja um dos tópicos mais relevantes para as empresas, pois sua redução é essencial para sua sobrevivência no mercado. Para competir num mercado cada vez mais competitivo é totalmente necessário atuar com qualidade e flexibilidade. Assim é de suma importância que os custos de um novo projeto sejam conhecidos antes mesmo que eles ocorram.

A engenharia de custos subsidia as empresas na tomada de decisões, com o gerenciamento de custos e com o orçamento referente ao desenvolvimento de um produto. É uma metodologia usada para prever e estimar o custo de uma atividade ou produto (STEWART *et al.*, 1995). A estimativa é essencial na fase de desenvolvimento do projeto, pois pode influenciar na tomada de decisão de produzi-lo ou não.

Muitos autores acreditam que entre 70% e 80% dos custos são definidos ainda na fase conceitual, entre os quais STEWART *et al.* (1995), NASA (2008), TAYLOR (1997) e MILEHAM *et al.* (1993). Uma escolha errada nesse estágio pode gerar um custo muito maior caso haja a necessidade de mudanças em momentos posteriores no processo de produção. As dificuldades em fazer as estimativas ainda na fase conceitual é destacado em PUGH (1992), BUXTON *et al.* (1994), MEISL (1988) e em RUSH and ROY (2000).

O conhecimento referente a engenharia de custos é apresentado em diversas linhas de pensamento, como a linha seguida pela organização AACE International (*Association for the Advancement of Cost Engineering International*) ou a linha seguida pela NASA (National Aeronautics and Space Administration). Esse trabalho tem como principal fonte o modelo desenvolvido pela NASA, com algumas complementações de outras linhas de pensamento.

2.1.1. PROCESSO DE ESTIMATIVA DE CUSTOS

O processo de estimativa de custos pode ser dividido em três fases:

- A definição do projeto;
- A metodologia estimativa de custo; e,
- A execução da estimativa em si.

2.1.1.1. A definição do projeto

A definição do projeto é dividida em três etapas:

- Receber a proposta e realizar o entendimento do projeto;
- Construir ou obter a estrutura analítica do projeto (EAP, também conhecido como WBS, *work breakdown structure*) que é a decomposição do projeto em estruturas menores; e,
- Obter ou participar do desenvolvimento de uma descrição técnica do mesmo.

O recebimento da proposta possui duas atividades principais interligadas: coletar e revisar todos os dados relevantes para a avaliação discutindo o cronograma, os dados, expectativas e necessidades de recursos com o demandante e avaliar as necessidades, objetivos e metas, avaliando o ambiente operacional e o ciclo de vida do projeto,

A construção e ou obtenção do EAP possui três atividades principais associadas: determinar se ele existe ou trabalhar para sua criação, criar um dicionário e definir os elementos do EAP e garantir que a estimativa de custos de EAP é consistente em diferentes tópicos, como por exemplo, no orçamento, no planejamento do projeto e nos contratos.

A obtenção ou participação do desenvolvimento de uma descrição técnica do projeto envolve duas atividades principais: descrever até um segundo nível ou mais características do projeto como configurações, fatores de qualidade e riscos, bem como descrever os marcos, o cronograma e a estratégia de gerenciamento do projeto.

2.1.1.2. A metodologia de custo

A Metodologia de custo é dividida em quatro tarefas:

- Desenvolver regras básicas e premissas;
- Escolher a metodologia de custo a ser utilizada;
- Escolher e construir o modelo;
- Coletar e fazer o tratamento de dados.

O desenvolvimento de regras básicas e premissas são feitos por três atividades:

- Estabelecer uma lista levando em consideração questões técnicas, programáticas e de encadeamento;
- Atingir um consenso entre as partes interessadas;
- Documentar de forma completa, já que essas serão utilizadas em todo o processo.

Existem diferentes referências sobre os tipos de metodologias de estimativa de custos. Segundo NASA (2008) existem 3 tipos de percursos metodológicos:

- Paramétricos;
- Análogos;
- Engenharia de construção (*bottom-up*).

É possível destacar também a existência da metodologia do “chute” baseada apenas no julgamento de profissionais mais experientes, a qual não apresenta um método e alta imprecisão, sendo assim não será tratada de forma mais detalhada.

Segundo HOLLMANN (1997) há 4 questões que são decisivas para a escolha da metodologia, o uso final da estimativa, o grau de detalhamentos das informações sobre o projeto, a precisão desejada para a previsão e a tipo de técnica a ser utilizada sendo estocásticas ou determinísticas.

Já AACE (1998) define que é possível utilizar uma matriz que separa em 5 classes de estimativas tendo como base 5 características, o grau de definição do projeto, o seu uso final, a metodologia, a precisão esperada e esforço para o preparo da mesma.

O grau de definição é quem delimita em que classe a previsão estará disposta, sendo as outras características relacionadas à primeira, sendo assim a matriz é utilizada para consulta.

Tabela 1 - Matriz de classificação de estimativas

Classe	Nível de definição do projeto	Uso final	Metodologia	Faixa de precisão esperada	Esforço para preparação
5	0% a 2%	Conceitual	Modelos paramétricos ou análogos	-50% a +100%	1
4	1% a 15%	Estudo de viabilidade	Modelos paramétrico, análogos ou engenharia de construção	-30% a +50%	2 a 4
3	10% a 40%	Orçamento, autorização ou controle	Modelos paramétrico, análogos ou engenharia de construção	-20% a +30%	3 a 10
2	30% a 70%	Controle e licitação	Modelos paramétrico, análogos ou engenharia de construção	-15% a +20%	4 a 20
1	50% a 100%	Estimativa de checagem ou para licitação	Engenharia de construção	-10% a +15%	5 a 100

Fonte: AACE (1998)

A metodologia de estimativas de custo paramétrica é baseada em dados históricos e expressões matemáticas relacionando o custo como variável dependente e os direcionadores de custos como variáveis dependentes em uma análise de regressão. Os modelos paramétricos podem ser classificados em simples ou complexos, segundo o ISPA (2008) modelos simples utilizam apenas uma variável como direcionadora de custos enquanto modelos complexos apresentam múltiplas relações de estimativas de custos ou se utilizam de algoritmos para se encontrar essa estimativa.

Em geral essa modelagem é escolhida quando poucas características chaves de dados são conhecidas e a premissa implícita é que as mesmas forças que afetaram o custo no passado irão afetar os custos no futuro.

A Figura 1 apresenta as sete etapas do processo de estimativa de custos paramétrica de acordo com NASA (2008) que são:

- Definir 'hipóteses' de estimativas;
- Coletar dados de 'relações';
- Avaliar e tratar os dados;
- Analisar os dados em busca de relações candidatas;
- Realizar a análise estatística;
- Testar as relações; e,
- Selecionar a relação de estimativa de custos.

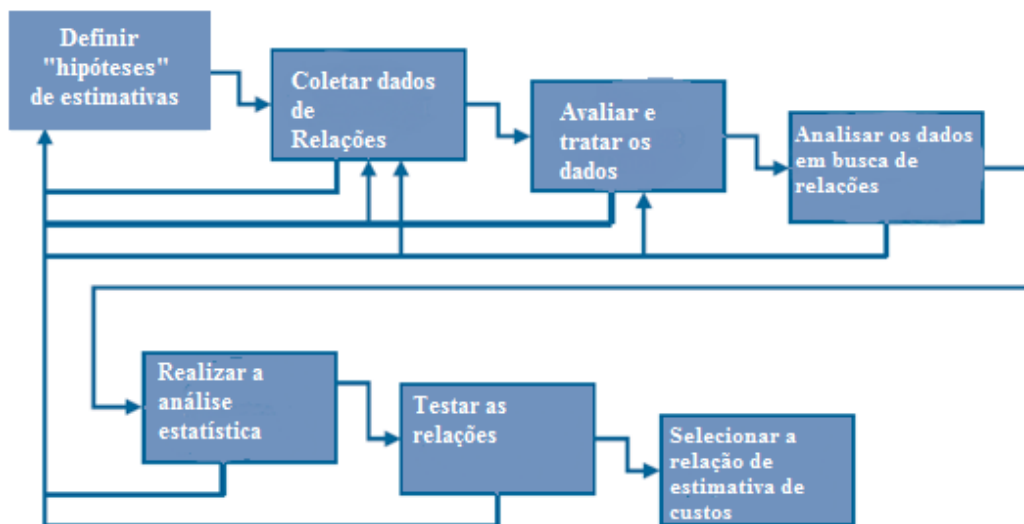


Figura 1 - Etapas no processo de Estimativa de custos paramétrica
Fonte: NASA (2008)

A apresentação da Figura 1 mostra-se interessante no sentido de evidenciar a idéia de rotina no processo, com exceção da última etapa, a qual envolve a escolha da relação, todas as outras podem levar novamente a primeira etapa e a pelo menos alguma outra das etapas anteriores caso os seus objetivos não tenham sido atendidos.

Nos modelos paramétricos o estimador pode criar suas próprias relações de estimativas de custos (*Cost estimate relationship*, CERs). Para se realizar uma análise de regressão o primeiro passo é determinar a relação entre as variáveis dependentes e independentes. Então os dados são ajustados usando técnicas como regressão linear e

regressões não lineares. Informações mais detalhadas sobre essas análises podem ser encontradas na próxima seção. Para um CER, a variável dependente será sempre o custo e as independentes serão os direcionadores de custos, que devem ser escolhidos por existir uma alta correlação entre ele e o custo e porque há razões conceituais para que essa relação seja investigada.

A análise de regressão não pode ser utilizada para confirmar causalidade, mas serve para testar a força e direção das relações quantitativas envolvidas. Ela é usada para estimar os testes de hipóteses considerando os parâmetros do modelo. As CERs estabelecidas anteriormente devem ser periodicamente examinadas para garantir que essas ainda sejam válidas e se a mostra de dados que está sendo estimada é aplicável ao modelo.

A Tabela 2 apresenta os pontos negativos e positivos da metodologia de custos paramétricos.

Tabela 2 - Pontos positivos e negativos do método de estimativa de custos paramétricas

Pontos positivos	Pontos negativos
Uma vez desenvolvidas as relações, são ótimas ferramentas para analisar a sensibilidade em diferentes situações	Em vários momentos é difícil outros entenderem as relações
Estatisticamente parecem preditores que fornecem informações sobre a confiança do estimador e sua habilidade preditiva	Necessidade de documentação de dados sem tratamento, ajuste de dados, desenvolvimento de equações, obtenção de resultados estatísticos e conclusões para validação e aceitação.
Elimina a dependência sobre o uso de observações atuais	A coleta apropriada de dados e a geração correta de relações estimativas de custos (CERs) são complicadas, consomem tempo e são caras.
A defesa da metodologia reside em uma correlação lógica, através de pesquisas disciplinadas, dados defensáveis e método científico.	Perde poder e credibilidade de previsibilidade fora do espaço de dados relevantes.

Fonte: NASA (2008)

A metodologia de estimativa de custos por analogia é feito com a comparação e extrapolação de dados de projetos passados que são tecnicamente representativos servindo de base para essa estimativa. Os dados são subjetivamente ajustados de acordo com a impressão de complexidade do novo produto em relação ao anterior.

Na Tabela 3 são apresentados os principais pontos positivos e negativos desse método.

Tabela 3 - Pontos positivos e negativos do método de estimativa de custos por analogia

Pontos positivos	Pontos negativos
Baseado em dados históricos atuais	Só considera um único dado para fazer a estimativa
Rapidez	Dificuldade de identificar a analogia apropriada
Facilidade de interpretação	Necessita normalização para garantir a precisão
Preciso para pequenas variações em relação ao projeto análogo	Baseia-se na extrapolação e ou no julgamento de especialista para o fator de ajuste

Fonte: NASA (2008)

A metodologia de engenharia de construção também conhecida como metodologia “*bottom-up*” de baixo para cima utiliza a estimativa de cada um dos elementos em seu nível mais detalhado para compor a estimativa de todo o projeto.

A Tabela 4 apresenta os principais pontos positivos e negativos na utilização dessa metodologia.

Tabela 4 - Pontos positivos e negativos do método de estimativa de custos *bottom-up*

Pontos positivos	Pontos negativos
Intuitiva	Custosa, exige um esforço significativo de tempo e mão de obra
Defensável	Não possui clara interpretação da sensibilidade
Credibilidade proveniente pela visibilidade da estimativa de cada elemento	Novas estimativas devem ser feitas para cada nova situação
Robusto, a estimativa geral não é afetada pelo erro de uma estimativa mal calculada individual	Não fornece um nível estatístico de confiança
Fornecer boas indicações de possíveis contribuidores do custo geral	Não fornece boas indicações de direcionadores de custos
Reutilizável, pode ser facilmente convertida em outras análises assim como em análises individuais	As relações entre os elementos precisam ser estruturadas pelo analista

Fonte: NASA (2008)

O ISPA (2008) estima que a economia de utilizar um modelo paramétrico é da ordem de 40% a 80% quando comparado com uma metodologia de engenharia de construção. São ferramentas mais simples de melhor usabilidade em fases conceituais do projeto, não é necessário saber o preço dos equipamentos e insumos e os dados são muito bem conhecidos após a calibração e validação dos resultados.

O modelo paramétrico será analisado de forma mais detalhada por se tratar do foco do trabalho. Para o ISPA (2008) a construção de um modelo paramétrico se dá em 7 passos, ou seja, desenvolvimento da base de dados, requisitos, resolução da

arquitetura e a disponibilidade de dados, desenvolvimento , calibração, validação e sua atualização.

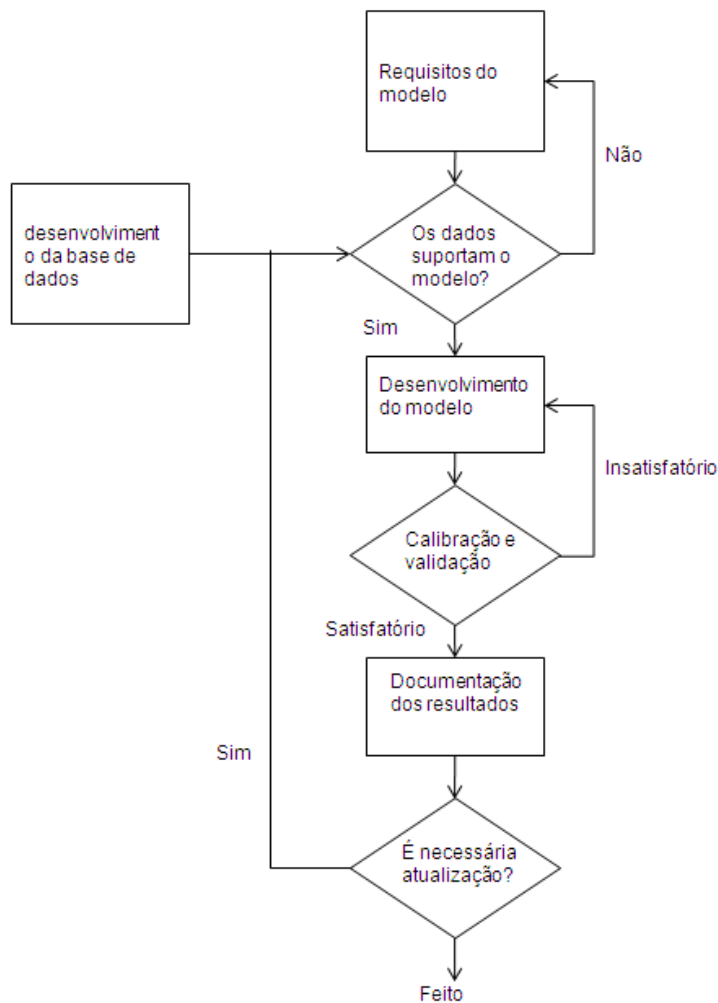


Figura 2 - Fluxo de atividades para o desenvolvimento para um modelo de estimativa de custos
Fonte: ISPA (2008)

O desenvolvimento do banco de dados é um fator chave para se encontrar um bom modelo paramétrico. Como já dito anteriormente, o modelo se baseia em dados históricos para realizar previsões futuras.

Todo modelo paramétrico apresenta pelo menos uma CER que é sempre uma relação matemática que envolve números sendo sua forma mais comum o custo como uma função das variáveis escolhidas como independentes. É importante citar que caso haja várias variáveis independentes essas devem apresentar baixa correlação entre si, evitando assim problemas de multicolinearidade, assunto que será tratado na seção 2.2.

Durante o desenvolvimento do modelo é importante especificar os métodos de estimação usados para se alcançar os resultados, identificar os elementos que serão estimados e definir os dados de entrada e os elementos da EAP.

O modelo deve ser calibrado e validado antes de ser usado no propósito de gerar estimativas. Modelos baseados em dados históricos já estão calibrados, no entanto, a validação se faz necessária para se garantir uma boa previsibilidade, robustez e o alinhamento com as premissas adotadas.

Da mesma forma a documentação é de extrema importância para se manter um histórico das mudanças ocorridas no modelo, dada a sua constante atualização e facilidade de realizar a mesma.

Os resultados encontrados podem servir para um grande número de situações. Algumas delas são propostas de preço, estimativas pouco precisas de custo e estudos dos impactos de mudanças no projeto. Pode também ser usado em gerenciamento de projetos, como uma base de comparação com outros métodos de estimativa de custo e em situações de concorrência, onde é importante a tentativa de replicar os custos do seu adversário.

Após definir a metodologia deve-se fazer a escolha ou construção do modelo levando em consideração os três passos seguintes:

- Revisar os modelos existentes e fazer uma seleção. Caso não exista nenhuma alternativa aceitável analise a opção de criar um modelo;
- Garantir que o modelo é validado e que contempla todos os custos; e,
- Estar preparado para a defesa do modelo escolhido.

Em relação à coleta e tratamento de dados, são quatro as atividades relacionadas:

- Identificar a necessidade de dados e a potencial fonte de dados;
- Revisar e pesquisar as fontes de dados para obter informações;
- Conduzir a análise do cronograma do projeto; e,
- Fazer o tratamento dos dados.

2.1.1.3. A execução da estimativa

A execução das estimativas é composta de cinco tarefas:

- Desenvolver a estimativa;
- Desenvolver e incorporar a avaliação de riscos;

- Documentar estimativas probabilísticas;
- Apresentar os resultados; e,
- Fazer a atualização das estimativas nas bases de dados.

Existem oito atividades associadas com desenvolvimento da estimativa:

- Popular o modelo com os dados tratados;
- Verificar se as regras básicas e premissas foram cumpridas;
- Garantir que a estimativa contempla todos os custos;
- Rodar o modelo para calcular a estimativa do custo;
- Ajustar a estimativa à fase do tempo;
- Ajustar a estimativa pela inflação;
- Conduzir qualquer verificação cruzada ou estimativa de reconciliação; e,
- Desenvolver ou atualizar uma série de custos para estimativas prévias ou independentes.

2.2. ANÁLISE MULTIVARIADA

É possível caracterizar a análise multivariada como a que “se refere a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas de cada indivíduo ou objeto sob investigação” (HAIR *et al.*, 2010). Seu objetivo é fazer medições, explicações e predições, além de testes de hipóteses e prever o grau de relação entre as variáveis. É utilizada quando se tem várias variáveis independentes ou quando existem várias variáveis dependentes correlacionadas entre si. Sua principal vantagem é o fato de apresentar uma análise conjunta, ao invés de várias análises individuais ou bivariadas.

2.2.1. CONCEITOS INICIAIS

Algumas conceituações iniciais sobre variáveis, escalas de medidas, erros de medida, significância e poder estatístico são necessários. As variáveis podem ser divididas em três grupos:

- A Variável Dependente é o efeito presumido, ou uma resposta a uma mudança em uma variável independente;

- A Variável independente é a causa presumida de uma mudança na variável dependente; e,
- A Variável Estatística, também chamada de *Variate*, é uma combinação linear, com pesos empiricamente determinados, aplicado a um conjunto específico de variáveis. Em outras palavras, é um modelo estatístico criado para explicar a relação entre variáveis dependentes e independentes.

2.2.1.1. Escala de medida

A escala de medida é a forma como as variáveis são mensuradas. Elas são essenciais para a escolha do método multivariado apropriado a ser utilizado e são classificadas em duas categorias, métricas e não métricas, cada uma com mais dois níveis. As escalas de medida estão resumidas na Figura 1:

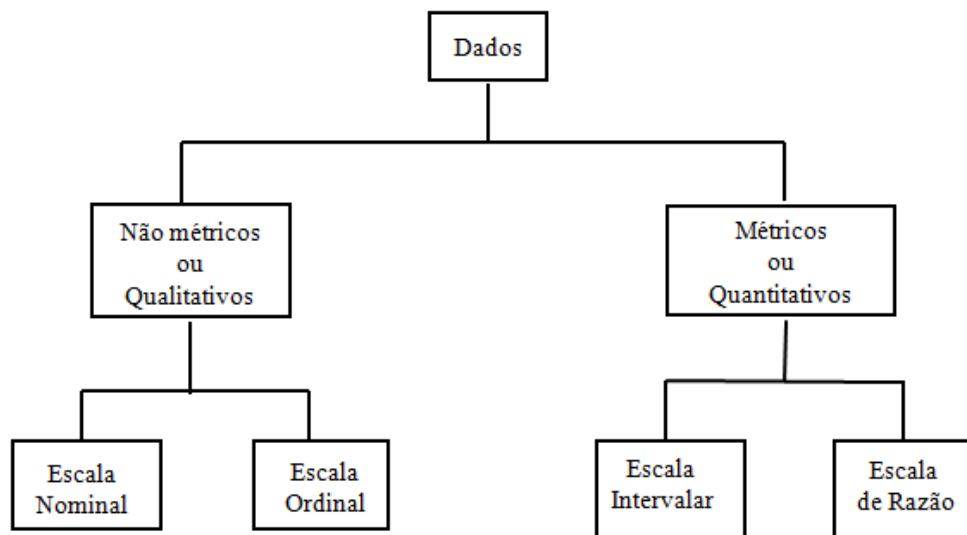


Figura 3 - Esquema de escala de medidas
Fonte: O autor

As escalas não métricas, também conhecidas como qualitativas, descrevem diferenças em tipo ou natureza, indicando assim se há a presença ou não de algum atributo que está sob observação. Podem ser nominais quando são usados números para rotular ou identificar indivíduos ou objetos e ordinais quando além de identificar esses números servem para ranquear ou ordenar os objetos ou indivíduos do estudo em relação à quantidade do atributo possuído.

As escalas métricas, também conhecidas como escalas quantitativas, refletem a diferença em quantidade ou em grau em relação a uma característica ou propriedade particular. Também são divididas em dois grupos, escalas intervalares e escalas de razão, nos quais as unidades de medida são constantes, sendo a única diferença entre elas o fato de que as intervalares possuem um ponto zero arbitrário, enquanto as de razão possuem um ponto zero absoluto. Essa categoria fornece uma precisão de medida alta, permitindo que praticamente todas as operações matemáticas possam ser feitas.

É importante compreender os diferentes tipos de medidas por duas razões, evitar a utilização de dados métricos como não métricos e vice-versa e para determinar quais técnicas multivariadas são mais aplicáveis aos dados.

2.2.1.2. Erro de medida

Erro de medida é “o grau em que os valores observados não são representativos dos valores verdadeiros” (HAIR *et al.*, 2010). Existem diferentes fontes de erros, dessa forma, é necessário que certo grau de erro de medida seja acrescido a cada uma das variáveis usadas em técnicas multivariadas.

Procurando reduzir o impacto desses erros de medida, duas características devem ser levadas em consideração: a validade e a confiabilidade. A validade é “o grau em que uma medida representa precisamente aquilo que se espera” (HAIR *et. al*, 2010). Caso a validade esteja garantida é preciso analisar a confiabilidade, o grau em que a medida está livre de erro medindo assim o valor verdadeiro, em outras palavras é o contrário do erro de medida.

Outra alternativa para mitigar os erros de medida é utilizar medidas multivariadas, também conhecidas como escalas múltiplas, nas quais diversas variáveis são reunidas para representar um conceito. O uso de várias variáveis como indicadores permite que a informação seja mais precisa assim evitando que toda a confiança seja dada em apenas uma característica, dividindo, dessa forma, entre várias.

O impacto proporcionado pelo erro de medida não está explícito, já que esse está embutido nos dados obtidos, assim é preciso buscar para o aumento da validade e confiabilidade das observações. Uma redução nos erros de medida pode melhorar e fortalecer resultados fracos.

2.2.1.3. Significância e poder estatístico

Com exceção da análise de agrupamento (*clusters*) e mapeamento perceptual, todas as demais técnicas multivariadas são baseadas em inferência estatística dos valores ou relações entre dados da população ou da amostra com dados aleatórios extraídos de uma população.

Para interpretar essas interferências é preciso definir níveis de erros estatísticos aceitáveis para a utilização de uma amostra, dado que é muito complicada a obtenção de todos os dados de uma população. A prática mais comum é determinar o erro Tipo I, chamado também de alfa (α) que é a probabilidade de se rejeitar a hipótese nula quando essa é verdadeira, sendo o valor de α o limite de erro permitido. Junto com o nível do Tipo I se determina o nível de erro Tipo II (β) que é a probabilidade de não se rejeitar a hipótese nula quando essa é falsa. Outra probabilidade que também é determinada é o poder do teste que representado por $1-\beta$, que é a probabilidade de rejeitar corretamente a hipótese nula.

Apesar do α estabelecer o nível de significância, é β quem determina o poder do teste, a probabilidade de sucesso em encontrar divergências caso essas realmente existam. No entanto, os erros Tipo I e Tipo II são inversamente relacionados, ou seja, enquanto a chance do erro Tipo I ocorrer diminui, a probabilidade de ocorrer um erro do Tipo II aumenta, diminuindo o poder do teste. Dessa forma deve se buscar um equilíbrio entre o nível de α e β .

A Tabela 5 apresenta os diferentes tipos de erros e suas probabilidades.

Tabela 5 - Tipos de erros

	H ₀ verdadeiro	H ₀ falso
Não rejeitar H ₀	1- α	β Erro Tipo II
Rejeitar H ₀	α Erro Tipo I	1- β Poder

Fonte: o Autor

O poder não é uma função apenas de α , ele é determinado também por outros dois fatores, o tamanho do efeito, a estimativa do grau que o fenômeno exerce na população, e o tamanho da amostra retirado da mesma. Estudos como de Cohen (1988)

sugerem que se devem buscar sempre níveis de α de pelo menos 5% com níveis de poder de 80%, para isso, os três fatores citados anteriormente devem ser considerados.

2.2.2. TÉCNICAS MULTIVARIADAS

Para a seleção da técnica mais adequada é necessário responder a três perguntas básicas:

- As variáveis podem ser divididas em dependentes e independentes?
- Se for possível, quantas variáveis são dependentes em cada análise?
- E como as variáveis, tanto dependentes como independentes, são medidas?

A resposta da primeira pergunta determina se serão usadas técnicas de dependência ou de interdependência. Técnicas de dependência se caracterizam por utilizarem uma variável ou um conjunto de variáveis como dependente, a ser prevista ou explicada por outras variáveis independentes. Em contra partida, técnicas de interdependência são aquelas onde nenhuma variável ou conjunto de variáveis é definida como dependente ou independente, todas as variáveis são analisadas de forma conjunta.

A Figura 4 apresenta um esquema com as principais técnicas multivariadas utilizadas.

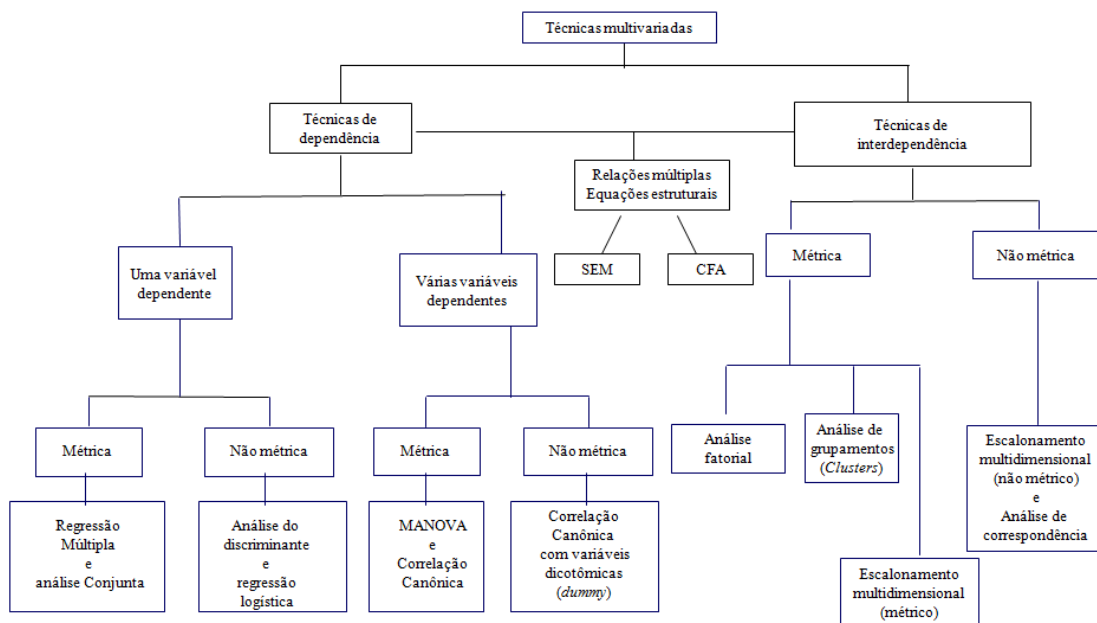


Figura 4 - Técnicas de análise multivariadas
Fonte: Hair et Al., 2010

Para a escolha da técnica de dependência apropriada deve levar em conta o número de variáveis dependentes e a escala de medida empregada pelas variáveis. Os procedimentos de dependência são muito similares entre si, dessa forma a Tabela 6 tem a intenção de definir esses procedimentos de acordo com a natureza e quanto ao número de variáveis dependentes e independentes.

Tabela 6 - Relação entre métodos de dependência

Correlação Canônica	$Y_1 + Y_2 + Y_3 \dots + Y_n = X_1 + X_2 + X_3 \dots + X_n$ (Métrico, não métrico) (Métrico, não métrico)
Análise Multivariada de Variância	$Y_1 + Y_2 + Y_3 \dots + Y_n = X_1 + X_2 + X_3 \dots + X_n$ (Métrico) (não métrico)
Análise de Variância	$Y_1 = X_1 + X_2 + X_3 \dots + X_n$ (Métrico) (não métrico)
Análise Discriminante Múltipla	$Y_1 = X_1 + X_2 + X_3 \dots + X_n$ (não métrico) (Métrico)
Análise de Regressão Múltipla	$Y_1 = X_1 + X_2 + X_3 \dots + X_n$ (Métrico) (Métrico, não métrico)
Análise Conjunta	$Y_1 = X_1 + X_2 + X_3 \dots + X_n$ (Métrico, não métrico) (não métrico)
Modelagem de Equações Estruturais	$Y_{1.} = X_{11} + X_{12} + X_{13} \dots + X_{1n}$ $Y_{2.} = X_{21} + X_{22} + X_{23} \dots + X_{2n}$ (Métrico) (Métrico, não métrico)

Fonte: HAIR *et al.*, 2010

Com os dados da Tabela 6, percebe-se que a correlação canônica pode ser vista como o modelo mais geral enquanto a modelagem de equações estruturais a mais especializada. A cada restrição imposta uma nova técnica se torna mais apropriada.

No caso de técnicas de interdependência, não há classificação das variáveis como dependentes ou independentes, pois todas são analisadas simultaneamente em busca de uma estrutura que englobe todo o conjunto de dados. Se a estrutura é o objeto de estudo deve-se utilizar análise fatorial, se os casos devem ser agrupados então se usa a análise de agrupamentos e caso a estrutura de objetos for o desejado, usa-se o escalonamento multidimensional, também conhecido como técnicas de mapeamento perceptual.

Cada uma das principais técnicas multivariadas será analisada individualmente, sendo as listadas a seguir:

- Análise fatorial;
- Regressão múltipla e correlação múltipla;
- Análise do discriminante múltipla e regressão logística;
- Análise da correlação canônica;

- Análise multivariada de variância e covariância (MANOVA e MANCOVA);
- Análise conjunta;
- Análise de agrupamentos (clusters);
- Mapeamento perceptual (escalonamento multidimensional);
- Análise de correspondência; e,
- Modelagem de equações estruturais e análise fatorial confirmatória.

2.2.2.1. Análise fatorial

A análise fatorial inclui a análise de componentes principais e a análise dos fatores comuns e é usada para “analisar as inter-relações de um grande número de variáveis e explicar essas variáveis em termos de suas dimensões inerentes comuns (fatores)” (HAIR *et al.*, 2010). Ela busca um meio de reunir as informações inerentes a diversas variáveis em um menor número de variáveis estatísticas, os fatores, reduzindo ao máximo a perda de informação. Essa técnica será melhor explicada na seção 2.2.3, mais a frente, por ser uma das técnicas utilizadas no estudo.

2.2.2.2. Regressão múltipla

A regressão múltipla é usada quando o problema apresenta uma variável dependente métrica relacionada com duas ou mais variáveis independentes métricas. Seu método mais utilizado é o dos mínimos quadrados e tem como objetivo prever as respostas da variável dependente em relação a mudanças nas variáveis independentes. Essa técnica será mais bem explicada na seção 2.2.4 por ser bastante utilizada no estudo.

2.2.2.3. Análise discriminante múltipla e regressão logística

A análise discriminante múltipla é usada quando a variável dependente é não métrica dicotômica ou multicotômica, as independentes são métricas e a amostra total pode ser separada em grupos baseados em variáveis dependentes não métricas que representam classes conhecidas. Seus objetivos são analisar a diferença entre grupos, assim como prever a probabilidade de um elemento pertencer a uma classe.

Os modelos de regressão logística, também conhecidos como análise logit, é uma mistura entre a regressão múltipla e a análise discriminante múltipla, pois usa variáveis independentes para prever uma única variável dependente, no entanto, essa variável dependente é não métrica. A análise logit pode utilizar variáveis não métricas e métricas como independentes e não exige a suposição de normalidade multivariada.

2.2.2.4. Correlação canônica

A correlação canônica é uma extensão da regressão múltipla, pois utiliza simultaneamente diversas variáveis dependentes métricas e diversas variáveis independentes métricas. Seu princípio é desenvolver uma combinação linear para cada conjunto de variáveis e maximizar a correlação entre os dois conjuntos.

2.2.2.5. Análise multivariada de variância e covariância

A análise multivariada de variância (MANOVA) é uma extensão da análise univariada de variância (ANOVA) e é usada para explorar de forma simultânea as relações entre duas ou mais variáveis dependentes métricas ou entre diversas variáveis independentes categóricas. Sua utilização se dá quando se busca testar hipóteses referentes à variância em grupos de duas ou mais variáveis dependentes métricas. Para remover o efeito de variáveis independentes métricas não controladas sobre variáveis dependentes pode-se usar a análise multivariada de covariância (MANCOVA).

2.2.2.6. Análise conjunta

A análise conjunta é uma técnica voltada para avaliação de produtos e serviços, seu objetivo é fazer avaliações complexas mantendo opções realistas para o respondente.

2.2.2.7. Análise de Agrupamentos

A análise de agrupamentos busca a geração de grupos significativos de dados, seu objetivo é classificá-los em segmentos mutuamente excludentes com base em características comuns. Difere em relação à análise de discriminante, pois seus grupos não são pré-definidos, ou seja, a técnica é utilizada para formar esses grupos.

Essa técnica é dividida em três etapas, primeiramente se faz uma busca por similaridades ou associações entre os dados para determinar quantos grupos compõem a amostra. Em seguida faz-se o agrupamento no qual os dados são particionados e por último é estabelecido o perfil das variáveis para determinar sua composição.

2.2.2.8. Mapeamento perceptual

O mapeamento perceptual também conhecido como escalonamento multidimensional tem como objetivo representar em espaços multidimensionais julgamentos de consumidores sobre preferências ou similaridades. Objetos selecionados como mais semelhantes ficam mais próximos entre si, no entanto, outras análises são necessárias para se determinar quais atributos definem a posição de cada objeto.

2.2.2.9. Análise de correspondência

A análise de correspondência facilita o mapeamento perceptual para dados com um conjunto de atributos não métricos, sendo sua forma mais simples uma tabela de contingência que nada mais é que uma tabulação cruzada de duas variáveis categóricas. Dessa forma transforma os dados não métricos em métricos e reduz a dimensão do mapeamento perceptual agindo de forma análoga a uma análise fatorial.

2.2.2.10. Modelagem de equações estruturais e análise fatorial confirmatória

A modelagem de equações estruturais (SEM) também conhecida como LISREL permite segmentar cada conjunto de variáveis dependentes e achar relações. É a técnica multivariada mais eficiente para a estimativa de uma série de equações de regressão múltipla simultaneamente. É dividida em dois componentes básicos, o modelo estrutural e o modelo de medida.

O modelo estrutural relaciona as variáveis independentes com as dependentes, se diferencia de outros métodos que permitem múltiplas variáveis dependentes, pois nos outros casos os modelos só permitem uma relação entre as variáveis dependentes e independentes. O modelo de medida é usado com diversas variáveis, indicadores, para cada variável dependente ou independente. Também conhecido como análise fatorial confirmatória, fornece a possibilidade de distinguir a contribuição de cada item de uma

escala e como a escala mede a confiabilidade. É semelhante a usar a análise fatorial nos itens da escala, pegar os resultados encontrados e gerar uma regressão.

2.2.3. ANÁLISE FATORIAL

Como já descrito anteriormente, segundo HAIR *et al.*(2010) a análise fatorial é uma técnica de interdependência cujo objetivo principal é definir a estrutura inerente entre as variáveis. Essa técnica fornece os meios para se analisar a estrutura inter-relacional em um grande número de variáveis, definindo conjuntos de variáveis chamados fatores. Esses fatores possuem alta intercorrelação, sendo considerados representações de dimensões dentro dos dados.

A análise fatorial pode ser exploratória ou confirmatória. A principal diferença é que a primeira não se oferecem restrições a priori sendo assim útil na estruturação de um conjunto de variáveis ou como método de redução de dados. Enquanto a segunda é feita com hipóteses já formadas sobre a estrutura dos dados, seja por referencial teórico ou por pesquisas passadas, sendo usada como uma forma de avaliação para confirmar a estrutura apresentada. Esse estudo se até a sua forma exploratória.

2.2.3.1. O processo de decisão

O processo de decisão da análise fatorial é dividido em sete estágios:

- Objetivos da análise fatorial;
- Planejamento de uma análise fatorial;
- Suposições em uma análise fatorial;
- Determinação de fatores e avaliação do ajuste geral;
- Interpretação dos fatores;
- Validação da análise fatorial; e,
- Usos adicionais dos resultados da análise fatorial.

2.2.3.2. Objetivos da análise fatorial

Seu objetivo é condensar a informação contida em diversas variáveis originais em um número menor de conjuntos de fatores com a perda mínima de informação. Ao alcançar seus objetivos, a análise fatorial é ajustada de acordo com quatro informações: a especificação da unidade de análise; a obtenção do resumo de dados e ou redução dos

mesmos; a seleção de variáveis e o uso de resultados da análise fatorial com outras técnicas multivariadas.

A especificação da unidade de análise se faz necessária, pois a análise fatorial pode identificar a estrutura de relações entre variáveis ou entre respondentes. Dessa forma existe uma diferenciação entre os dois casos, no primeiro, é utilizada uma matriz de correlação de variáveis e é chamada análise fatorial R, no segundo é utilizada uma matriz de correlação dos respondentes sendo chamada de análise fatorial Q.

O resultado obtido com o resumo de dados, apesar de semelhante, difere da redução de dados. Enquanto o resumo tem como resultado estruturas que descrevem os dados em um número muito menor de conceitos que as variáveis inicialmente utilizadas, a redução estende o conceito até a obtenção de um escore fatorial para cada fator substituindo seu valor original. O resumo tem uma ideia de conceito enquanto a redução pode ser feita selecionando-se as variáveis mais significativas ou com a criação de novas.

A seleção de variáveis é importante porque se escolhem as dimensões que serão encontradas pela análise fatorial além de evitar o problema de geração de resultados espúrios com a entrada de *inputs* de forma indiscriminada.

A análise fatorial é um excelente marco para iniciar outras técnicas multivariadas, pois além usar as inter-relações de forma direta, ainda evita que variáveis altamente correlacionadas afetem os procedimentos posteriores, pois essas possivelmente se encontraram dentro do mesmo fator. Dessa forma reduz problemas relativos à grande número de variáveis e de altas intercorrelações entre elas.

2.2.3.3. Planejamento de uma análise fatorial

O planejamento de uma análise fatorial leva em consideração três fatores: o cálculo dos dados de entrada; o número de variáveis e suas características e o tamanho da amostra.

O cálculo dos dados de entrada envolve basicamente a escolha do tipo de análise entre R e Q, sendo Q o caso menos utilizado e que exige um maior tratamento e cuidado com os cálculos, principalmente para diferenciá-lo de outras técnicas multivariadas. No caso do tipo R, mais comum, se utiliza uma matriz normal de correlação.

Em relação às características das variáveis, deve se levar em consideração quais podem ser utilizadas e quantas devem ser incluídas na análise. A única restrição para o

tipo de variável é que a correlação tem que ser calculada em todas elas, o que não é um problema em relação às variáveis métricas, para variáveis não métricas é aconselhável definir variáveis dicotômicas para seu lugar. A quantidade de variáveis incluídas deve ser minimizada sem comprometer a composição dos fatores. Deve-se também identificar variáveis chaves que reflitam os fatores previstos hipoteticamente para ajudar na validação dos fatores determinados.

O tamanho da amostra deve ser preferencialmente maior que cem observações, e ter como regra cinco a dez vezes mais observações do que dados a serem analisados. Deve-se buscar um conjunto de variáveis mais parcimonioso que leve em consideração conceitos e práticas.

2.2.3.4. Suposições na análise fatorial

As suposições feitas acerca da análise fatorial são de natureza conceitual e de natureza estatística. Diferente de outras técnicas multivariadas, na análise fatorial a exigência conceitual é mais forte que as estatísticas, pois essa se centra no caráter e composição das variáveis.

Em relação a questões conceituais, duas suposições devem ser consideradas: a que existe alguma estrutura subjacente no conjunto de variáveis escolhidas e a que a amostra deve ser homogênea com relação à estrutura fatorial.

Em relação a questões estatísticas, os desvios em relação a normalidade, a homocedasticidade e a linearidade são críticos, pois só interferem nas correlações observadas.

Após a aceitação das premissas básicas é necessário garantir que as variáveis são suficientemente correlacionadas para a composição dos fatores, para isso, podem ser usadas medidas gerais ou medidas específicas de intercorrelação. Nas medidas gerais destacam-se a inspeção visual, o teste de esfericidade de Barlett e a medida de adequação da amostra (MSA). Para análises individuais podem se estender os métodos de exame visual e MSA.

2.2.3.5. Determinação de fatores e avaliação do ajuste geral

Para a determinação de fatores e a avaliação do ajuste geral, deve-se escolher o método de extração, ou análise dos fatores comuns, ou análise de componentes e o número de fatores usados na explicação da estrutura desenvolvida para os dados.

A escolha do método de extração deve considerar os objetivos da análise fatorial assim como a relação entre as variáveis. Como essa análise faz o agrupamento de diferentes variáveis em fatores é necessário entender o quanto de variância de uma variável é compartilhado com as outras variáveis que compõem esse mesmo fator. A variância pode ser dividida em três: variância comum, variância específica e variância de erro. A análise de componentes é usada quando se tem como objetivo resumir a maior parte da informação original num menor número de fatores para previsão. A análise de fatores comuns é usada para identificar fatores que as variáveis têm em comum.

Em relação à quantidade de fatores a serem extraídos, nenhum método quantitativo exato foi desenvolvido, deve-se combinar conceitos e evidências empíricas. Alguns critérios podem ser úteis para decidir o momento de estabilizar o modelo:

- Critério da raiz latente;
- Critério a priori;
- Critério de percentagem de variância; e,
- Critério do teste de scree.

2.2.3.6. Interpretação dos fatores

Três processos são considerados os mais atraentes para a interpretação de uma estrutura fatorial assim como a escolha da solução final. A interpretação fatorial é circular, assim podem ser necessárias diversas interações até que a solução final seja obtida.

O primeiro passo é a estimativa da matriz fatorial, na qual a matriz fatorial inicial não rotacionada contendo as cargas fatoriais para cada variável é estimada, sendo carga fatorial a correlação entre a variável e o fator.

Em seguida é feita a rotação de fatores com a intenção de simplificar e aumentar a significância dos resultados, além de em diversos casos retirar ambiguidades existentes.

Por último é feita a interpretação e reespecificação de fatores, onde as cargas fatoriais já rotacionadas são avaliadas em relação a sua essência na formação da estrutura fatorial. Alguns problemas podem gerar a necessidade de reespecificar o modelo fatorial:

- Eliminação de variáveis na análise;
- Desejo de se utilizar um método rotacional distinto;

- Necessidade de se extrair um número diferente de fatores; e,
- Anseio por mudar o método de extração.

A rotação de fatores pode ser feita de duas maneiras, os métodos rotacionais ortogonais e os métodos de rotação oblíqua. Não existe nenhuma razão analítica para escolha específica de um método, diversas considerações devem orientar essa escolha.

Na interpretação de fatores é preciso decidir quais cargas fatoriais devem ser consideradas. Para isso almeja-se uma garantia de significância prática e se realizam avaliações da significância estatística. A interpretação de uma matriz fatorial se dá em cinco etapas:

- Examinar a matriz fatorial de cargas;
- Identificar as cargas significantes para cada variável;
- Avaliar as comunalidades das variáveis;
- Reespecificar o modelo fatorial se necessário; e,
- Rotular os fatores.

2.2.3.7. Validação da análise fatorial

A avaliação da análise fatorial envolve o grau de generalidade dos resultados em relação à população e a influência de casos individuais sobre os resultados gerais. Algumas questões importantes sobre o delineamento de pesquisa e sobre as características dos dados podem se destacar:

- Utilizar uma perspectiva confirmatória, para avaliar a repetitividade dos resultados;
- Avaliar a estabilidade da estrutura fatorial; e,
- Detectar observações influentes.

2.2.3.8. Usos adicionais dos resultados da análise fatorial

A utilização de usos adicionais só se faz necessária caso o objetivo do trabalho seja a aplicação em seguida de outras técnicas multivariadas, pois nesse caso alguma forma de redução de dados será utilizada. São três as opções de redução de dados:

- Seleção de variáveis substitutivas;
- Criação de escalas múltiplas; e,

- Cálculo dos escores fatoriais.

2.2.4. REGRESSÃO MÚLTIPLA

A regressão múltipla é “uma técnica estatística que pode ser usada para se analisar a relação entre uma única variável depende (critério) e várias variáveis independentes” (Hair *et al.*,2010). Ela pode ser representada pela equação 1

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n + e \quad (1)$$

Sendo:

Y = variável dependente

b_0 = intercepto e é uma constante

b_1 = coeficiente da regressão associado a variável X_1

b_2 = coeficiente da regressão associado a variável X_2

X_1 = variável independente 1

X_2 = variável independente 2

e = Erro da previsão

O coeficiente de regressão é a variação estimada na variável dependente por uma variação unitária no valor da variável independente. O intercepto pode ter duas interpretações, pode significar o valor na ausência da característica associada às variáveis quando as variáveis podem apresentar valores nulos ou é simplesmente uma medida para a obtenção de melhores resultados. O erro de previsão é a diferença entre o valor real e o valor estimando pela regressão.

A avaliação do modelo de previsão costuma ser feito pelo coeficiente de determinação, conhecido como R^2 . Ele é calculado pela razão entre a soma total dos quadrados subtraída da soma de erros quadrados e a soma dos quadrados explicados pela regressão. Sua relações podem ser vistas na equação 2 e na equação 3.

$$\sum_{i=1}^N (Y_i - \bar{Y})^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^N (Y_i - \bar{Y})^2 \quad (2)$$

$$SS_T = SS_E + SS_R$$

$$R^2 = \frac{SS_T - SS_E}{SS_R} \quad (3)$$

Sendo:

Y = variável dependente

\bar{Y} = valor esperado de Y , seu valor médio

\hat{Y} = valor previsto pela regressão para Y

R^2 = Coeficiente de determinação

SS_T = soma total dos quadrados

SS_E = soma de erros quadrados

SS_R = soma dos quadrados explicados

A elaboração de uma análise de regressão múltipla pode ser dividida em seis estágios:

- Objetivos;
- Planejamento de pesquisa;
- Análise das suposições;
- Estimativa e avaliação do ajuste geral do modelo;
- Interpretação da variável estatística; e,
- Validação dos resultados.

2.2.4.1. Objetivos da regressão múltipla

O primeiro passo para a realização de uma regressão múltipla é avaliar se o problema proposto pode ser solucionado por esse método. Em seguida deve-se decidir seu propósito que pode ser de caráter preditivo ou explicativo. Uma mesma regressão apresenta em si os dois resultados, variando apenas a forma como os resultados serão interpretados.

A escolha das variáveis a serem utilizadas na regressão é de suma importância, e deve sempre levar em consideração três fatores, a teoria, o erro de medida e o erro de especificação.

2.2.4.2. Planejamento de pesquisa

Para o planejamento de pesquisa, três características das amostras escolhidas são determinantes: o tamanho, aspectos especiais de relação entre as variáveis e a natureza das variáveis independentes.

O tamanho da amostra é importante, pois é essencial na determinação do poder estatístico e para a generalização dos resultados encontrados.

Algumas variáveis independentes podem não seguir os pressupostos básicos de serem métricas e possuírem uma relação linear com a variável dependente. Dessa forma alguns tratamentos podem ser feitos, sendo possível destacar:

- A transformação de variáveis;
- A incorporação de dados não métricos com variáveis dicotômicas;
- A representação de efeitos curvilíneos com polinômios; e,
- A representação de efeitos de interação ou moderadores.

A natureza da variável se refere ao fato de ela ser fixa ou aleatória. Apesar de diferentes, as análises estatísticas podem ser feitas da mesma forma em ambos os casos.

2.2.4.3. Análise das suposições

Para que os coeficientes de uma regressão múltipla sejam válidos é necessário que as suposições da regressão sejam atendidas. As suposições de uma regressão múltipla a serem atendidas:

- A linearidade do objeto de estudo;
- A variância dos erros ser constante, conhecida como homocedasticidade;
- A independência dos elementos de erro; e,
- A normalidade dos elementos de erro.

A principal diferença entre a avaliação das variáveis de forma individual e a regressão múltipla é o fato da variável estatística depois de determina a influência de forma coletiva na previsão da variável dependente. Dessa forma, é necessário testar as suposições tanto para as variáveis individuais como para a variável estatística.

A necessidade de se examinar todas as variáveis se dá pelo fato de o mesmo com violações de suposições suas relações podem não sofrer com má interpretação e o fato da necessidade de caso haja violação na variável estatística quais são as possíveis ações corretivas.

A análise dos resíduos é a principal forma de se identificar possíveis não atendimentos as premissas, e dessa forma necessitar algum tipo de tratamento nos dados. Resíduos são entendidos como a diferença entre valores observados e os valores

previstos para a variável dependente. É altamente recomendado que os resíduos sejam padronizados para evitar problemas de escala, pois dados com ordem de grandeza maior tem a tendência de apresentarem maiores valores residuais, dessa forma podendo comparar diferentes variáveis. A padronização mais utilizada são os resíduos “estudentizado”, cujos valores se tornam valores t , sendo assim numa mesma ordem de grandeza.

Segundo HAIR *et al.*, (2009) “a representação gráfica dos resíduos versus as variáveis independentes ou previstas é o método básico para identificar violações de suposições”. O gráfico mais utilizado é o gráfico de dispersão entre os resíduos e os valores dependentes previstos.

Na Figura 5 são apresentados os diferentes gráficos de dispersão onde estão representadas diferentes violação de pressupostos. O gráfico nulo é a representação na qual todas as premissas são atendidas, os resíduos estão dispersos de forma aleatória, com dispersão aparentemente igual, ao redor do valor nulo e sem nenhuma demonstração de uma concentração de valores acima ou abaixo desse valor. Os demais gráficos de dispersão representam diferentes tipos de violações. Para se analisar de forma individualizada as variáveis independentes e saber qual necessita de algum tratamento especial pode-se utilizar os gráficos de regressão parcial.

Linearidade é o grau de variação de uma variável dependente em relação a uma variável independente. Um padrão curvilíneo nos resíduos indica que existe a possibilidade de melhora de resultados caso seja tomada umas das ações corretivas, transformação dos dados, inclusão indireta de relações não lineares no modelo de regressão ou uso de métodos específicos para acomodar esse feito curvilíneo.

A heterocedasticidade é a existência de variâncias diferentes nos resíduos. Além da análise gráfica podem ser feitos outros testes, como o de Levine que aprecia se a variância para um par de variáveis é igual. Esse teste evita a influência da não normalidade para se averiguar a existência de heterocedasticidade. As duas formas de corrigir essa violação são fazer uma regressão ponderada ou fazer transformações de estabilização de variância.

Em relação à independência dos termos de erro, na regressão é suposto que os valores previstos não apresentam qualquer relação com outras previsões, não sendo sequenciados e nem possuindo um padrão. A ocorrência desse fenômeno em dados que apresentam sazonalidade é bastante comum. Para corrigir esse problema podem ser

feitas transformações como as primeiras diferenças, a adição de variáveis indicadoras ou o uso de modelos apropriados para esse tipo de dados.

A normalidade pode ser aferida pelo histograma de resíduos com uma comparação com a curva normal, no entanto, para amostras pequenas ele se torna pouco eficiente. O uso de gráficos de probabilidade normal é outro método que apresentam melhores resultados. Transformações também podem ser feitas para corrigir essa violação.

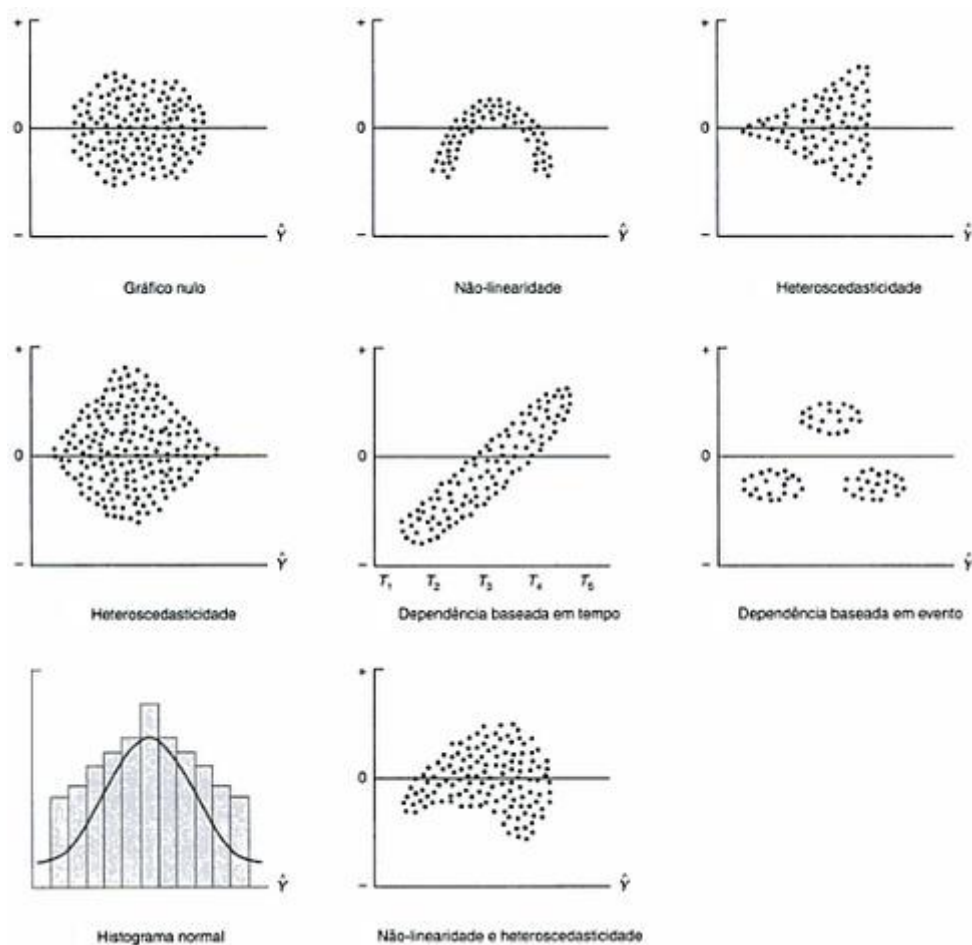


Figura 5 - Análise gráfica dos resíduos
 Fonte: HAIR *et al.*, 2009

2.2.4.4. Estimação e avaliação do ajuste geral do modelo

Nesse estágio, deve-se decidir o melhor método a ser utilizado para especificar o modelo de regressão, avaliar a significância do modelo na previsão da variável dependente e buscar observações que possam ter uma influência não desejada sobre os resultados.

Na especificação do modelo existem três grupos de técnicas, as de especificação confirmatória, os métodos de busca sequencial e a abordagem combinatória. No primeiro grupo as variáveis independentes a serem incluídas são totalmente escolhidas de forma teórica, dessa forma o pesquisador tem todo o controle sobre a entrada de variáveis independentes. Nos métodos de busca sequencial, as variáveis pré-escolhidas são incluídas ou retiradas do modelo seguindo uma regra até que se alcancem os níveis estatísticos desejados. As abordagens utilizadas são a estimação *Stepwise*, *forward* (adição) e *backward* (eliminação). A abordagem combinatória tem como principal método fazer a regressão entre todos os possíveis subconjuntos.

A estimação *stepwise* é a mais utilizada para a seleção de variáveis, a Figura 6 apresenta um esquema com seu fluxo de atividades.

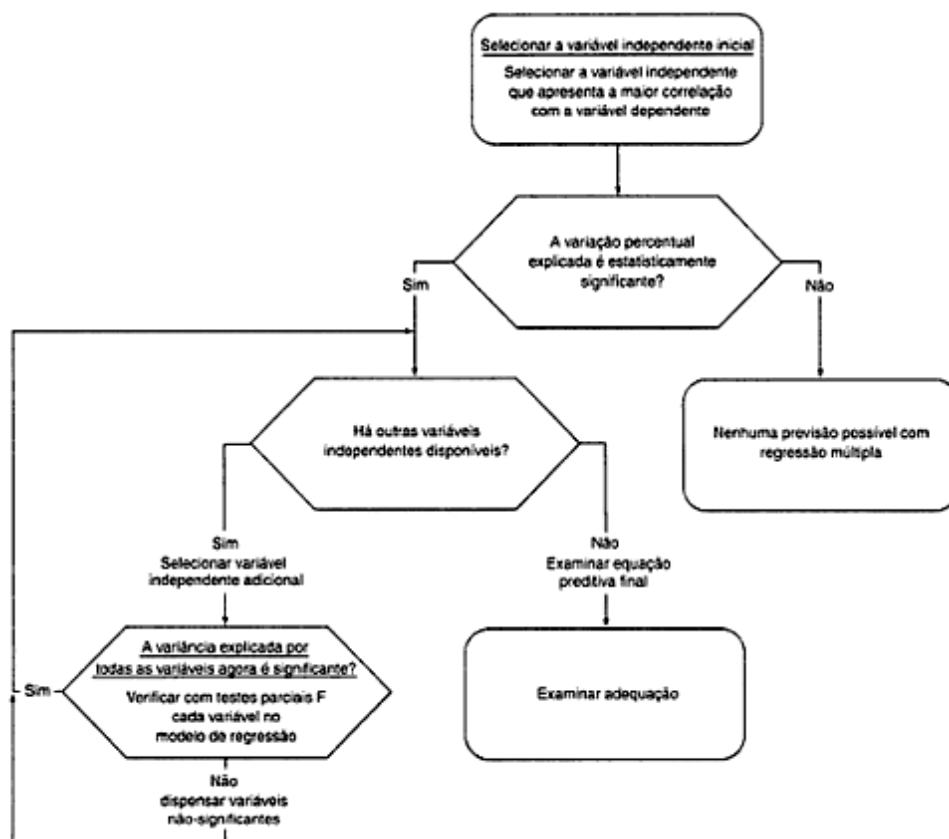


Figura 6 - Fluxograma do método de estimação *stepwise*
Fonte: HAIR *et al.*, 2009

Após a escolha de qual método será utilizado, deve-se testar a variável estatística satisfaz as suposições de regressão, repetindo o que foi apresentado no terceiro estágio. Caso sejam encontradas violações significativas será necessária à utilização de um dos

métodos de correção em algumas das variáveis independentes e logo após realizar novamente a regressão.

Em seguida deve-se avaliar a significância estatística do modelo encontrado. Para a análise geral aplica-se o teste do coeficiente de determinação que usa a razão F que afere se R^2 é significativamente maior que zero. Deve-se também realizar o ajuste de R^2 ao maior número de variáveis independentes e conseqüentemente sua diminuição do grau de liberdade.

Com a regressão geral selecionada, realizam-se testes de significância dos coeficientes encontrados. Para isso deve-se usar um intervalo de confiança, o erro padrão, a variação esperada dos coeficientes estimados, o tamanho da amostra e o nível de significância α . Então é feito o teste para se aferir se o valor de cada coeficiente é igual a zero, sendo assim não significativo ou diferente de zero e significativo.

Por último procuram-se observações que possam estar influenciando o modelo que podem ser classificadas em observações atípicas, pontos de alavancagem e observações influentes. Deve-se buscar justificativas para esses casos, e se evitar a eliminação dessas, no entanto, deve-se analisar a existência de uma mudança impactante nos resultados com a retirada de apenas uma observação.

Ações corretivas consistem em evitar erros em observações ou dados de entrada, e em eliminar casos válidos, mas que apresentam uma situação extraordinária. Caso exista uma observação excepcional sem explicação forte, deve-se estudar mais a fundo esse dado e fazer até dois modelos, um com a entrada e o outro sem. Caso a observação seja comum quando individualmente analisada, mas é excepcional em uma análise combinada, essa representa uma mudança conceitual e seu valor deve ser mantido para os estudos.

2.2.4.5. Interpretação da variável estatística

Para interpretar a variável estatística da regressão deve-se utilizar os coeficientes estimados. No caso da escolha de métodos sequenciais ou combinatórios é preciso estudar também a multicolinearidade, pois essa pode afetar o resultado das últimas variáveis entrantes no modelo.

Os coeficientes de regressão podem ser usados para a previsão de dados, assim como, para uma análise mais profunda que envolva também a explicação da natureza e no impacto das variáveis independentes na previsão. Para se evitar o problema de

comparar coeficientes que estão relacionados a variáveis que se apresentam em ordem de grandeza diferentes, é possível padronizá-los gerando assim a regressão beta. Essa padronização normaliza os dados, colocando-os com escala e variabilidade comuns, preferencialmente com média zero e desvio padrão igual a um. Os resultados desse procedimento são chamados de coeficientes beta (β) e eliminam o problema de se utilizar diferentes unidades de medida. Para sua utilização, dois cuidados devem ser levados em consideração, só podem ser usados dessa forma quando a multicolinearidade for mínima e os coeficientes betas não podem ser utilizados no contexto em que foram gerados.

A correlação entre as variáveis independentes é um problema, algumas transformações de variáveis podem gerar alta multicolinearidade. Assim é preciso avaliar o seu grau, determinar o seu impacto nos resultados e quando for o caso, aplicar as ações corretivas. Duas formas diretas de medir a multicolinearidade são a tolerância e o fator de inflação de variância. Seus principais impactos são sobre a predição, sobre a estimação dos coeficientes e sobre as explicações sobre os efeitos desses.

Algumas opções corretivas da multicolinearidade são, omitir uma ou mais variáveis que apresentam alta correlação entre si e inserir outras variáveis independentes menos correlacionadas para a melhora do modelo. É possível usar variáveis altamente correlacionadas apenas para modelos de previsão, sem a interpretação dos coeficientes encontrados. Outra forma é usar as correlações simples a fim de entender suas relações entre variáveis dependente e independentes ou usar métodos de maior complexidade como uma regressão Bayesiana ou uma regressão sobre componentes principais para se extrair o real efeito das variáveis independentes.

2.2.4.6. Validação dos resultados

Após se escolher o modelo mais adequado, deve-se analisar a sua representatividade para a população geral, sua generalização e sua transferibilidade. A melhor forma de se fazer isso é comparando o modelo encontrado com modelos teóricos já existentes ou com um conjunto de dados já validados sobre o mesmo tópico que serviria como uma espécie de grupo controle. No entanto, nem sempre existem teorias disponíveis ou resultados anteriores validados, assim alguns métodos empíricos podem ser usados para validar o modelo.

Dentre esses métodos são destacados o da utilização de amostras adicionais ou particionadas, do calculo da estatística PRESS, da comparação de diferentes modelos de regressão e da previsão com o modelo em um novo conjunto de valores para as variáveis.

3 SONDAS MARÍTIMAS

A atividade de sondagem se baseia no arrendamento de uma sonda a um representante ou uma companhia de petróleo, bem como suas operações para prospecção ou produção de petróleo.

A operação de sondas é complexa, englobando unidades marítimas ou terrestres de produção ou perfuração, sendo estas de grande ou pequeno porte com graus bastante variados de complexidade tecnológica. O custo diário do aluguel de sondas é muito instável, pois é muito influenciado pela situação do mercado, pelo preço do barril de petróleo e também devido ao seu tipo ou categoria.

Em função do aumento da demanda por equipamentos usados na fase de exploração de petróleo, os custos passaram por um aumento vertiginoso. Todas as matérias-primas da área de produção e exploração de petróleo, bem como todos os equipamentos sofreram tal efeito, contudo as sondas de perfuração *offshore* foram o recurso cujo custo mais aumentou nos últimos anos.

Na Tabela 7, podemos contemplar o aumento percentual de vários recursos desta área entre os anos de 2000 e 2007.

Tabela 7 - Aumento dos custos dos insumos e serviços de E&P entre 2000 e 2007

Insumo – Serviço de E&P	% de aumento
Sondas marítimas	356
Aço	149
Engenharia e gerência de projeto	103
Barco de instalação <i>offshore</i>	95
Depósito e fabricação	78
Equipamento	77
Mão de obra para construção	73
Materiais	49
Sondas em terra	42

Fonte: IHS - CERA

Adverso ao que ocorre com as plataformas de produção de gás e óleo, nenhuma empresa petrolífera possui frota própria de sondas de perfuração. As empresas que produzem gás e óleo afretam as embarcações e usam os serviços das empresas proprietárias de sondas por meio de contratos de duração variável.

A companhia de petróleo que afreta uma sonda à proprietária passa a ser a operadora daquele equipamento, visto que tem a função de definir quais poços serão perfurados, suas especificações e localização, durante a vigência do contrato.

As empresas proprietárias de sondas de perfuração têm duas opções de produção, ou fecham um contrato de operação de longo prazo e após isso procuram um estaleiro para efetuar a construção ou realizam uma compra especulativa, ficando exposto ao risco da demanda do mercado, iniciando a construção sem ainda ter um contrato firmado. Nessa situação a empresa está sujeita a um enorme risco de mercado, para um ativo de elevado investimento.

3.1. O MERCADO MUNDIAL DE SONDAS MARÍTIMAS

Em 2008, o mercado de sondas de perfuração, quando analisado em escala mundial era constituído por aproximadamente 140 empresas cujas sedes se espalham por todo o mundo. A frota de sondas de perfuração marítimas é internacional, móvel e é de posse concentrada em grandes empresas. A Tabela 8 apresenta os dados referentes ao ano de 2008 sobre a relação das empresas e a quantidade de sondas.

Tabela 8 - Relação de companhias e quantidade de sondas marítimas - 2008

	Companhias	Quantidade de Sondas contratadas	Frota total	% de utilização
Total	143	844	1197	67,9%
10 maiores	10	494	624	79,2%
% do total	7,0%	58,5%	50,2%	
20 maiores	20	625	826	75,7%
% do total	14,0%	74,1%	66,5%	
100 menores	100	100	217	46,1%
% do total	69,9%	11,8%	17,5%	

Fonte: Rigzone

Os números apresentados na Tabela 8 deixam evidente a concentração do mercado. As 10 maiores empresas do setor, dadas a quantidade total de sondas, detêm 58,5% das sondas contratadas nos dias de hoje. Em 2008, portanto, o cenário apresentado era que apenas 7,0% das companhias que possuem sondas de perfuração possuem 58,5% do mercado contratado, em contrapartida as 69,9% menores empresas possuem somente 11,8% da frota.

A recente alta dos preços de gás e óleo no mercado mundial possibilitou a exploração de diversos prospectos localizados principalmente em águas profundas ao

redor do mundo. Assim, grande parcela dos investimentos da indústria de sonda de perfuração tem focado nesta nova fronteira operacional. As últimas consideráveis descobertas de jazidas de petróleo, localizam-se em lâminas d'água cada vez maiores gerando assim uma previsão da demanda por sonda em lâminas d'água ultra-profundas em detrimento à demanda por sondas de outros segmentos. Os novos investimentos estão sendo alocados principalmente em navios-sonda e em plataformas semissubmersíveis em contra partida a utilização assim como o investimento em sondas do tipo *Jack-up* tem caído nos últimos anos, em decorrência deste tipo de embarcação estar limitada a lâminas d'águas rasas.

3.2. AS ESPECIFICAÇÕES DE UMA SONDA DE PERFURAÇÃO

Uma sonda de perfuração é o conjunto de equipamentos e acessórios que permitem a perfuração do poço de petróleo. Tal sonda é formada por diversos sistemas que possibilitam a geração de energia, a movimentação de cargas, a segurança do poço, a circulação de fluídos e o constante monitoramento de diversas condições no processo de perfuração.

A exploração e produção realizada em campos marítimos é mais cara e complexa envolvendo investimentos em aquisição ou construção da plataforma, perfuração, no desenvolvimento de poços, instalação de linhas flexíveis e dutos submarinos e infraestrutura de escoamento do petróleo ou gás para o continente (THOMAS, 2004).

Nas sondas marítimas o processo de perfuração é similar ao das terrestres, e seus equipamentos localizam-se sobre uma plataforma cujas características dependem da profundidade e condições marítimas. São, em sua maioria, unidades móveis que se movimentam por meio de propulsão própria ou com auxílio de rebocues, e devem flutuar com segurança, no entanto, devem permanecer estacionadas nas coordenadas do poço para operar (THOMAS, 2004).

Existem variados tipos de sondas, conforme evidenciado anteriormente, os quais possuem capacidades e especificações distintas. A análise organizará estas sondas em quatro subconjuntos: plataformas fixas, *Jack-ups*, navios-sonda e semissubmersíveis.

Na Figura 3 é possível visualizar os quatro tipos de sondas analisadas.

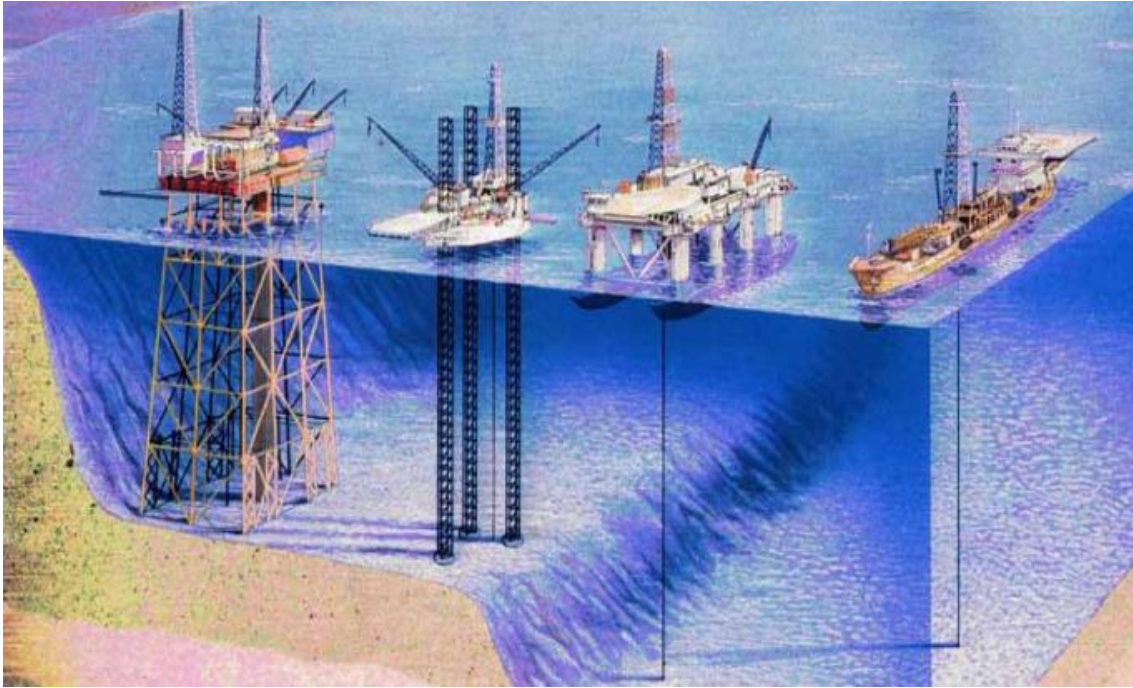


Figura 7 - Principais tipos de sondas *offshore*: fixas, *Jack-up*, semissubmersíveis e navio sonda.
Fonte: BOEMRE

As plataformas fixas são estruturas cujo material utilizado em sua construção pode ser de concreto ou aço e tem como característica marcante o fato de ser presas no fundo do mar.

Após a descoberta da reserva de hidrocarbonetos e tomada a decisão pela continuidade do projeto, a plataforma deve ser construída para perfurar outros poços, dando início à produção. Por serem fixadas no fundo do mar, as plataformas operam apenas em estreitas lâminas d'água e em geral são de grande porte quando comparadas com os outros tipos de sondas.

Os *Jack-ups*, também conhecidos como plataformas autoeleváveis (PA), operam em lâminas d'água de até 300 metros de profundidade. Sua principal característica é a presença de 3 ou 4 estruturas entrelaçadas ou tubulares conhecidas como pernas, que se apoiam no fundo do mar e suspendem a plataforma a uma altura acima do nível da água, distância conhecida como *air gap*, suficiente para evitar a interferência de ondas e o movimento das marés.

Seu casco, na maioria das vezes, possui forma retangular ou triangular e se desloca verticalmente ao longo das pernas, por meio de um sistema de macacos hidráulicos ou engrenagens. Dessa maneira, permite a flutuação para navegar ou o apoio sobre as pernas quando perfurando. Os equipamentos de segurança do poço situam-se no nível da plataforma, pouco acima da superfície do mar (THOMAS, 2004).

Tais plataformas podem ser de dois tipos: *Slot* ou *Cantilever*. No tipo *Slot* a torre se apoia sobre um rasgo localizado no casco da embarcação, possuindo pouca capacidade de deslocamento lateral e longitudinal; no tipo *Cantilever* há duas vigas deslizantes sobre trilhos nas quais a torre se movimenta para fora da borda do casco, permitindo maior área livre, maior flexibilidade de operação e a perfuração de vários poços a partir de uma mesma localização (THOMAS, 2004).



Figura 8 - Sonda do tipo *Jack-up* (*Cantilever*)
Fonte: Rigzone

Os navios sonda (NS) surgiram a partir de adaptações em navios de carga geral, porém ao longo dos anos, foram projetados com objetivo específico. É um navio de perfuração, com autonomia de locomoção e grande velocidade quando comparado com outras estruturas. Embora críticos deste tipo de unidade considerem sua instabilidade seu principal ponto negativo, a evolução tecnológica vem conseguindo diminuir essa limitação. Não possui restrição quanto ao limite de lâmina d'água operável.

Possui casco de navio normal, porém com uma abertura central, através da qual é executada a atividade de perfuração. Apresentam grande capacidade de estocagem, desenvolvem uma alta velocidade quando em movimento e são autopropelidos, no entanto, são sensíveis às condições ambientais, restringindo sua atuação em mares agitados em função do tempo em que ficam fora de operação (*down time*) (THOMAS, 2004).



Figura 9 - Navio Sonda
Fonte: Rigzone

As sondas semissubmersíveis surgiram a partir de 1961 e são embarcações apoiadas sobre flutuadores, lastreados quando em operação, para garantir a estabilidade e a deixando menos vulneráveis a ação dinâmica das ondas. Tais sondas podem ter autopropulsão do tipo diesel-elétrico ou não. Possuem capacidade de atuar em condições severas de mar e a grandes profundidades.

As sondas semissubmersíveis são o segmento de sondas que apresenta maior crescimento. Esta estrutura flutuante possui colunas que, quando enchidas com água do mar, submergem a unidade até uma determinada profundidade. Apesar de movida pela atividade do mar, possui grande parte de sua estrutura imersa, o que, combinado com oito grandes âncoras, fazem das semissubmersíveis instalações bastante estáveis. THOMAS (2004).



Figura 10 - Sonda semissubmersível
Fonte: Rigzone

Em caso de lâminas d'água maiores empregam-se as chamadas unidades flutuantes, semissubmersíveis ou navios-sonda, cujos equipamentos de controle de poço situam-se no fundo do mar, e que são mantidas no local desejado por um sistema de ancoragem ou de posicionamento dinâmico.

No sistema de ancoragem utiliza-se de 8 a 12 linhas, com cabos e/ou correntes - com diâmetro variável em virtude da profundidade e esforços atuantes – conectados às ancoras.

No posicionamento dinâmico (*Dynamic Position, DP*), ao invés de ancoras, é utilizado um conjunto de impulsores laterais e hélices de propulsão. Um sistema de computadores é encarregado dos cálculos para correção do posicionamento da plataforma, deslocando-a para compensar as forças ambientais (correnteza, vento) e a mantendo dentro de um raio máximo aceitável delimitado em relação ao poço, visando evitar movimentações excessivas da coluna de perfuração. Assim facilita o projeto dos *risers*, tubulação que liga a plataforma ao fundo do mar (poço ou linhas de produção) (THOMAS, 2004).

Unidades com DP são mais recentes (a primeira foi o navio SEDCO 445 concluído em 1971), mais caras e mais sofisticadas. Seu uso é justificado em condições adversas de trabalho (possível necessidade de desconexão da sonda ao poço e rápido abandono da locação), em águas muito profundas ou áreas nas quais o fundo do mar esteja bastante congestionado, por dutos, poços e linhas em geral (D'ALMEIDA, 2000).

Ao construir uma sonda de perfuração, alguns pontos sobre equipamentos devem ser considerados, conforme o ambiente em que esta sonda irá operar:

- Profundidade da lâmina d'água e profundidade da perfuração;
- Pressão interna, tamanho de *risers*, especificação de BOP;
- Espaço na plataforma e capacidade de carregamento variável; e,
- Existência de posicionamento dinâmico.

A lâmina d'água é o principal fator limitante dependendo do ambiente em que a sonda irá realizar suas operações, em suma, transfigura-se na capacidade máxima de operação de uma estrutura em função da profundidade do mar na região em questão.

Basicamente se refere à distância entre o fundo do oceano e a mesa rotativa da plataforma. Em uma profundidade de cerca de 200 metros observa-se o domínio predominante do uso de *Jack-ups*. Em uma lâmina d'água intermediária as semissubmersíveis são as estruturas mais comuns.

O foco de atenção no mercado de exploração de petróleo e gás, na contemporaneidade, são as águas profundas. Estão sendo feitos grandes investimentos nas duas principais estruturas capazes de explorar nestes ambientes, os navios-sondas e as semissubmersíveis.

A capacidade de carregamento é capacidade total de carga da sonda, englobando os revestimentos, a coluna de perfuração, os *risers* e as tubulações. Sondas com baixa capacidade possuem problemas na armazenagem e no transporte de equipamentos utilizados na atividade de perfuração, chegando em alguns casos, ter a necessidade de se contratar barcos de apoio para auxiliar no processo. Uma maior capacidade de carregamento permite à sonda ganhos de eficiência devido à possibilidade de realizar operações de forma simultânea.

Tradicionalmente, a ancoragem restringiu as embarcações a lâminas d'água de até 5000 pés, aproximadamente 1.500 metros, de profundidade, o que, certamente, desestimula a seleção de tal tecnologia em relação ao sistema de posicionamento dinâmico para sua utilização em projetos de novas sondas. No entanto, recentes empreendimentos da Petrobrás, como o teste piloto em Tupi, utilizam embarcações com ancoragem em lâminas d'água consideravelmente maiores que as citadas.

O que é evidenciado atualmente, provocando todo o debate acerca do assunto, é que existem importantes variações em termos de custos operacionais e de capital que diferenciam os dois sistemas.

Em um contexto econômico, observa-se que a ancoragem possui vantagens comparativas em projetos exploratórios mais longos e em lâminas d'água de até 5500 pés, cerca de 1700 metros, isto incluindo a fase de desenvolvimento. Em contrapartida, o posicionamento dinâmico estaria mais apto a explorar em lâminas mais profundas, uma vez que seria quase impossível, perfurar em águas ultraprofundas sem esse sistema.

3.2.1. PRINCIPAIS COMPONENTES DE UMA SONDA DE PERFURAÇÃO MARINHA

Uma sonda de perfuração pode ser dividida em duas partes, o casco e os equipamentos presentes na parte superior da sonda, sendo conhecido na indústria como *topside*.

3.2.1.1. Casco

O casco é o corpo do navio sem mastro ou qualquer outro aparelhamento acessório ou arranjo, sua forma não é sempre a mesma, sendo definida levando em consideração a resistência mínima a propulsão, a mobilidade desejada e a estabilidade da plataforma.



Figura 11 - Casco de um navio
Fonte: Arte naval

Em medidas métricas o casco é definido por seu comprimento, largura e altura. Seus principais direcionadores de tamanho são a capacidade de armazenamento de óleo, o espaço no deck, a estabilidade no mar e em ambientes hostis, a capacidade máxima de

carga suportada para os equipamentos, a capacidade de variação de carga e o tipo de sonda.

3.2.1.2. Equipamentos do topo da sonda

Os equipamentos de topo da sonda marítima são em certo ponto semelhantes aos encontrados em uma sonda terrestre, mas adaptados ao ambiente marinho.

A Figura 12 apresenta um esquema de uma sonda de perfuração.

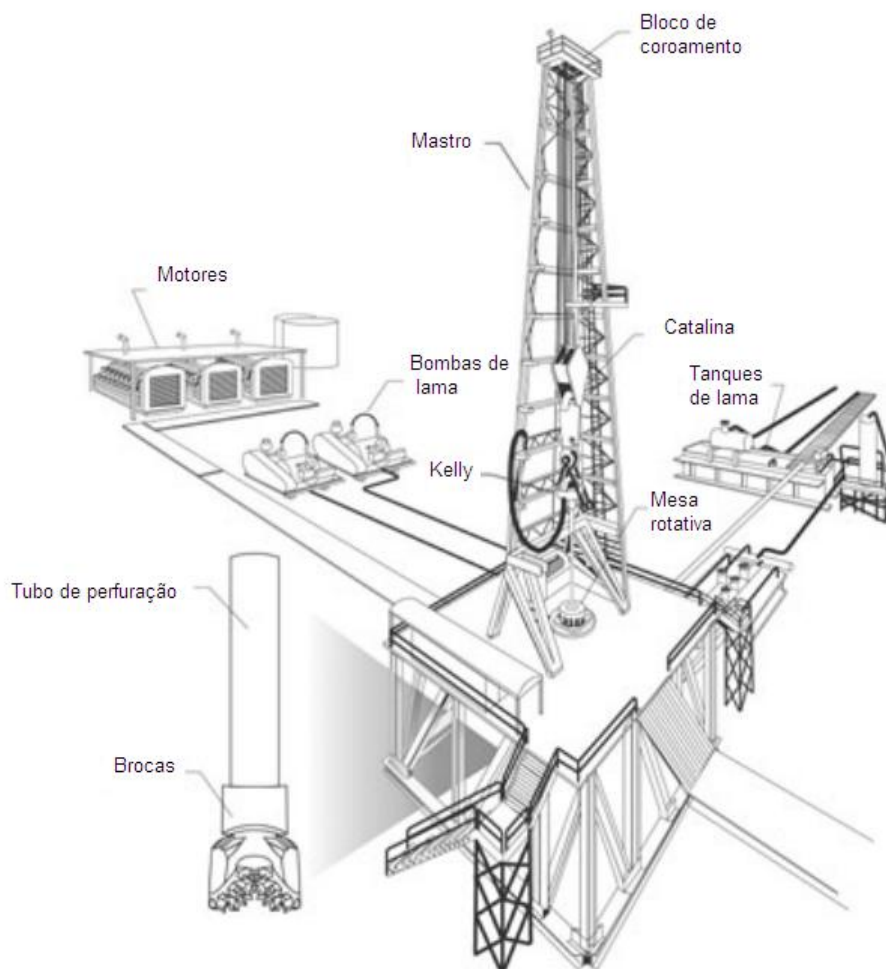


Figura 12 - Esquema de uma sonda de perfuração
Fonte: HALLIBURTON (2007)

Os equipamentos de topo podem ser separados em 7 sistemas assim denominados: O sistema de sustentação de cargas, o sistema de movimentação de cargas, o sistema de rotação, o sistema de circulação, o sistema de segurança de poço, o sistema de monitoração e o sistema de geração e transmissão de energia.

O sistema de sustentação de cargas, cujo principal equipamento é a torre ou o mastro, sustenta o peso da coluna de perfuração além de ser usado na manobra de substituição de brocas. É composto também pela subestrutura local e pelo estaleiro. A subestrutura local sustenta o mastro e o local onde é realizada boa parte do trabalho da plataforma. O estaleiro é o local onde ficam os tubos usados e a serem usados durante a perfuração.



Figura 13 - Mastro de uma sonda marítima
Fonte: ANP

O sistema de movimentação de cargas, utilizado para a movimentação de colunas de perfuração, colunas de revestimento, entre outros equipamentos. É formado pelo conjunto de cabos de perfuração, guincho, bloco de coroamento, catalina, gancho e *swivel*.

A Figura 14 a seguir apresentada, ilustra o sistema de movimentação de cargas.

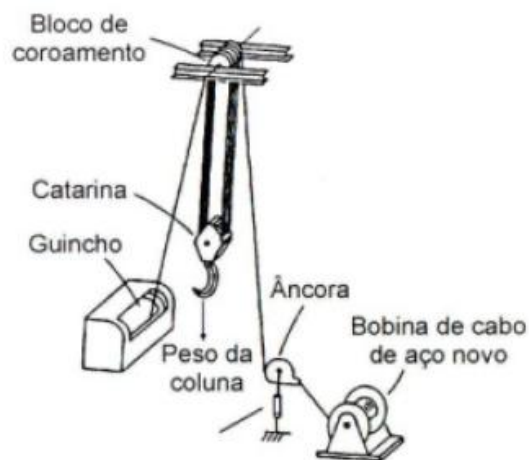


Figura 14 - Sistema de movimentação de cargas
Fonte: THOMAS(2004)

O sistema de rotação em sondas convencionais é composto pela mesa de rotação, sendo essa responsável pela ação da broca na rocha, o *kelly*, tubo de forma quadrangular ou hexagonal, que recebe a tração da mesa rotativa e a transfere para os tubos de perfuração, e pelo *swivel*, responsável pela ligação entre o *kelly* e a catalina.

Existe outro equipamento chamado *top drive*, que substitui o *swivel*, o *kelly* e a mesa rotativa, fazendo a conexão entre a parte fixa e a rotativa, dando torque à coluna e permitindo a entrada da mangueira de injeção de fluido de perfuração. Com o *top drive* ganha-se o comprimento do *kelly*, o que permite que três seções de canos de perfuração (*drill pipes*) sejam encaixadas na coluna por vez, otimizando o processo de perfuração, uma vez que menos paralisações são necessárias. O *top drive* é um equipamento de grande relevância no preço da sonda.

A Figura 15 apresenta um esquema de um *top drive*.

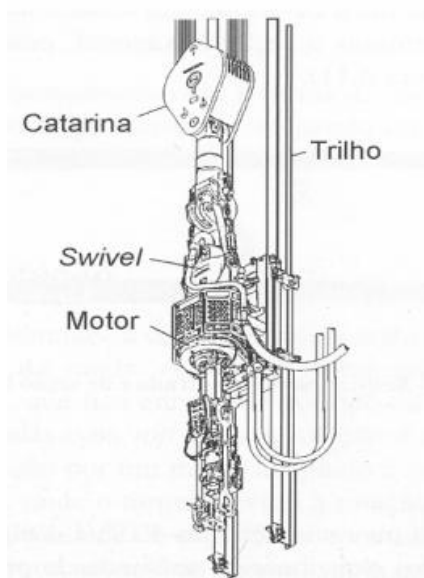


Figura 15 - *Top drive*
Fonte: THOMAS (2004)

O sistema de circulação é composto pelos tanques de lama, pelas bombas de lama, pelo tubo bengala, pelo *swivel* e pelo subsistema de tratamento. É responsável pelo bombeamento do fluido de perfuração, também conhecido como lama, pela coluna de perfuração e pela broca, que além de auxiliar na atividade de perfuração, ainda trás consigo, os resíduos gerados no processo, sendo tratado após sua utilização para que possa ser reutilizado.

A figura 16 apresenta um esquema composto pelo sistema de rotação e pelo sistema de circulação.

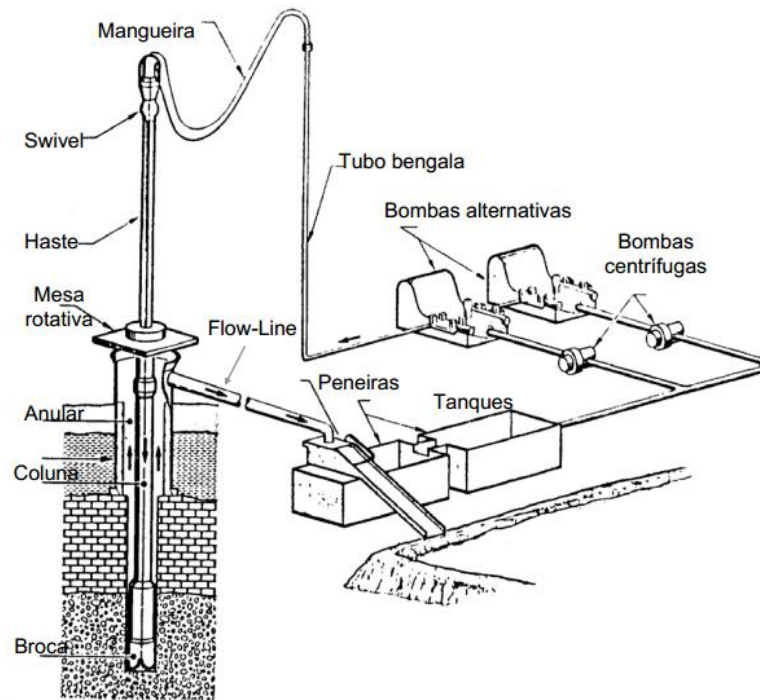


Figura 16 - Esquema do sistema de rotação e sistema de circulação
 Fonte: Thomas (2004)

O sistema de segurança de poço é composto por dois grupos principais, os equipamentos de segurança de cabeça de poço e equipamentos complementares e tem como objetivo prevenir que eventuais erupções cheguem ao topo.

Dentre os equipamentos complementares se destaca o *blow out preventer* (BOP), que é usado para evitar o vazamento descontrolado do poço, sendo composto por um conjunto de válvulas que são acionadas quando em situação de risco.

Sua estrutura pode ser observada na Figura 17 e na Figura 18.

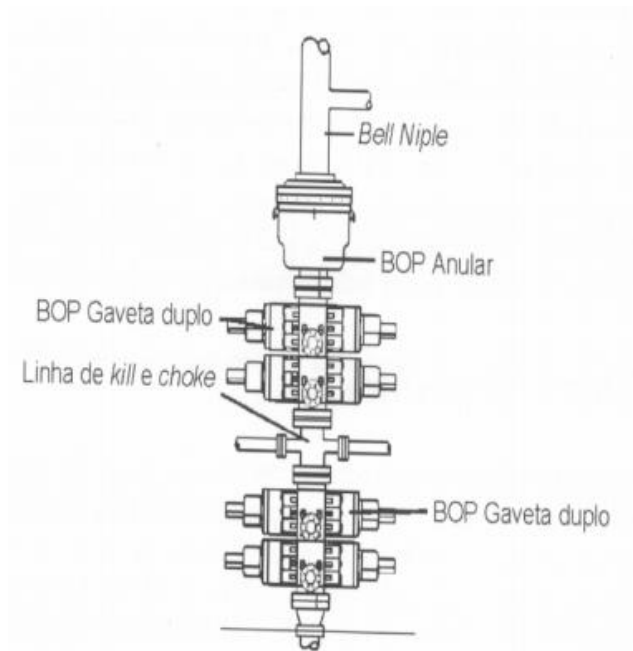


Figura 17 - Esquema gráfico de um BOP
 Fonte: <http://petroleoinfonet.blogspot.com.br>



Figura 18 – Fotografia de um BOP
 Fonte: <http://dailyhurricane.com>

O sistema de monitoração é constituído pelos equipamentos de controle da atividade, como manômetros, balanças, tacômetros, dentre outros. Tem por função acompanhar as atividades que são desenvolvidas durante a perfuração de poços.

O sistema de geração e transmissão de energia é constituído pelos geradores de energia e pela respectiva rede de transmissão. Quase sempre são utilizados motores a diesel, no entanto, ocorre por vezes, a utilização de motores a gás natural, quando este está presente no reservatório explorado.

Além dos sistemas descritos, alguns outros equipamentos possuem grande destaque na estimativa do custo de uma sonda. Como já citado acima, a presença de um equipamento de posicionamento dinâmico no lugar de equipamentos de ancoragem costuma encarecer o projeto, pois se trata de um equipamento alto valor, que permite a exploração em águas mais profundas.

A Figura 19 apresenta um esquema de funcionamento de um sistema de posicionamento dinâmico.

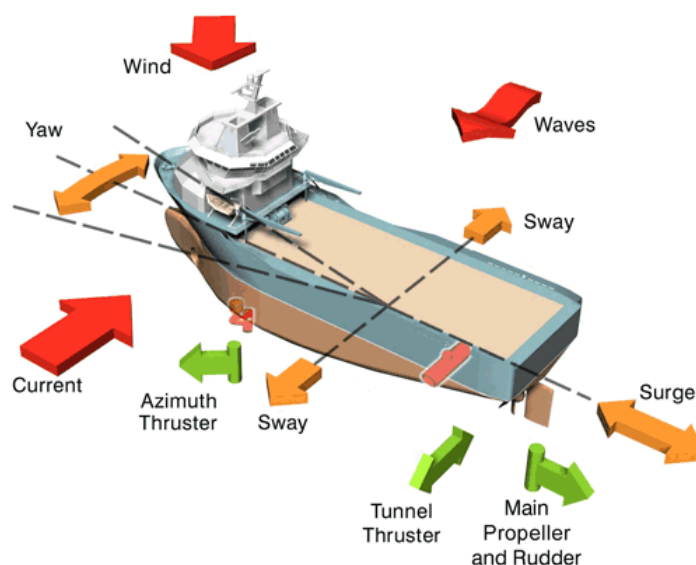


Figura 19 - Funcionamento do posicionamento dinâmico
Fonte: www.marineinsight.com

A capacidade de teste extensivo de poço (*extended well testing*, EWT) também é uma característica relevante, pois avalia a capacidade e as características do reservatório, dessa forma reduzindo os riscos envolvidos.

A presença de duas torres de perfuração, conhecida como capacidade de dupla atividade (*dual activity*) é característica que merece destaque, pois é a capacidade de realizar ao mesmo tempo operações de perfuração, assim como, fazer conexões de tubos acelerando todo o processo, e, por conseguinte, reduzindo os custos.

Pode-se observar na Figura 20 uma foto aérea de um navio sonda com dupla atividade.



Figura 20 - Navio sonda de dupla atividade
Fonte: Oil and Gas Journal (1999)

4 METODOLOGIA

O processo de estimativa de custo, conforme descrito no capítulo 2, envolve três fases, ou seja, a definição do problema, a escolha da metodologia e a execução. Essas fases, no entanto, para esse estudo podem ser desdobradas em 6 etapas:

- Definição do problema, os objetivos e a abordagem adotada;
- Desenvolvimento de premissas e regras básicas;
- Seleção da metodologia que será utilizada;
- Desenvolvimento do modelo a ser utilizado;
- Coleta e tratamento dos dados;
- Realização da estimativa.

A definição do problema e sua delimitação foram feitas logo no início da pesquisa e já foram descritas no capítulo 1, tendo por objetivo o desenvolvimento de um modelo paramétrico, usando análise multivariada para fazer uma estimativa dos custos de uma sonda de perfuração marítima.

As premissas assumidas e regras básicas têm por propósito delimitar o escopo, o contexto e o ambiente em que a pesquisa foi feita. Elas foram:

- Limitar a amostra a navios sondas e a sondas semissubmersíveis;
- Não houve restrição quanto ao local de construção, tampouco empresas contratantes ou operadoras;
- A amostra se limitou a todas as sondas construídas entre os anos 2000 e 2010;
- A moeda corrente foi o Dólar Norte Americano (US\$);
- Foram desconsiderados os custos operacionais das sondas depois de construídas;
- Para o tratamento da inflação no custo, considerou-se a composição de 30% para o casco e 70% para os equipamentos da sonda;
- O custo de construção é a única variável dependente tratada como métrica em todas as situações; e,

- O nível de significância aplicado em todos os testes é de 5%.

A metodologia de estimação paramétrica também foi definida no capítulo 1. Os modelos de análise foram desenvolvidos usando a relação entre variáveis e através de método de regressão múltipla e análise fatorial não sendo, portanto usado nenhum modelo já existente. Outros detalhes desses métodos podem ser encontrados no Capítulo 2.

O resultado do modelo se deu comparando os resultados e resíduos de uma regressão univariada com uma regressão multivariada e com outra regressão multivariada após o uso da análise fatorial.

Os dados foram coletados no banco de dados do *Market Survey System* administrado pelo IHS Petrodata e constituiu uma amostra inicial de 103 observações para cada uma das variáveis. Os dados, inicialmente coletados, contaram com 80 características técnicas para cada navio sonda e sonda semissubmersível. Com esse grande número de características tornou-se interessante a utilização de técnicas de redução de variáveis como o método dos componentes principais, que busca manter o máximo de informação condensada em um menor número de variáveis.

As características apuradas foram divididas em dois grupos, as características qualitativas, como nome da sonda e fabricante, local onde foi construída a sonda ou se possui algum equipamento específico, e, as quantitativas, como comprimento e largura do casco e capacidade máxima de carregamento. O detalhamento das características é apresentado no apêndice.

4.1. ESTUDOS PRELIMINARES

Os estudos preliminares compreenderam os tratamentos feitos antes de serem realizadas as análises multivariadas. Eles foram divididos em dois tipos, os referentes à análise das variáveis de forma individualizadas e os que envolvem a relação entre duas variáveis.

Esses estudos preliminares tiveram por objetivo melhorar a base de dados, dando o tratamento necessário quando ocorreram problemas e violações das premissas da regressão. Observa-se que é importante haver um prévio contato com os dados por parte do pesquisador para que possa ter maior familiaridade no momento da modelagem.

4.1.1. ANÁLISE DE DADOS COM APENAS UMA VARIÁVEL

O objetivo da análise de dados com apenas uma variável foi fazer o tratamento inicial de modo a deixar a amostra preparada para os estudos seguintes, como também, buscar por valores extremos e entender a estrutura dos dados, para que fosse possível identificar certas características como:

- Se os dados são homogêneos o bastante ou se é necessário separá-los em grupos distintos de análise;
- Existência de valores extremos; e,
- Encontrar curvas de distribuições padrão para serem usadas simplificando assim os estudos mais avançados.

Foram retiradas as variáveis que não acrescentavam valores significativos ao modelo. Isso se deu por possuírem valores constantes em todas as observações dadas ou porque apresentaram informação em menos de 60% dos casos.

Em seguida retirou-se da amostra os dados que não apresentavam informações completas, dessa forma foram retiradas da amostra as sondas cujas informações correspondiam a menos de 60% das características, ou seja, 48 ou menos características.

Após esse primeiro tratamento foram realizados estudos das variáveis de forma individual. O custo que é a variável depende foi priorizada, no entanto, algumas das variáveis independentes também foram analisadas.

O custo é uma variável monetária sendo, portanto, a soma de todos os custos individuais de cada elemento e serviço que foram usados na construção da sonda marítima.

Este trabalho se refere a uma estimativa paramétrica, que se utiliza de alguns direcionadores de custos e modelos matemático-estatísticos, que fazem com que o custo seja uma função daqueles direcionadores.

Os dados coletados se referem às sondas construídas entre o ano 2000 e 2010, sofrendo assim influência da inflação com o passar do tempo, seja para os equipamentos, seja para o casco. Por apresentarem características muito distintas, o casco e os equipamentos apresentam uma dinâmica diferente de normalização dos seus valores.

Assim foram usados índices distintos para ajustar o custo da sonda a uma mesma data base. Para o casco usou-se o índice relativo a navios de grande porte e para os equipamentos do topo da sonda utilizou-se o índice de máquinas usadas no setor de petróleo e gás fornecidos pelo *Bureau of Labour Statistics* (BLS).

O primeiro objetivo foi trazer todos os valores a uma mesma data e assim poder compará-los de forma correta. Usando dados históricos o custo foi segmentado, sendo 30% referente ao casco e 70% referentes aos equipamentos, essa proporção foi encontrada pelos dados do banco da ODS-Petrodata.

Os dados usados nos índices de deflação constam na Tabela 9.

Tabela 9 - Índices de inflação usados na deflação dos custos

Ano	Casco semissubmersível	Casco navio sonda	Equipamentos
2002	93	89	90
2003	100	100	100
2004	131	130	120
2005	146	143	135
2006	180	175	168
2007	223	213	202
2008	252	230	219
2009	186	160	179
2010	211	186	199
2011	241	212	221
2012	244	214	228
2013	250	218	233
2014	253	220	239
2015	257	223	244

Fonte: o autor

A data base escolhida foi janeiro do ano de 2003, dessa forma o seu índice é igual a 100% ou 1. Os valores dos custos encontrados para outras períodos foram todos trazidos aos valores referentes ao ano de 2003.

Em relação ao casco foi usado além do tipo de sonda, pois muda o índice para o casco, o ano de sua construção. Exemplificando, se o casco de um navio sonda foi construído em 2007 deve-se deflacionar o valor de 30% do custo da sonda a um índice de 213.

Para os equipamentos utilizou-se a média dos índices entre a data de contratação e de entrega da sonda, sendo essa média usada como índice de inflação, que será usado

no denominador para deflacionar o custo. Exemplificando, se uma sonda fosse contratada em 2005 e entregue em 2009 o valor do índice seria 180, que é a média entre os valores do índice para cada um dos anos entre a contratação e a entrega.

Assim a equação para o cálculo do valor corrigido é:

$$\frac{C_c}{C_r} \times \frac{I_{dc}}{I_{dts}}$$

Onde:

C_c = Custo corrigido;

C_r = Custo na data em que ocorreu;

I_{dc} = Índice de deflação do Casco; e,

I_{dts} = Índice de deflação dos equipamentos do topo.

Na Tabela 9 é possível observar que a inflação apresentou grande volatilidade durante os anos estudados. Houve um aumento no valor dos 3 índices em todos os anos, exceto para o ano de 2009 onde houve uma redução. Isso pode ser explicado pela crise econômica mundial, que fez com que o preço do petróleo caísse muito mediante a expectativa de uma possível recessão, o que reduziria o crescimento mundial, assim como, a expectativa de crescimento do consumo, da necessidade de combustíveis e matérias primas e, por conseguinte de exploração de petróleo. Após esse ano, a crise perdeu força e os preços voltaram a crescer, porém não na mesma intensidade de antes, onde o mercado estava superaquecido, demorando cerca de cinco anos para retornar ao mesmo patamar de preços de 2008. Foram comparados os valores antes e depois da correção em relação as suas estatísticas descritivas.

Em seguida a busca por valores anormais foi realizada junto com a verificação de normalidade. Foram feitos histogramas com curvas normais sobrepostas, assim, foi possível analisar a presença de dados anômalos, como também a existência de normalidade.

Para um estudo mais preciso na busca por valores anômalos, o gráfico Boxplot foi o escolhido. O Boxplot é um gráfico em forma de caixa construído com base em 5 números, o valor mínimo, o primeiro quartil (Q1), a mediana, o terceiro quartil (Q3) e o valor máximo. Para o teste de valores anômalo, calcula-se a medida interquartil (AIQ), que é a diferença entre Q3 e Q1 e se utiliza como limite inferior $Q1 - 1,5 \times AIQ$ e como

limite superior $Q3 + 1,5 \times AIQ$. Os valores que ultrapassarem esses limites são candidatos a valores anômalos.

Os valores atípicos foram retirados da amostra da variável dependente para os estudos seguintes.

As variáveis independentes quantitativas também foram exploradas, no entanto, seus dados não foram tratados, apenas estudados. Estes resultados se encontram no Apêndice.

As variáveis qualitativas relevantes foram transformadas em quantitativas por meio de variáveis dicotômicas (*dummy*), representando a existência ou ausência de uma característica a elas referenciada. Para essas variáveis foram feitos gráficos de pizza para entender o comportamento da amostra, dada à característica estudada. Estes gráficos se encontram no Apêndice.

4.1.2. ANÁLISE DE DADOS COM DUAS VARIÁVEIS

Na análise com duas variáveis, estudou-se a relação entre a variável dependente e cada uma das variáveis independentes. Nesse estágio buscou-se averiguar a existência de uma relação linear dos elementos, assim como dos coeficientes, podendo também ser chamada de bilinearidade.

A equação 5 apresenta a relação de linearidade entre as variáveis:

Onde:

y = variável dependente;

b_0 = coeficiente linear; e,

b_1 = coeficiente angular.

x = variável independente.

Para desenvolver a relação bilinear foi preciso estudar a distribuição das variáveis independentes em relação a variável dependente. Essa análise da distribuição segue os seguintes passos:

- Buscar por valores anômalos;

- Ter uma visualização dos dados; e,
- Quantificar a relação percebida entre as variáveis.

Para as variáveis independentes quantitativas, em primeiro lugar, foram feitos os estudos de correlação em relação a variável dependente, dessa forma tentando encontrar os direcionadores de custos além de tentar compor relações para as estimativas de custos (CERs). A correlação entre as variáveis independentes também foram calculadas com o intuito de encontrar as possíveis causas de multicolinearidade.

Gráficos de dispersão foram feitos com o objetivo de encontrar algum tipo de relação entre as variáveis independentes qualitativas e a variável custo, buscando analisar a linearidade entre elas, assim como a existência de valores anômalos.

Foi feita a ANOVA com o intuito de averiguar se a média da variável custo das observações com determinada característica são significativamente diferentes das médias encontradas sem essa característica. Também foi feito o teste de Levine para verificar a homocedasticidade entre os diferentes grupos formados.

Por último foram estudadas as relações entre as variáveis independentes quantitativas e qualitativas. Os resultados se encontram no Capítulo 5.

4.2. ANÁLISES MULTIVARIADAS PARA A ESTIMATIVA DE CUSTOS

São os modelos onde existe a maior concentração de informações, sendo dessa forma os mais completos. Antecedendo a análise multivariada, 4 passos devem ser observados:

- Busca por valores anômalos, que em geral é feito por algoritmos, já que tabelas e gráficos se tornam pouco úteis quando se está analisando várias variáveis independentes;
- Busca por colinearidade entre as variáveis independentes;
- Visualização dos dados para se entender sua estrutura e responder a questões como, a homogeneidade dos dados, a confirmação da existência de valores anômalos, como os dados estão relacionados e se existe alguma sugestão de relação; e,
- Quantificação das relações observadas.

As análises multivariadas foram realizadas em três etapas:

- Regressão simples;
- Regressão com múltiplas variáveis; e,
- Regressão com fatores ou constructos.

O primeiro passo foi a realização de uma regressão simples entre a característica que apresentou a maior correlação linear e a variável custo. O objetivo desse passo é ter um valor base de comparação para analisar a melhora dos resultados com o desenvolvimento de novos modelos mais complexos. Os resultados se encontram no capítulo 5 e os diagnósticos dessa regressão no Apêndice.

No segundo passo foram feitas regressões múltiplas, utilizando o método *stepwise*. Os resultados foram comparados e foi definido o melhor modelo de regressão múltipla. O critério de escolha do melhor modelo é o maior R^2 , e ter todas as premissas da regressão atendidas.

A terceira etapa foi dividida em duas fases. Foi realizada a análise dos componentes principais para se estimar os indicadores e, em seguida, foi gerada uma regressão linear entre esses indicadores e a variável dependente. Os resultados encontrados foram comparados com os anteriormente obtidos. Maiores detalhes dos procedimentos e os resultados encontrados estão apresentados no Capítulo 5.

Os passos 2 e 3 apresentaram os diagnósticos da regressão, assim como, a validação dos resultados. Na validação foram utilizados 10 dados que foram propositalmente excluídos nos estudos preliminares, dessa forma, a verificação dos resultados encontrados, assim como, a comparação entre a capacidade de previsão dos modelos foram observadas. Os resultados encontrados foram comparados e o melhor modelo foi escolhido.

Todos os estudos de regressão, análise multivariada de análise fatorial foram realizadas no programa SPSS 18.

5 RESULTADOS

5.1. ESTUDOS PRELIMINARES

Antes mesmo de começar os estudos estatísticos, foram retiradas da amostra as variáveis que não contribuíam estatisticamente ou porque não apresentaram casos suficientes ou porque o valor era o mesmo em todas as observações, possuindo assim pouca significância para o modelo.

Em relação às observações, dos 103 casos 2 foram retirados da amostra por não apresentarem dados suficientes para a análise, pois o número de variáveis independentes era muito pequeno, podendo comprometer o estudo. Mais 10 observações foram separadas para compor um grupo usado na validação dos resultados, restando 91 valores amostrais.

Após esses estudos, o trabalho se limitou a tratar de 22 variáveis independentes e realizar estudos sobre elas. Como definido no Capítulo 4, elas podem ser divididas em dois grupos, qualitativas e quantitativas, sendo 12 no primeiro grupo e 10 no segundo.

A Tabela 10 apresenta esses dados de forma organizada.

Tabela 10 - Lista de variáveis independentes selecionadas

Variáveis Qualitativas	Variáveis Quantitativas
Tipo de Sonda	Profundidade da lâmina d'água (pés)
Se foi construída de forma especulativa	Profundidade de perfuração (pés)
Se apresenta posicionamento dinâmico	Carga máxima suportada pela sonda (ton)
Se é capaz de realizar dupla atividade	Capacidade de carga variável da sonda (ton)
Se está adaptado à ambiente hostil	Comprimento do casco (pés)
Se possui top drive	Largura do casco (pés)
Em que estágio da construção que a sonda se encontra	Capacidade do alojamento (pessoas)
Capacidade de teste extensivos de poços (EWT)	Capacidade de água de perfuração (BBL)
Se está adaptado ao mar do norte	Capacidade de água potável (BBL)
Se possui a capacidade de zero desperdício de fluidos	Carga máxima suportada pela torre (libras)
Se está habilitado para a atividade na Noruega	
Local de construção	

Fonte: o autor

5.1.1. ANÁLISE COM UMA VARIÁVEL

5.1.1.1. Análise da variável dependente - Custos

Serão apresentados os resultados referentes aos estudos individuais sobre a variável custo, que possui como unidade de medida o Dólar americano, comparando-se as estatísticas descritivas, assim como, o histograma da variável dependente antes e após ser tratada em relação à inflação.

A Tabela 11 apresenta as estatísticas descritivas para os dois casos.

Tabela 11 - Estatísticas descritivas dos custos bruto e corrigido

	Custo bruto (US\$)	Custo corrigido (US\$)
Média	\$621.000.000,00	\$317.000.000,00
Variância	22.000.000.000.000.000,00	4.770.000.000.000.000,00
Desvio-padrão	148.000.000,00	69.000.000,00
Assimetria	-0,238	0,018
Erro Padrão Assimetria	0,253	0,253
Relação Assimetria/erro	-1,06043	14,1028
Curtose	1,352	0,612
Erro Padrão Curtose	0,500	0,500
Relação Curtose/erro	0,370122	0,81692
Mediana	\$640.000.000,00	\$322.960.800,00
Moda	\$750.000.000,00	\$238.146.100,00
Mínimo	\$250.000.000,00	\$158.061.400,00
Máximo	\$1.150.000.000,00	\$555.995.300,00
Amplitude	\$900.000.000,00	\$397.933.900,00
Quantidade	91,00	91,00
1st Quartil	\$537.000.000,00	\$259.348.500,00
3rd Quartil	\$720.000.000,00	\$365.901.900,00
Amplitude Interquartil	\$183.000.000,00	\$106.553.400,00

Fonte: o autor

Pelos valores dispostos na tabela 11 é possível notar que o desvio-padrão diminuiu após a correção dos valores, isso leva a crer que exista uma influência da inflação que aumentava a dispersão dos dados. Essa afirmativa ganha força quando se comparam o valor máximo e a amplitude de cada amostra. O valor máximo reduziu de US\$ 1.150 milhões para cerca de US\$ 555 milhões, praticamente a metade, e a

amplitude de US\$ 900 milhões para próximo de US\$ 400 milhões, indicando que a amostra se tornou menos dispersa.

Além disso, os valores da assimetria e curtose se aproximaram do valor nulo, inclusive com a nova assimetria não sendo significativamente diferente de zero, logo houve uma melhora em relação à normalidade dos dados.

As Figuras 21 e 22 contêm os histogramas para o custo bruto e para o custo corrigido respectivamente.

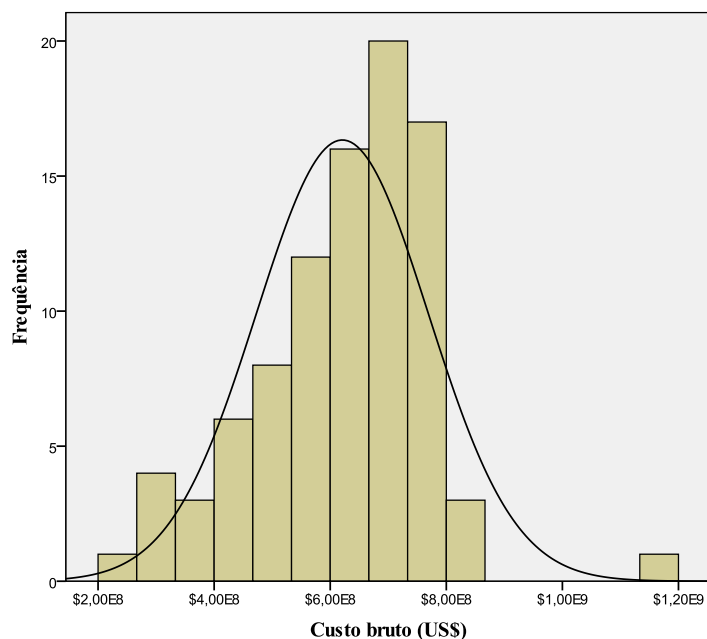


Figura 21 - Histograma de Custo bruto

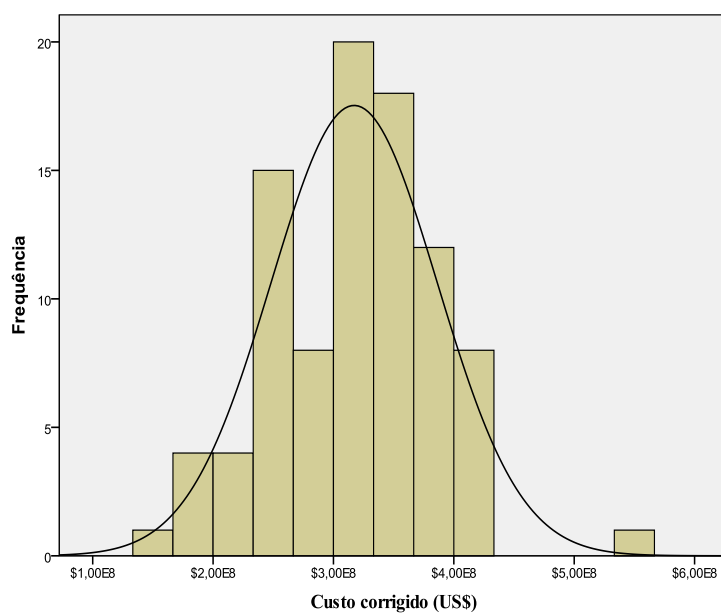


Figura 22 - Histograma de custo corrigido pela inflação

Os histogramas constantes nas Figuras 21 e 22 permitem observar que o valor corrigido apresenta uma assimetria mais central que o valor não corrigido, dessa forma a correção deixou os dados mais próximos de uma função normal, o que é benéfico para o modelo. Portanto, todos os estudos que seguem utilizarão apenas os dados sobre os custos já corrigidos pela inflação.

Ainda observando os referidos histogramas fica clara a presença de um valor anômalo em ambas as amostras, assim, essa anomalia foi estudada antes de se realizar as outras análises de normalidade.

Para essa análise será feito um Boxplot representado pela Figura 23.

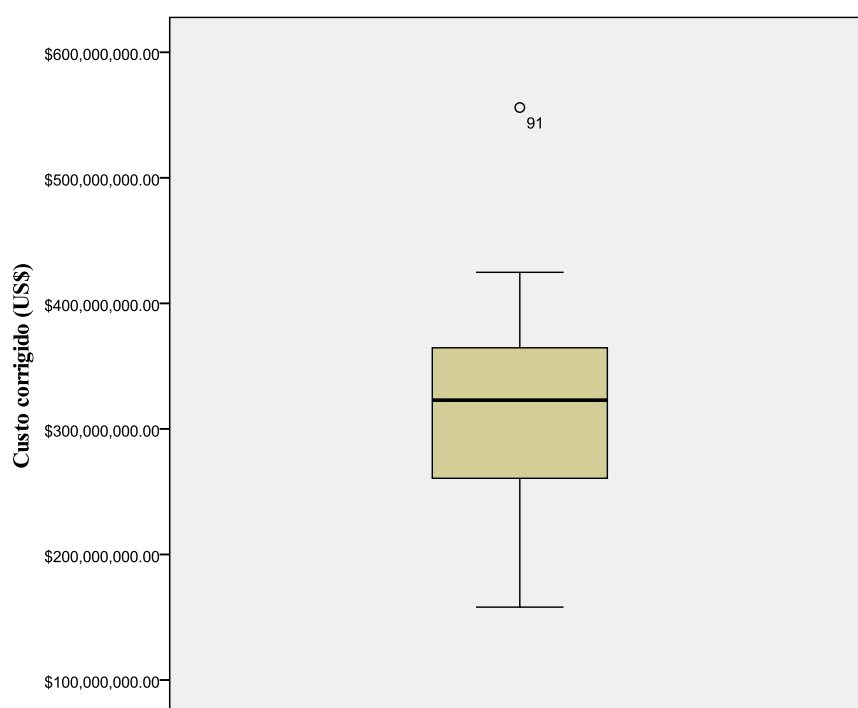


Figura 23 - Boxplot do custo corrigido

A Figura 23 evidencia a existência de um valor muito superior aos demais. Para entender melhor a origem dessa anomalia foi feito o estudo somente dessa observação. Todas as suas características estavam dentro do padrão dos outros, no entanto essa sonda tinha uma característica especial de ser de uma classe chamada *iceclass fullywinterized*, cuja propriedade é estar preparada para ambientes muito hostis em relação ao frio. Como essa característica foi encontrada de forma isolada, optou-se por retirar esse dado da amostra evitando assim, que esse traga algum viés negativo para a análise geral dos dados.

Após a retirada do dado anômalo, foi promovido um novo estudo das estatísticas descritivas e gerado um novo histograma.

Tabela 12 - Estatísticas descritivas dos custos corrigidos com e sem valores anômalos

	Custo corrigido	Custo corrigido sem valores anômalos
Média	\$317.000.000,00	\$314.550.123,33
Variância	4.770.000.000.000.000,00	4.172.745.508.523.830,00
Desvio-padrão	69.000.000,00	64.596.791,78
Assimetria	0,018	-0,432
Erro Padrão Assimetria	0,253	0,254
Relação Assimetria/erro	14,1028	-1,70163
Curtose	0,612	-0,527
Erro Padrão Curtose	0,500	0,503
Relação Curtose/erro	0,81692	-1,04739
Mediana	\$322.960.800,00	\$322.960.800,00
Moda	\$238.146.100,00	\$238.146.100,00
Mínimo	\$158.061.400,00	\$158.061.400,00
Máximo	\$555.995.300,00	\$424.753.200,00
Amplitude	\$397.933.900,00	\$266.691.800,00
Quantidade	91,00	90,00
1st Quartil	\$259.348.500,00	\$258.551.225,00
3rd Quartil	\$365.901.900,00	\$363.982.800,00
Amplitude Interquartil	\$106.553.400,00	\$105.431.575,00

Os dados da Tabela 12 permitem visualizar, que com a retirada do valor anômalo a média diminuiu cerca de US\$ 2,5 milhões e o desvio-padrão reduziu em mais de US\$ 4 milhões, isso mostra o impacto que um dado com um valor muito divergente dos demais pode ter sobre a amostra. No entanto, a nova amostra não ficou mais próxima de uma distribuição normal e apesar da curtose ter diminuído sua assimetria se afastou do valor central.

A Figura 24 apresenta o novo histograma da variável custo sem o valor anômalo.

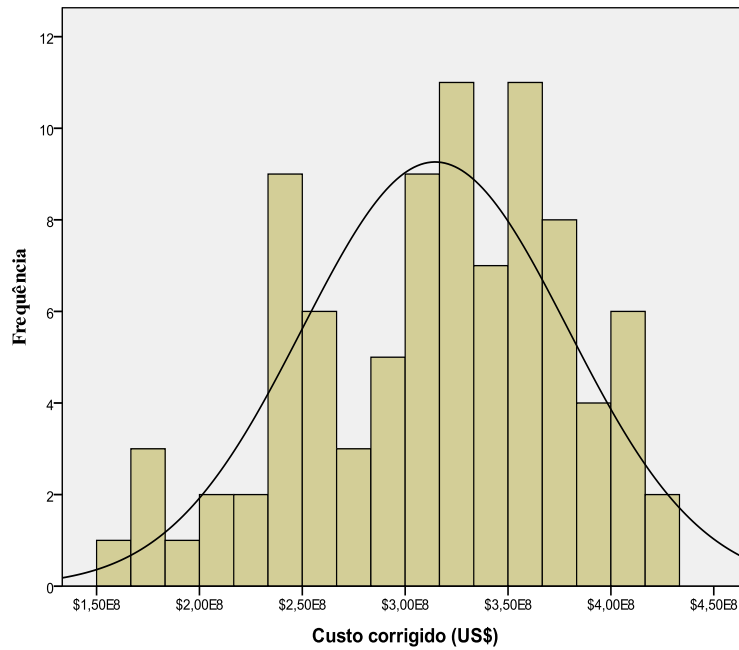


Figura 24 - Histograma do Custo corrigido e sem dados anômalos

No histograma apresentado na Figura 24 aparentemente a amostra não é normal e existe a presença de assimetria negativa. Aparentemente o problema dos dados extremos foi resolvido, não há nenhuma evidência no histograma que possa acusar a existência de um novo dado anômalo.

No intuito de corroborar essa afirmação, um novo gráfico Boxplot foi feito para a nova amostra, conforme representado na Figura 25.

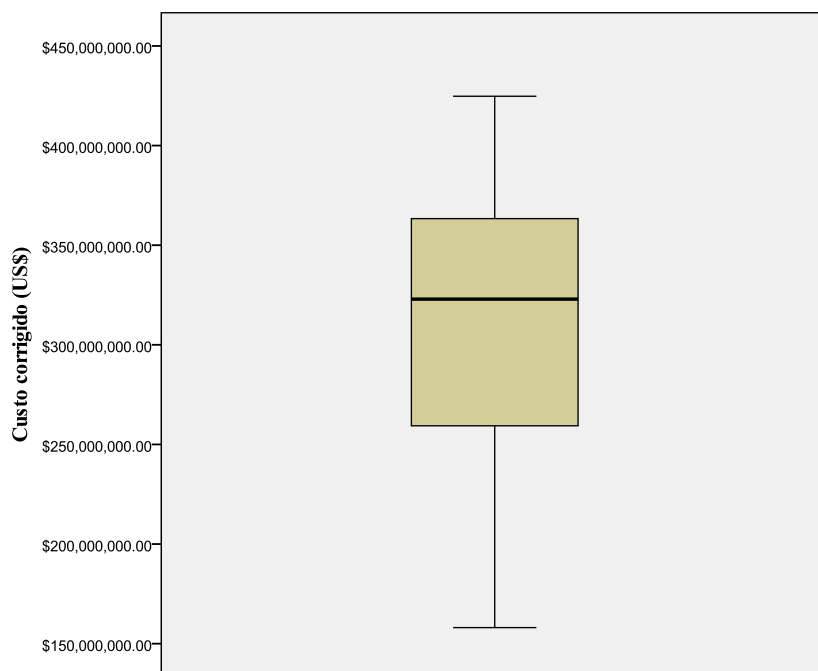


Figura 25 - Boxplot do Custo corrigido sem dados anômalos

O Boxplot confirmou que não havia mais dados anômalos na amostra, sendo assim essa passou a ser usada nos estudos seguintes.

Em relação à normalidade dos dados, foi feito o teste de Kolmogorov-Smirnov no qual se tem como hipótese nula que os dados seguem uma distribuição normal. Caso o valor de significância seja maior que 0,05 não se deve rejeitar a hipótese nula de que os dados seguem uma distribuição normal.

A Tabela 13 apresenta o resultado do teste de Kolmogorov-Smirnov para os custos corrigidos.

Tabela 13 - Teste de normalidade Kolmogorov-Smirnov para a variável dependente

Variável	Kolmogorov-Smirnov	Significância
Custo corrigido	0,088	0,085

Fonte: o autor

O nível de significância é de 0,085. Dessa forma, não foi possível rejeitar a hipótese nula e a amostra pode não ser considerada como não seguindo uma distribuição normal.

5.1.2. ANÁLISES COM DUAS VARIÁVEIS

Nessa seção serão feitas as análises sobre as relações entre a variável dependente e as independentes e também a relação entre as variáveis independentes entre si.

5.1.2.1. Variáveis quantitativas

Para as variáveis quantitativas, o primeiro passo foi o estudo da correlação, nesse caso, por serem variáveis métricas, apenas o método de Pearson foi utilizado.

A Tabela 14 apresenta os resultados encontrados.

Tabela 14 - Lista de correlação entre Custo e variáveis quantitativas

Variável	Custo corrigido	Significância
Profundidade da lâmina d'água (pés)	0,506	0,000
Profundidade de perfuração (pés)	0,449	0,000
Carga máxima suportada pela sonda (ton)	0,704	0,000
Capacidade de carga variável da sonda (ton)	0,672	0,000
Comprimento do casco (pés)	0,673	0,000
Largura do casco (pés)	-0,509	0,000
Capacidade do alojamento (pessoas)	0,635	0,000
Capacidade de água de perfuração (BBL)	0,274	0,009
Capacidade de água potável (BBL)	0,262	0,013
Carga máxima suportada pela torre (libras)	0,332	0,001

Fonte: o autor

Observando a Tabela 14 nota-se que todos os dados rejeitam a hipótese nula de ausência de correlação, logo todas as variáveis independentes qualitativas tem correlação significativa com a variável dependente. Assim, todas estariam aptas a entrar no modelo de regressão. Os melhores resultados foram encontrados para a carga máxima suportada pela sonda com o valor de 0,704, sendo então a variável escolhida para ser usada na regressão simples.

Para detectar um possível problema de multicolinearidade, assim como, a possibilidade de uma análise fatorial, foram estudadas as correlações entre as variáveis independentes.

Nas Tabelas 15 e 16 são apresentadas as correlações. Os dados foram divididos em duas tabelas devido sua quantidade, facilitando assim a sua visualização.

Tabela 15 - Correlação entre as variáveis independentes quantitativas (I)

Variável	Profundidade de perfuração	Carga máxima da sonda	Variação de carga suportável	Comprimento do casco
Profundidade da lâmina d'água (pés)	0,559	0,526	0,490	0,392
Profundidade de perfuração (pés)	1	0,472	0,491	0,390
Carga máxima suportada pela sonda (ton)		1	0,930	0,918
Capacidade de carga variável da sonda (ton)			1	0,946

Tabela 16 - Correlação entre as variáveis independentes quantitativas (II)

Variável	Largura do casco	Capacidade do alojamento	Água de perfuração	Água potável	Carga máxima suportada pela torre (ton)
Profundidade da lâmina d'água (pés)	-0,231	0,608	-0,002	0,168	0,621
Profundidade de perfuração (pés)	-0,299	0,599	0,176	0,174	0,546
Carga máxima suportada pela sonda (ton)	-0,833	0,704	0,405	0,262	0,222
Capacidade de carga variável da sonda (ton)	-0,869	0,748	0,429	0,381	0,262
Comprimento do casco (pés)	-0,907	0,710	0,346	0,404	0,216
Largura do casco (pés)	1	-0,538	-0,298	-0,397	-0,083
Capacidade do alojamento (pessoas)		1	0,236	0,346	0,484
Capacidade de água de Perfuração (BBL)			1	0,271	0,398
Capacidade de água potável (BBL)				1	0,270
Carga máxima suportada pela torre (ton)					1

Fonte: o autor

Analisando os dados das Tabelas 15 e 16, percebe-se que as correlações apresentam valores bem distintos, assim, deve-se ter cautela principalmente ao se utilizar as variáveis que possuam correlação acima de 0,7 no mesmo modelo.

Para estudar a linearidade, a homocedasticidade e a presença de valores anômalos, serão feitos os gráficos de dispersão entre o custo corrigido e as variáveis independentes quantitativas excluindo as variáveis capacidade de água de perfuração e capacidade de água potável por apresentarem correlação abaixo de 0,3 com a variável dependente.

Em relação à linearidade é analisado se os dados dispersos podem ser explicados por uma reta ou se outro tipo de relação é mais evidente. Para a homocedasticidade é analisado se os valores apresentam dispersão semelhante por todo o gráfico. Para se analisar a existência de valores anômalos deve-se envolver a maioria dos dados com

uma elipse. Caso existam dados que estejam muito afastados dos outros, então existe uma grande chance deles serem caracterizados como anômalos.

A primeira relação a ser estudada pelo gráfico de dispersão ocorre entre os custos corrigidos e a profundidade da lâmina d'água, e está representado na Figura 26.

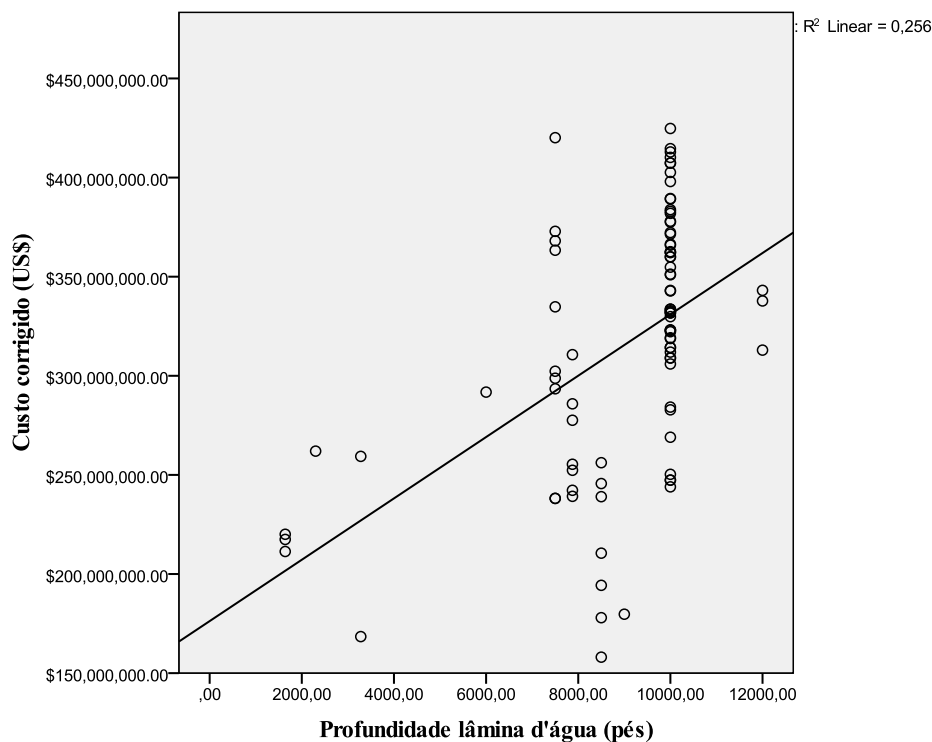


Figura 26 - Gráfico de dispersão entre os custos corrigidos e a profundidade da lâmina d'água

Pela Figura 26, nota-se que não existe homocedasticidade, pois a dispersão e concentração dos dados aumentam muito com o acréscimo do valor da lâmina de água. Uma relação linear positiva de primeira ordem pode ser percebida pela linha de tendência, porém, com R^2 apresentando de baixo valor 0,256. A presença de valores anômalos devido à diferença de desvio-padrão ao longo da amostra não é clara.

A segunda relação estudada é entre os custos corrigidos e a profundidade de perfuração.

A Figura 27 apresenta essa relação através de um gráfico de dispersão.

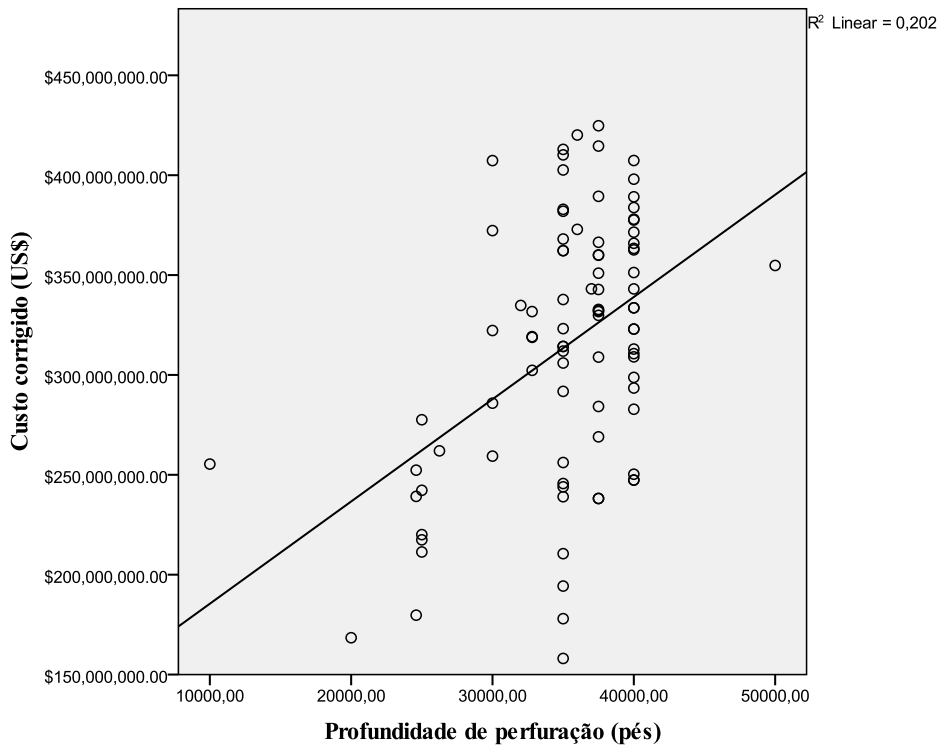


Figura 27 - Gráfico de dispersão entre os custos corrigidos e a profundidade perfuração

Pela Figura 27 é possível notar uma distribuição mais homogênea em relação à variância, no entanto, não fica clara a existência de homocedasticidade nesses dados. A presença de valores anômalos é clara em pelo menos dois dados e uma relação linear simples positiva pode ser observada apesar de seu baixo valor de R^2 igual a 0,202.

A terceira relação a ser estudada ocorre entre o custo corrigido e a carga máxima suportada pela sonda. A Figura 28 apresenta o gráfico de dispersão dessa relação.

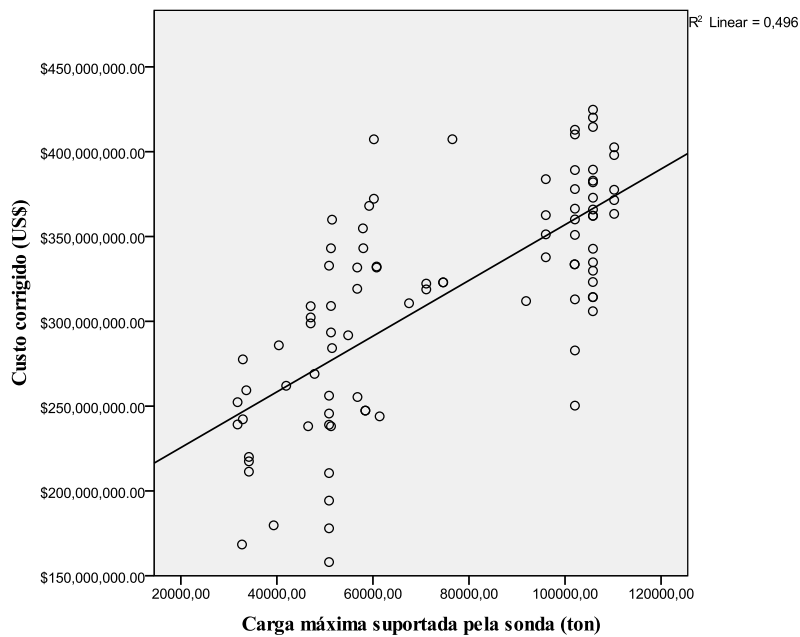


Figura 28 - Gráfico de dispersão entre os custos corrigidos e a carga máxima da sonda

Os dados apresentam boa distribuição em relação à variância, à não existência de valores anômalos e à linearidade, o valor encontrado para R^2 , foi o mais alto entre as variáveis estudadas R^2 igual a 0,496. No entanto, existem dois grupos claros de dados, os com valores de carga máxima suportada pela sonda próximos a 50.000 toneladas e os com a carga máxima próxima a 110.000 toneladas. Intuitivamente, o tipo de sonda deve ser considerado nesse caso para diferenciar esses valores em dois grupos distintos. Assim será feito um gráfico Boxplot para a visualização dessa análise, apresentado na Figura 29.

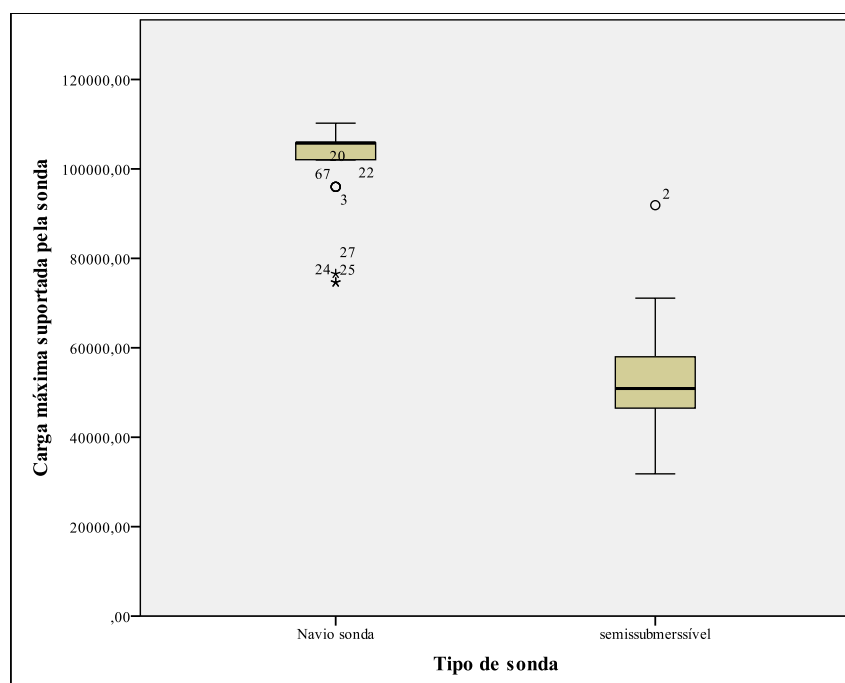


Figura 29 - Boxplot da carga máxima da sonda pelo tipo de sonda

Pela Figura 29 nota-se uma grande diferença entre os valores médios de carga máxima da sonda para os dois diferentes tipos de sondas. Isso deve ser considerado no momento de se fazer a regressão, avaliando a possibilidade de inclusão de um efeito moderador.

A Figura 30 apresenta novamente o gráfico de dispersão, no entanto, esse além de apresentar a relação entre o custo corrigido e a carga máxima suportada pela sonda, também considera a distinção em relação ao tipo de sonda. Duas linhas de tendência estão presentes, uma para cada grupo de dados.

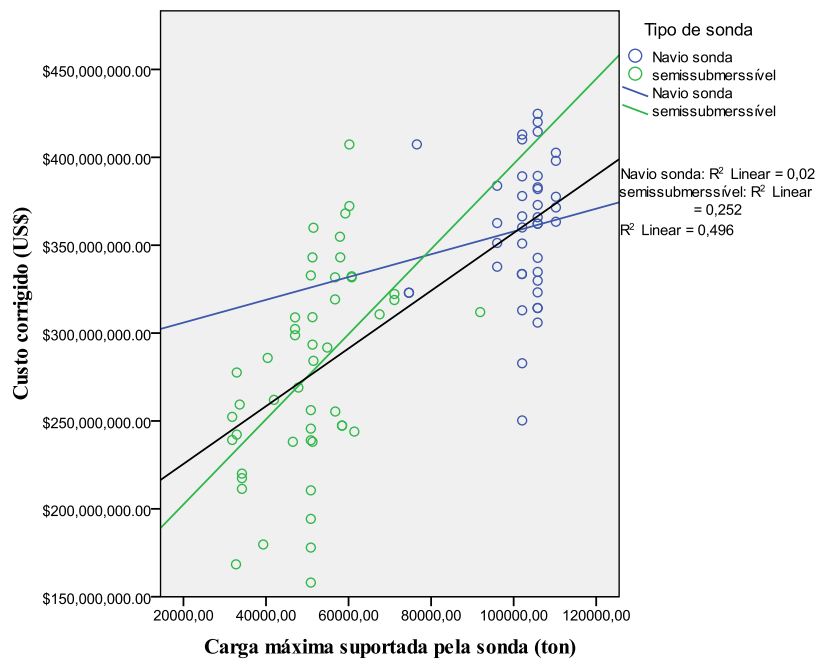


Figura 30 - Gráfico de dispersão entre os custos corrigidos, a carga máxima e o tipo de sonda

Corroborando o que havia sido apresentado no gráfico Boxplot, existe uma clara divisão entre os valores da carga máxima suportada pela sonda dado o tipo de sonda. As linhas de tendência informam que o coeficiente angular, assim como o coeficiente linear são divergentes entre cada grupo e também em relação ao conjunto total de dados. Vale destacar que o R^2 diminuiu de valor após a separação dos dados sendo seu valor para toda a amostra de 0,496, enquanto para o grupo das sondas semissubmersíveis o R^2 é de 0,252 e para o grupo de navios sonda igual a 0,02.

A próxima relação estudada será entre o custo corrigido e a capacidade de carga variável da sonda e está apresentada na Figura 31.

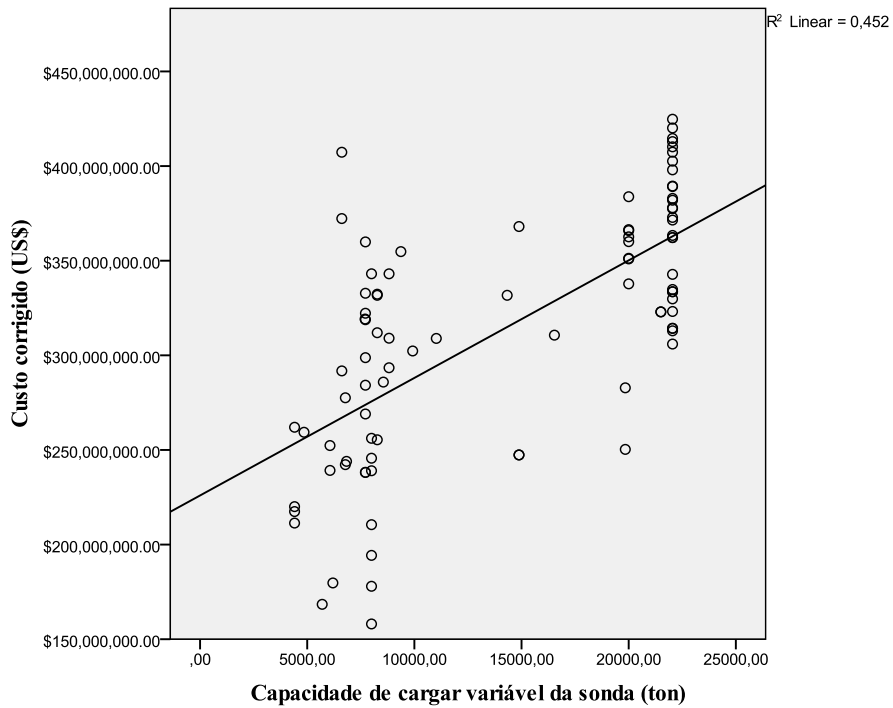


Figura 31 - Gráfico de dispersão entre os custos corrigidos e capacidade de carga variável da sonda

Na figura 31, não fica evidente a presença de dados anômalos, não se pode afirmar que não há homocedasticidade. Em relação à linearidade, o valor de R^2 referente à linha de tendência é relativamente alto, sendo de 0,452 e os dados apresentam correlação positiva.

Em relação à dispersão dos dados, semelhante ao ocorrido com a variável anterior, pela Figura 31 é possível separar dois grupos distintos, os que se encontram com variação de carga entre 5.000 e 10.000 toneladas e o grupo com valores próximos a 20.000 toneladas. Novamente, o tipo de sonda é a característica conceitualmente mais influente. Assim será feito um gráfico Boxplot para essa análise, apresentado na Figura 32.

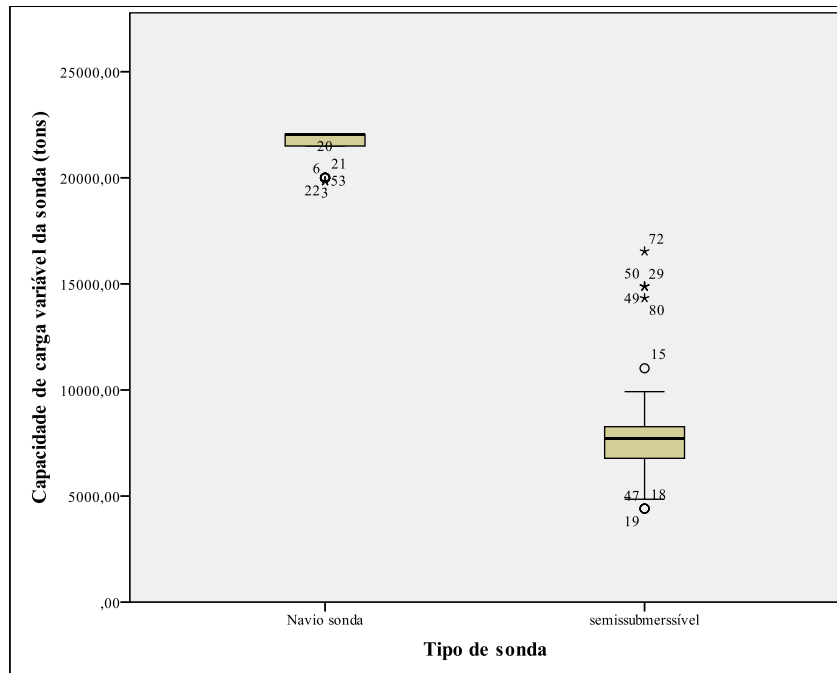


Figura 32 - Boxplot da capacidade de carga variável da sonda pelo tipo de sonda

Da mesma forma que ocorrido na Figura 29, na Figura 32 fica evidenciado a grande diferença entre os valores médios da capacidade de carga variável da sonda entre os dois tipos de sondas. Caso essa variável seja aplicada ao modelo, passar a ser interessante o estudo do efeito moderador do tipo de sonda sobre ela, averiguando se esse é ou não significativo.

A Figura 33, assim como a figura 30, apresenta o gráfico de dispersão entre o custo corrigido, a capacidade de carga variável da sonda e o tipo de sonda.

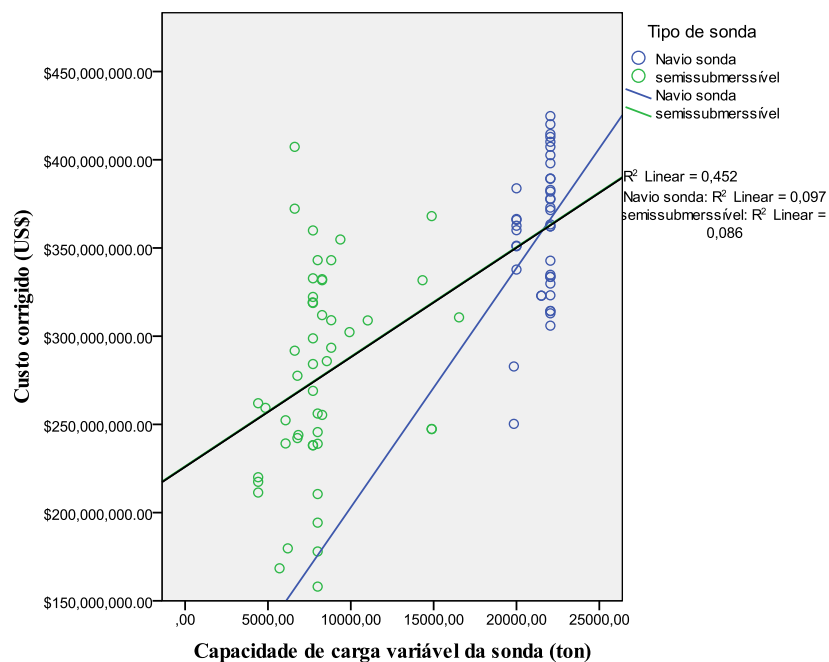


Figura 33 - Gráfico de dispersão entre os custos corrigidos e capacidade de carga variável da sonda

As linhas de tendência foram traçadas, e, novamente, os valores encontrados para o R^2 dos grupos formados pelo tipo de sonda foram muito menores que o da amostra completa. Levando em consideração todos os números, o R^2 foi de 0,452, para os valores referentes às sondas semissubmersíveis foi de 0,086 e para os navios sonda 0,097.

Dessa forma, essas variáveis quando analisada levando-se em conta toda a amostra, apresentam um alto valor, não pela relação com o custo da sonda em si, mas porque estão sendo influenciadas por outra variável, no caso o tipo de sonda.

O mesmo comportamento é esperado para as variáveis comprimento e largura do casco, pois pelo próprio formato dos tipos de sondas estudadas, conceitualmente existe uma diferença entre os dois tipos. Como esses são muito correlacionados com a variável da carga máxima suportável, seus estudos foram suprimidos.

As duas características restantes a serem estudadas são a capacidade do alojamento e a carga máxima suportada pela torre. Conceitualmente essas características não são afetadas pelo tipo de sonda.

A Figura 34 apresenta a relação entre o custo corrigido e a capacidade do alojamento apresentada em forma de um gráfico de dispersão com a linha de tendência traçada.

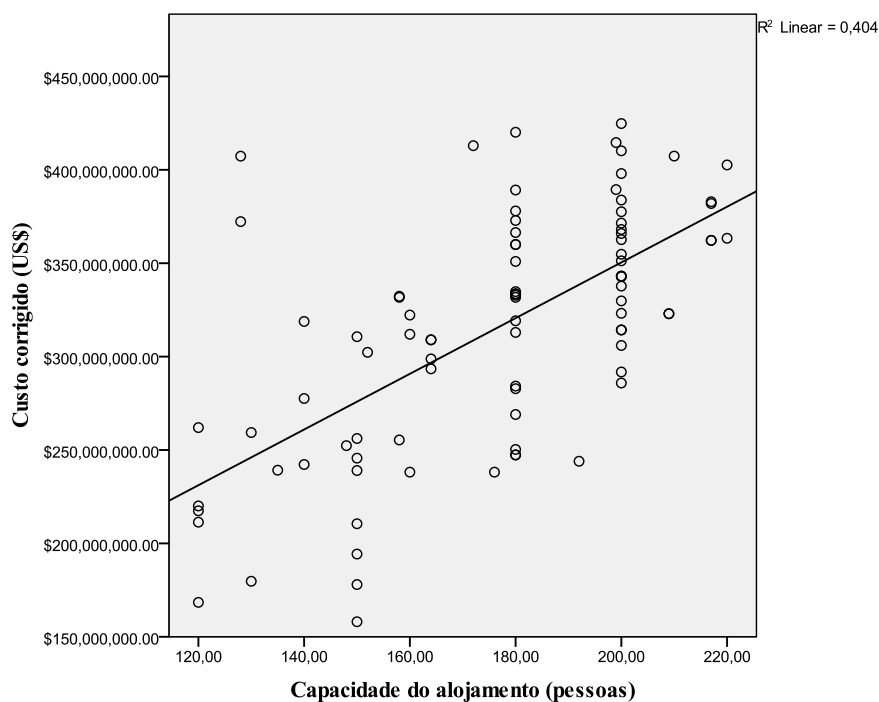


Figura 34 - Gráfico de dispersão entre os custos corrigidos e capacidade do alojamento

Os dados se encontram de forma dispersa e aparentemente existe uma concentração maior de dados ao final da amostra, assim como, uma menor dispersão ao redor da linha de tendência. O R^2 encontrado foi de 0,404, um valor relativamente alto. Não é clara a existência de dados anômalos.

A última relação analisada é entre o custo corrigido e a carga máxima suportada pela torre, que está apresentada na Figura 35 através do gráfico de dispersão e a linha de tendência.

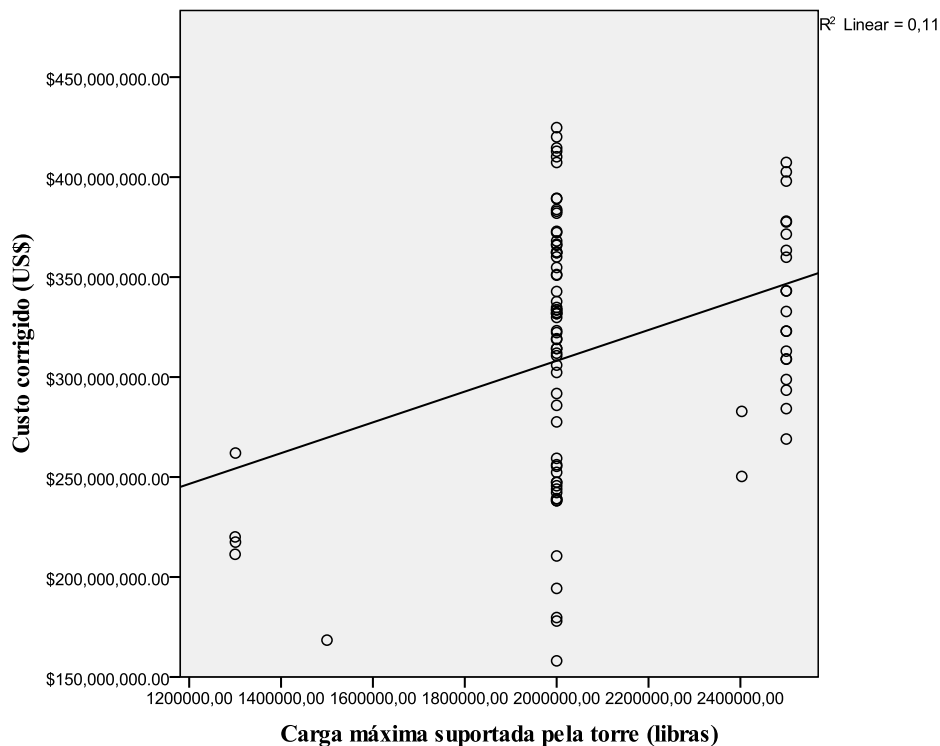


Figura 35 - Gráfico de dispersão entre os custos corrigidos e a carga máxima suportada pela torre

Os dados estão claramente segmentados em 3 grupos, torres com carga máxima abaixo de 1.500.000 libras, torres com carga máxima de 2.000.000 libras e torres com carga máxima entre 2.400.000 e 2.500.000 libras. Existe uma relação linear positiva com R^2 de 0,11, um valor baixo. Pela distribuição da amostra não é razoável analisar a homocedasticidade e a existência de dados anômalos.

5.1.2.2. Variáveis qualitativas

Os estudos envolvendo a variável dependente e as variáveis independentes não métricas será feito separando-se o custo corrigido em grupos, de acordo com as características apresentadas em cada variável.

As variáveis qualitativas foram transformadas em dicotômicas, assim elas passaram a apresentar o valor 0 para a ausência da característica estudada e 1 para sua presença e assim, utilizá-las na regressão. Por exemplo, se a sonda possui top drive, então o valor da variável é 1, caso não o tenha esse valor é 0.

Algumas variáveis necessitam de um tratamento mais elaborado, pois, apresentam mais de duas opções. A variável estágio da construção que a sonda se encontra foi redefinida em duas opções. “Entregue” com valor 1 para as que já foram finalizadas e “Em construção”, com o valor 0, para as que já estão em processo de construção, assim como, para aquelas que só tenham sido encomendadas.

Outra variável que sofreu uma transformação diferenciada foi a local de construção, pois apresenta 5 opções. Assim, foram feitas 5 novas variáveis dicotômicas, que fazem referência a cada uma das regiões, sendo elas, Sudeste Asiático, Oriente, Mar Cáspio, Oriente Médio e Noroeste Europeu. Essas variáveis substituirão a variável região de construção para todos os estudos realizados em seguida.

Continuando os estudos relacionais entre as variáveis independentes qualitativas e a variável dependente, foram feitos testes de comparação das médias dos custos com ou sem a característica analisada. Assim, pode-se notar quais são as variáveis que significativamente representam uma variação no custo da sonda quando presentes. Para isso foram realizadas análises da variância simples (ANOVA) entre os dois grupos. Também foi testada a homocedasticidade entre os dois grupos através do Teste de Levene e os resultados encontrados para essas duas análises estão resumidos na Tabela 17.

Tabela 17 - Resumo da significância e da estatística F para a *one way* ANOVA

Variável	Estatística F	Significância	Estatística Levene	Significância
Tipo de Sonda	59,432	0,000	8,294	0,005
Se foi construída de forma especulativa	0,365	0,547	2,436	0,122
Se apresenta posicionamento dinâmico	4,305	0,041	0,34	0,854
Se é capaz de realizar dupla atividade	21,586	0,000	1,828	0,18
Se está adaptado à ambiente hostil	0,753	0,388	2,584	0,112
Se possui top drive	2,610	0,110	2,882	0,093
Em que estágio da construção a sonda se encontra	0,719	0,399	3,576	0,062
Capacidade de teste extensivos de poços (EWT)	32,187	0,000	7,999	0,006
Se está adaptado ao mar do norte	0,000	0,982	2,714	0,103
Se possui a capacidade de zero desperdício de fluidos	2,711	0,103	0,736	0,393
Se está habilitado para a atividade na Noruega	0,059	0,808	0,211	0,647
Fabricada no sudeste asiático	23,325	0,000	0,106	0,745
Fabricada no Oriente	37,492	0,000	0,249	0,619
Fabricada no mar Cáspio	5,201	0,025	0,051	0,823
Fabricada no Oriente médio	1,584	0,212	2,566	0,113
Fabricada no noroeste europeu	0,017	0,896	3,958	0,050

Fonte: o autor

Pela Tabela 17 é possível inferir que em apenas 7 das 16 variáveis houve uma mudança significativa na média do custo corrigido, com ou sem a característica em questão, essas foram: Tipo de Sonda, Posicionamento Dinâmico, Dupla Atividade, Capacidade de EWT e as variáveis relacionadas ao local de construção: Sudeste Asiático, Oriente e Mar Cáspio. Sendo essas as primeiras escolhas para entrada no modelo.

Das variáveis escolhidas, duas apresentam problemas em relação à homocedasticidade, o Tipo de Sonda e a Capacidade de EWT. O teste de Levene é um teste de inferência estatística que busca a existência de igualdade entre a variância de duas amostras. A hipótese nula é de que os dados apresentam a mesma dispersão, sendo rejeitada, se o valor da significância calculado for menor que a significância de controle definida nas premissas apresentadas nos Capítulos 2 e 4, como 5% para todos os testes. Nos dois casos a hipótese nula foi rejeitada, apesar de ANOVA ser robusta para amostras de tamanhos semelhantes, ainda assim é preciso ter cautela ao se utilizar essas variáveis em uma regressão.

Apesar de não serem variáveis métricas, as variáveis qualitativas se comportam como tal quando transformadas em dicotômicas. Assim, será estudada a correlação entre a variável dependente e as independentes qualitativas. Esse estudo foi feito de duas formas, usando a correlação de Pearson, supondo a existência de uma relação linear e que as variáveis sejam quantitativas e a correlação de Kendall's tau_b, que supõem uma relação não linear e se baseia em pares de variáveis categóricas.

A tabela 18 apresenta os resultados encontrados.

Tabela 18 - Análise da correlação entre a variável dependente e as variáveis independentes qualitativas

Variável	Correlação Pearson		Correlação Kendall's tau_b	
	Custo corrigido	Significância	Custo corrigido	Significância
Tipo de Sonda	-0,635	0,000	-0,542	0,000
Se foi construída de forma especulativa	0,064	0,547	0,023	0,789
Se apresenta posicionamento dinâmico	0,216	0,041	0,16	0,067
Se é capaz de realizar dupla atividade	0,444	0,000	0,373	0,000
Se está adaptado à ambiente hostil	0,092	0,388	0,050	0,556
Se possui top drive	0,170	0,110	0,171	0,049
Em que estágio da construção a sonda se encontra	0,090	0,399	0,102	0,243
Capacidade de teste extensivos de poços (EWT)	0,518	0,000	0,446	0,000
Se está adaptado ao mar do norte	0,002	0,982	-0,045	0,603
Se possui a capacidade de zero desperdício de fluídos	-0,173	0,103	-0,149	0,087
Se está habilitada para a atividade na Noruega	0,026	0,808	0,018	0,833
Fabricada no sudeste asiático	-0,458	0,000	-0,375	0,000
Fabricada no Oriente	0,547	0,000	0,452	0,000
Fabricada no mar Cáspio	-0,236	0,025	-0,157	0,071
Fabricada no Oriente médio	-0,133	0,212	-0,132	0,129
Fabricada no noroeste europeu	0,014	0,896	-0,112	0,891

Fonte: o autor

Analisando os dados da Tabela 18 nota-se que os valores significantes foram, praticamente, os mesmo para os dois tipos de estudo, sendo diferente apenas para as variáveis Posicionamento Dinâmico e Mar Cáspio, que pelo método de Pearson são significantes, enquanto que, pelo método de Kendall's tau_b, não o foram. Outro dado interessante é o fato de que as variáveis que possuem correlação significativa para a correlação de Pearson serem exatamente as mesmas que apresentaram variação

significativa entre as médias das duas amostras pela ANOVA, mostrando que os resultados encontrados são coerentes.

Serão estudadas as correlações entre as variáveis independentes qualitativas, visando mitigar o problema da multicolinearidade.

Na Tabela 19 serão apresentados apenas os resultados mais relevantes.

Tabela 19 - Correlação entre as variáveis independentes qualitativas

Variável	Variável	Correlação
Tipo de sonda	Capacidade de teste extensivos de poços (EWT)	-0,531
Tipo de sonda	Fabricada no sudeste asiático	0,567
Tipo de sonda	Fabricada no Oriente	-0,679
Se apresenta posicionamento dinâmico	Fabricada no mar Cáspio	-0,812
Se está adaptado à ambiente hostil	Se está adaptado ao mar do norte	0,676
Se está adaptado à ambiente hostil	Se está habilitado para a atividade na Noruega	0,588
Se está adaptado ao mar do norte	Se está habilitado para a atividade na Noruega	0,475
Se possui top drive	Fabricada no Oriente Médio	-0,525
Fabricada no sudeste asiático	Fabricada no Oriente	-0,835

Fonte: o autor

A Tabela 19 apresenta apenas os casos onde o valor da correlação ultrapassou o valor de 0,5. Foram selecionados apenas 9 casos diante desse critério, sendo válido destacar, que para os dois métodos utilizados, Pearson e Kendall's tau_b, os valores encontrados foram os mesmos.

Ao analisar a entrada das variáveis no modelo, as correlações devem ser consideradas e se deve ter cautela em relação à multicolinearidade.

Para melhor entender os dados, serão feitos estudos de análise cruzada para algumas correlações encontradas, pois, os valores encontrados foram significantes, apesar de não haver nenhum fundamento teórico que o sustente.

A relação entre o Tipo de Sonda e o local de construção da sonda não possui fundamento teórico para a alta correlação encontrada. Dessa forma, a Figura 36 ilustrará essa relação através de uma análise cruzada, representada num gráfico de barras para que sejam aferidas maiores informações.

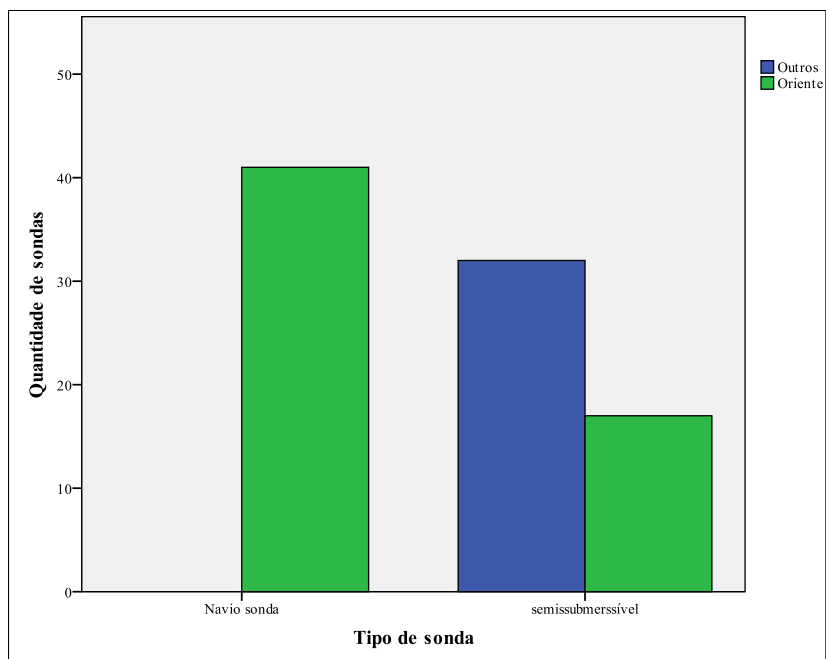


Figura 36 - Relação entre o tipo de sonda e oriente

A Figura 36 indica que todos os navios sondas, nessa amostra, foram feitos na região Oriente, por isso, há a correlação positiva entre os dados. Isso deve ser levado em consideração para a execução do modelo, visto que essa informação tem um caráter muito restritivo, pois, a amostra assume que todos os navios sonda são feitos na região Oriente, o que não é verdade absoluta no mundo real. Logo isso deve ser usado com cautela.

A segunda relação a ser estudada é o Tipo de Sonda e a Capacidade de EWT. A Figura 37 mostra graficamente essa relação.

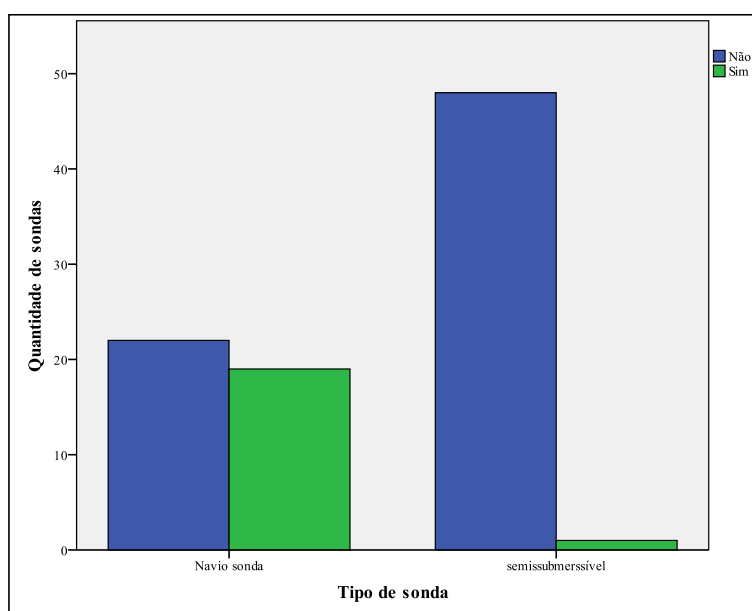


Figura 37 - Relação entre o tipo de sonda e a capacidade de EWT

Pela Figura 37 é possível notar que a maioria das sondas que possuem a capacidade de EWT é do tipo navio sonda. Essa alta correlação se deu pelos dados da amostra.

As correlações relativas a sondas fabricadas no Oriente-Médio e no Mar Cáspio são de difícil interpretação, dada a baixa quantidade de dados de sondas construídas nesses locais, na amostra. Apenas três sondas da amostra foram feitas no Oriente-Médio e duas no Mar Cáspio, dessa forma qualquer característica que esteja presente apresenta uma alta correlação. Os estudos univariados, referentes às variáveis independentes, encontram-se no apêndice.

Os demais valores de correlação encontrados eram esperados, dada a semelhança entre as características presentes nas variáveis.

Ao analisar os dois tipos de variáveis, é possível selecionar como variáveis quantitativas mais representativas a carga máxima suportada pela sonda, a capacidade de carga variável da sonda, a profundidade da lamina d'água e a profundidade de perfuração. No entanto, é preciso destacar a grande influência da variável tipo de sonda sobre a variável carga máxima suportada pela sonda e capacidade de carga variável da sonda.

Em relação às variáveis qualitativas, sete se mostraram significativas: o Tipo de Sonda, a existência de Posicionamento Dinâmico, a Capacidade de EWT, se possui top drive e se foi construída no Oriente, no Sudeste Asiático, ou no Mar Cáspio. Aqui também cabe a ressalva sobre a variável Tipo de Sonda, que tem influência sobre a Capacidade de EWT e sobre as variáveis referentes ao local de construção da sonda, com destaque para a variável Oriente.

A possibilidade de se separa em duas amostras diferentes dada a característica Tipo de sonda foi descartada após terem sido realizados os primeiros estudos. As duas novas amostras de custos corrigidos não se comportam como uma distribuição normal, o número de dados anômalos cresceu e há a perda de informação por se tratarem de amostras menores. Dessa forma esse estudo envolvendo a divisão da amostra foi descartado.

5.2. ANÁLISES MULTIVARIADAS

Conforme mencionado no Capítulo 4, as análises multivariadas foram feitas em três etapas, a construção de uma regressão linear simples, a regressão múltipla e a regressão múltipla após a análise fatorial.

5.2.1. REGRESSÃO LINEAR SIMPLES

A primeira etapa será a construção de um modelo com uma única variável. A variável escolhida foi a carga máxima suportada pela da sonda, por apresentar a maior correlação com a variável dependente e por ter uma alta correlação com várias outras variáveis quantitativas.

As Tabelas 20, 21 e 22, apresentam os resultados da regressão simples.

Tabela 20 - Sumário da regressão do custo corrigido pela variação de carga suportável

R	R ²	R ² ajustado	Erro padrão do estimador
0,704	0,496	0,490	4,612 E7

Fonte o autor

Tabela 21 - Análise da variância da regressão do custo corrigido pela variação de carga suportável

	Somas dos quadrados	Graus de liberdade	Média dos quadrados	F	Sig.
Regressão	1,842 E+17	1	1,704 E+17	86,601	0,000
Resíduo	1,872 E+17	88	2,284 E+15		
Total	3,714 E+17	89			

Fonte o autor

Tabela 22 - Coeficientes da regressão do custo corrigido pela variação de carga suportável

Variáveis	Coeficientes	Erro padrão	Beta	t valor	Significância
Constante	1,927 E+08	1,397 E+07		13,792	0,000
Carga máxima suportada pela da sonda	1.642,493	176,499	0,704	9,306	0,000

Fonte o autor

O valor de R é o mesmo da correlação linear de Pearson 0,704, no entanto o valor que será comparado com outros modelos será o R² ajustado, igual a 0,490, pois leva em consideração os graus de liberdade e o número de variáveis independentes na regressão.

A regressão é válida, pois a estatística F apresenta um valor alto, 86,601 e a significância encontrada é 0,000, menor que os 5% definidos. Analisando os

coeficientes individualmente, é possível ver o valor de $1,927E+08$ para a constante com um t-valor de 13,792 e significância de 0,000, assim, esse valor é estatisticamente diferente de zero. O valor do coeficiente da carga máxima suportada pela sonda é 1.642,49 com t-valor de 9,306 e significância de 0,000, sendo assim também significativamente diferente de zero. Em relação a essa variável existe também o coeficiente Beta que representa o valor padronizado do coeficiente da regressão, cujo objetivo é comparar os coeficientes quanto a sua relevância, mesmo quando as variáveis possuem ordem de grandezas distintas.

Foram feitos os gráficos de probabilidade normal, assim como, o gráfico de dispersão para análise de resíduos padronizados. A Figura 38 testa a normalidade dos resíduos.

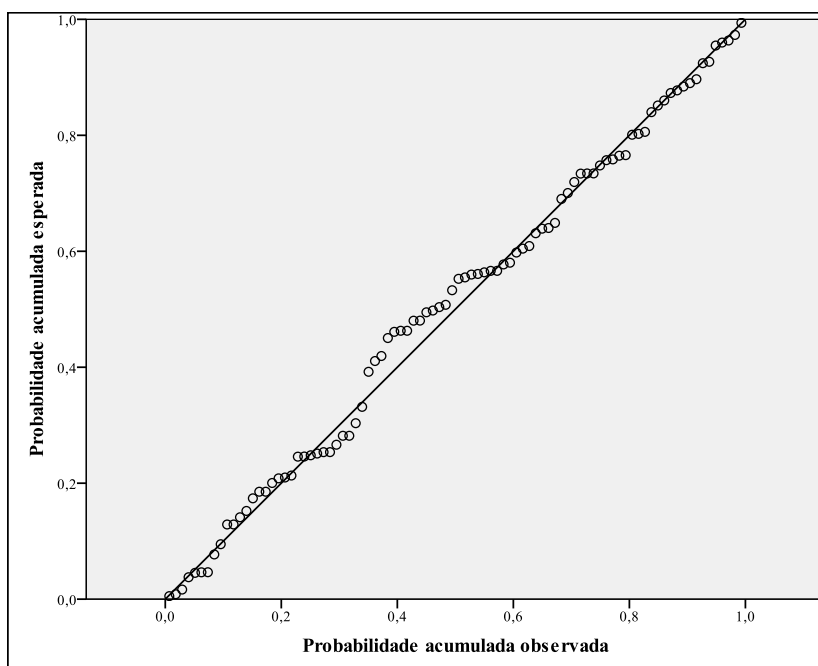


Figura 38 - Gráfico de probabilidade normal: resíduos padronizados da regressão simples

Como apresentado na Figura 38, os valores estão ao longo da diagonal sem apresentar desvios muito grandes nem de forma sistemática. Assim é possível afirmar que os resíduos são representativos de uma distribuição normal satisfazendo a suposição de normalidade.

A Figura 39 apresenta os resíduos em forma de gráfico de dispersão. No eixo das abscissas se encontram os valores previstos e no eixo das ordenadas os valores dos resíduos normalizados. Esse gráfico foi feito para analisar as premissas de linearidade e homocedasticidade.

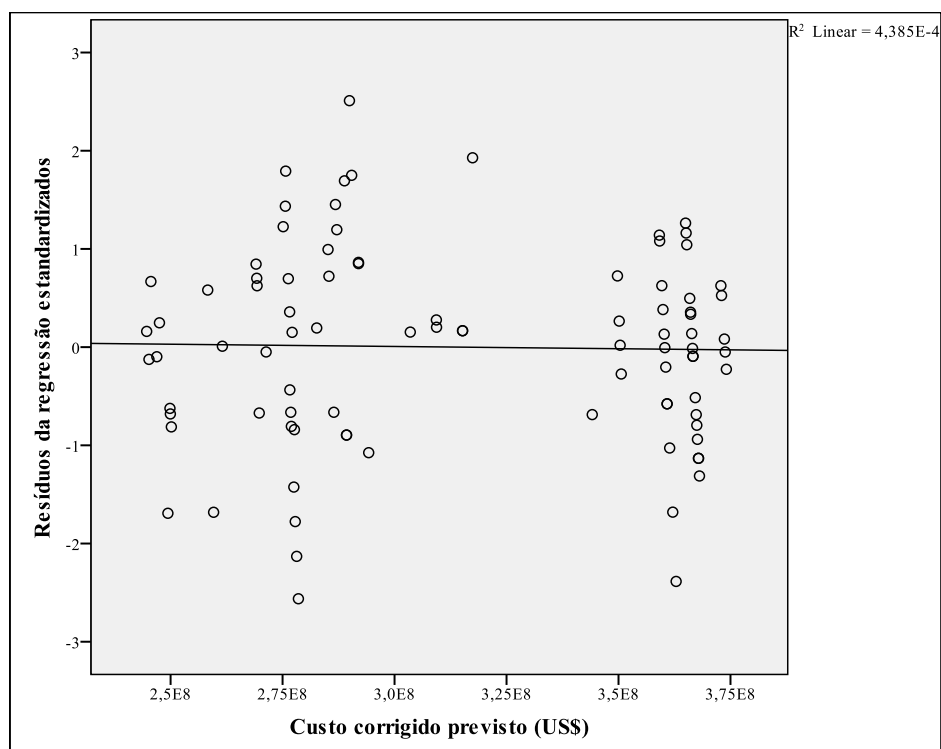


Figura 39 - Gráfico de dispersão dos resíduos padronizados da regressão simples

Os resíduos apresentam um padrão de dispersão decrescente, indicando que pode existir a heterocedasticidade, no entanto, eles não seguem nenhum padrão linear, o que afasta a hipótese de não haver linearidade.

É possível observar também, que os resíduos não estão uniformemente distribuídos pela amostra, dessa forma, a regressão não é boa, pois existe influência sobre os resíduos. Como esse estudo é apenas base de comparação não serão feitos tratamentos para a melhoria da regressão.

5.2.2. REGRESSÃO MÚLTIPLA

A segunda etapa consiste na realização de uma regressão múltipla entre a variável custo corrigido e as demais variáveis apresentadas na Seção 5.1.

A regressão múltipla será realizada com o método *stepwise*, incluindo todas as variáveis. Este método está descrito, de forma mais detalhada na seção 2.2.4.4.

Antes de ser feita a regressão é preciso criar novas variáveis que representam o efeito moderador de variável Tipo de Sonda sobre as variáveis carga máxima suportada pela sonda, capacidade de carga variável da sonda, comprimento e largura do casco,

Oriente e Capacidade de EWT. Essas variáveis foram criadas multiplicando a variável Tipo de Sonda com as outras que aparentemente sofrem o seu efeito.

A Tabela 23 apresenta os passos feitos pelo método *stepwise* até se encontrar o melhor modelo.

Tabela 23 - Variáveis entrantes e removidas pelo método *stepwise*

Variáveis entrantes	Variáveis removidas
Carga máxima suportada pela sonda	
Capacidade do alojamento	
Habilitado para atividade na Noruega	
Capacidade de EWT	
Carga máxima suportada pela torre	
Fabricado no sudeste asiático	
	Carga máxima suportada pela sonda
Adaptado a ambiente hostil	

Fonte: o autor

O critério de entrada usado pelo método foi a variável apresentar significância abaixo de 0,05 e para retirada de alguma variável no modelo, após sua entrada, foi o valor de significância superar 0,10.

A regressão com melhor resultado levou em consideração as seguintes variáveis: capacidade do alojamento, Habilitado para atividade na Noruega, Capacidade de EWT, carga máxima suportada pela torre, se foi produzida no Sudeste Asiático e se está adaptado a ambiente hostil. A variável carga máxima suportada pela sonda foi a primeira a entrar no modelo por apresentar a maior correlação com a variável dependente, porém, com a entrada de outras variáveis, seu poder preditivo acabou perdendo significância, sendo excluída pelo critério adotado.

Assim, como a variável carga máxima suportada pela sonda, outras variáveis que não entraram no modelo podem ter uma correlação alta com a variável dependente. O fato de não terem entrado no modelo ocorreu por essas também serem correlacionadas com outras variáveis independentes que entraram no modelo. Como o método *stepwise* busca a parcimônia, variáveis muito correlacionadas acabam por agregar pouco valor preditivo ao modelo, sendo assim preteridas por outras com perfil mais diferenciado.

As tabelas 24, 25 e 26 apresentam os resultados da regressão múltipla. A Tabela 24 expõe o sumário da regressão múltipla, a Tabela 25 a análise de variância e a Tabela

26 os coeficientes da regressão, seu erro padrão e seus valores, seus Betas e sua significância.

Tabela 24 - Sumário da regressão múltipla do custo corrigido pelo método *stepwise*

R	R ²	R ² ajustado	Erro padrão do estimador
0,846	0,715	0,694	3,572 E+7

Fonte o autor

Tabela 25 - Análise da variância da regressão múltipla do custo corrigido pelo método *stepwise*

	Somas dos quadrados	Graus de liberdade	Média dos quadrados	F	Sig.
Regressão	2,655E+17	6	4,425E+16	34,684	0,000
Resíduo	1,059E+17	83	1,276E+15		
Total	3,714E+17	89			

Fonte: o autor

Tabela 26 - Coeficientes da regressão da regressão múltipla do custo corrigido pelo método *stepwise*

Variáveis	Coeficientes	Erro padrão	Beta	t valor	Significância
Constante	- 7,556 E+07	4,066 E+07		-1,858	0,067
Capacidade do alojamento	1.128.061,62	181.753,42	0,481	6,207	0,000
Habilitado para a atividade na Noruega	8,105 E+07	1,675 E+07	0,470	4,839	0,000
Capacidade de EWT	4,312 E+07	9.991.272,48	0,279	4,315	0,000
Carga máxima suportada pela torre	92,75	20,53	0,400	4,519	0,001
Fabricado no sudeste asiático	-4,485 E+07	1,012+07	-0,313	-4,433	0,000
Adaptado à ambiente hostil	-3,340 E+07	1,181 E+07	-0,251	-2,827	0,006

Fonte: o autor

O valor de R encontrado de 0,846 foi maior que o anterior de 0,704, enquanto o R² ajustado aumentou de 0,49 para 0,694. A regressão multivariada representa de forma melhor a variável custo corrigido que a regressão simples, realizada com a variável capacidade máxima da sonda.

A estatística F reduziu seu valor de 86,601 para 34,684, no entanto, a significância se manteve em 0,000, menor que os 5% definidos, sendo assim, a regressão é válida.

Analisando os coeficientes individualmente, os valores encontrados foram - 7,556 E+07 para a constante com t-valor de -1,858 e significância de 0,067, 1.128.061,62 para a capacidade do alojamento, com t-valor de 6,207 e significância de 0,000, 8,105 E+07 para Habilitado para atividade na Noruega, com t-valor de 4,839 e

significância de 0,000, 4,312 E+7 para a Capacidade de EWT com t-valor de 4,315 e significância de 0,000, 92,75 para a carga máxima suportada pela torre com t-valor de 4,519, -4,485 E+07 para o Sudeste Asiático com t-valor de -4,433 e significância de 0,000 e -3,340 E+07 para a Adaptado a ambiente hostil o com t-valor de -2,827.

Para todas as variáveis selecionadas, os coeficientes foram significativamente diferentes de zero, no entanto, para a constante, o t-valor ficou acima do limite de 0,05, mas devido à proximidade com o limite, o seu valor será mantido.

É necessário destacar o grande valor do erro padrão da constante, 4,066 E+07, valor muito próximo do coeficiente em si, que tem relação direta com a menor significância observada. Os outros erros padrão possuem valores mais suaves, não gerando nenhuma percepção especial.

Analisando o valor de Beta das variáveis, é possível comparar, e, assim ranquear, as variáveis de acordo com sua importância no modelo. As mais importantes são aquelas que apresentam o maior valor absoluto de Beta, sendo que as mais relevantes são a capacidade do alojamento, a Habilitado para a atividade na Noruega e a carga máxima suportada pela torre.

Após a regressão é necessário fazer o diagnóstico da regressão para saber se essa respeitou as premissas existentes. A primeira suposição que será testada é a não existência de multicolinearidade.

A Tabela 27 apresenta o estudo da multicolinearidade através dos valores de Tolerância e VIF.

Tabela 27 - Estatísticas de colinearidade da regressão múltipla

Variáveis	Tolerância	VIF
Capacidade do alojamento	0,573	1,746
Habilitado para a atividade na Noruega	0,364	2,748
Capacidade de EWT	0,822	1,217
Carga máxima suportada pela torre	0,438	2,281
Fabricado no sudeste asiático	0,690	1,449
Adaptado à ambiente hostil	0,437	2,286

Fonte: o autor

Valores de Tolerância próximos a 0,200, assim como, valores de VIF acima de 5,000, são fatores de atenção. O valor da Tolerância mede o quanto da variância de uma variável não está sendo explicada pelas outras variáveis da regressão. Dessa forma, a

multicolinearidade presente não é grande o suficiente para distorcer os resultados de maneira significativa, ou seja, as variáveis escolhidas para a regressão não apresentam valores preocupantes.

A variável que possui a maior correlação com a variável dependente não entrou no modelo. Isso pode se explicado pela sua alta correlação com outras variáveis entrantes que, pelo princípio da multicolinearidade leva a escolha de variáveis menos correlacionadas.

Para o estudo das demais premissas, será feito um gráfico de probabilidade normal, assim como um gráfico de dispersão para a análise de resíduos padronizados.

A Figura 40 testa a normalidade dos resíduos via gráfico PP.

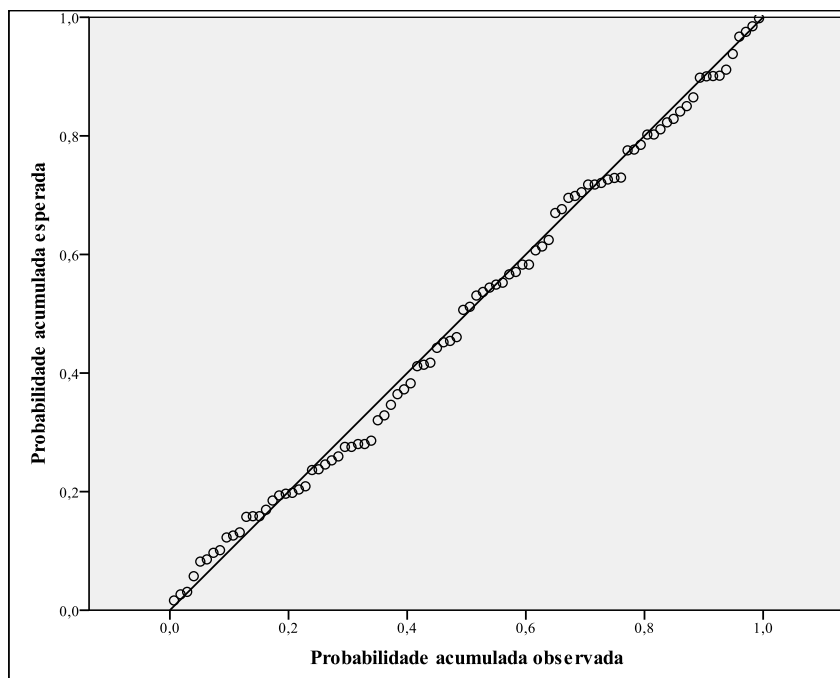


Figura 40 - Gráfico de probabilidade normal: resíduos padronizados da regressão múltipla

Como mostrado da Figura 40, os valores estão ao longo da diagonal sem apresentarem grandes desvios, nem mesmo de forma sistemática. Assim, é possível afirmar que os resíduos são representativos de uma distribuição normal satisfazendo a suposição de normalidade.

A Figura 41 apresenta os resíduos em forma de gráfico de dispersão. No eixo das abscissas estão os valores previstos padronizados e no eixo das ordenadas os valores dos resíduos padronizados. Esse gráfico foi feito para analisar a premissa de homocedasticidade dos resíduos e a linearidade, além da existência de valores anômalos.

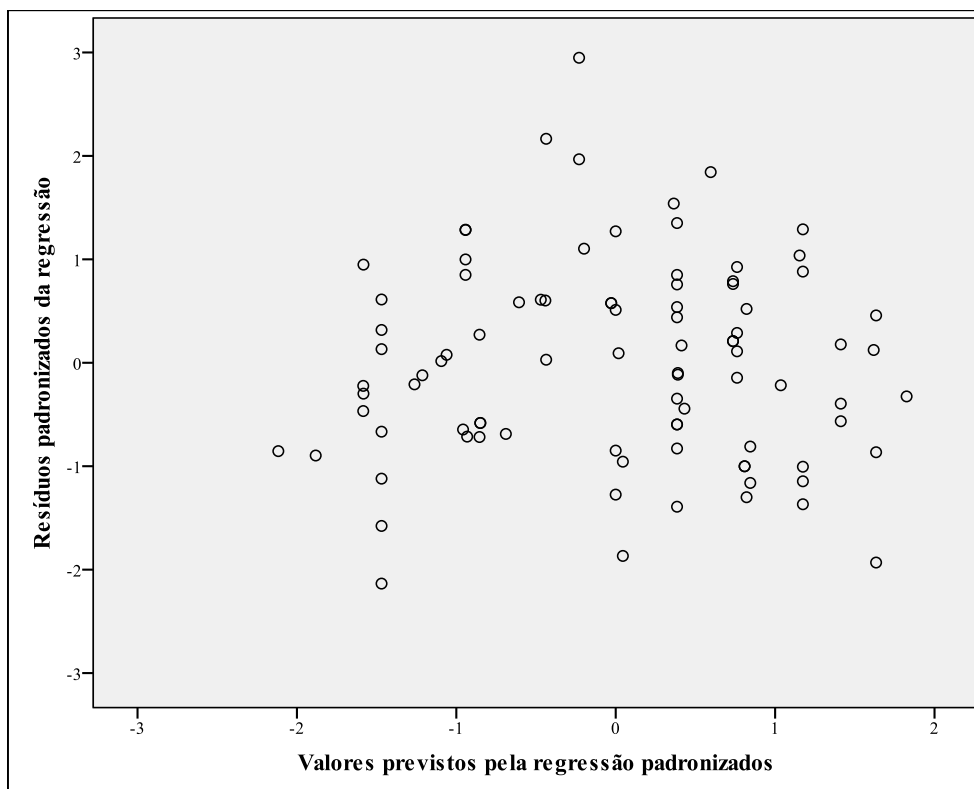


Figura 41 - Gráfico de dispersão dos resíduos padronizados para a regressão múltipla

Diferente do ocorrido no caso da regressão simples, os resíduos não apresentam nenhum padrão de dispersão, nem de linearidade, reduzindo as chances de ocorrência de heterocedasticidade e de não haver linearidade.

Não serão analisados os gráficos de regressão parcial, pois na seção 5.2 as relações entre as variáveis independentes e a dependente foram minuciosamente estudadas.

Existe uma sonda que o valor previsto ficou muito aquém do valor real, porém, não é possível caracterizá-lo como um valor anômalo, pois, ele não ultrapassa o limite de 3 desvios-padrão, além de não apresentar nenhuma característica relevante para sua exclusão.

Outro método que poderia ser utilizado é o da especificação confirmatória, onde todas as variáveis selecionadas entram no modelo. Assim, as variáveis mais importantes, conforme definido pela teoria seriam incluídas no modelo de forma direta.

Como o objetivo final é a previsão do custo para uma estimativa do custo da sonda, a multicolinearidade pode ser aceita, no entanto, serão perdidas as informações relativas às variáveis e seus coeficientes, restando apenas a previsão.

Porém, todos os casos gerados apresentaram valores de R^2 corrigido menores que os encontrados pelo método *stepwise*, dessa maneira, não serão apresentados os

dados referentes a outros métodos, pois esses, além de apresentarem resultados menos satisfatórios, ainda perdiam uma quantidade significativa de informação por conta da alta multicolinearidade.

5.2.3. ANÁLISE FATORIAL

Na última etapa do estudo, será feita uma análise fatorial com a intenção de reduzir o número de variáveis, mas ao mesmo tempo, mantendo o máximo possível de suas informações. Será utilizado o método dos componentes principais para se encontrar os fatores, que em seguida, serão utilizados na composição de uma regressão linear múltipla.

A análise fatorial terá os mesmos *inputs* que a regressão múltipla, a ideia é comparar se os fatores oferecem mais poder preditivo que as variáveis na sua forma original.

Os testes KMO e o teste de esfericidade de Bartlett's foram feitos, respectivamente, para averiguar se existe correlação suficiente para se realizar uma análise fatorial e para se rejeitar a hipótese que a matriz de correlação é uma matriz identidade, o que inviabilizaria o procedimento. Os valores encontrados estão representados na Tabela 28.

Tabela 28 - Teste de KMO e de Bartlett's

KMO		0,752
Teste de Bartlett	Chi.quadrada	869,43
	Graus de liberdade	66
	Sig.	0,000

Fonte: o autor

O teste KMO com valor calculado acima de 0,70 garante a existência de correlação suficiente entre as variáveis e a hipótese nula foi rejeitada, além disso, a matriz não é identidade.

A comunalidade é a proporção da variância de uma variável observada, que é explicada pelos componentes extraídos, sendo que seu valor pode variar de 0 a 1. Na Tabela 29 se encontram os valores para as comunalidades iniciais e extraídas.

Tabela 29 - Comunalidades

	Inicial	Extraída
Adaptado a ambiente hostil	1,000	0,832
Adaptado ao mar do norte	1,000	0,781
Capacidade de carga variável da sonda	1,000	0,949
Capacidade de água potável	1,000	0,604
Capacidade do alojamento	1,000	0,764
Carga máxima suportada pela sonda	1,000	0,901
Carga máxima suportada pela torre	1,000	0,859
Comprimento do casco	1,000	0,940
Descarga zero	1,000	0,500
Capacidade de dupla atividade	1,000	0,543
Especulativa	1,000	0,595
Estagio da Construção	1,000	0,786
Fabricada no mar Cáspio	1,000	0,835
Fabricada no nordeste europeu	1,000	0,831
Fabricada no Oriente	1,000	0,861
Fabricada no Oriente Médio	1,000	0,694
Fabricado no sudeste asiático	1,000	0,867
Habilitado para atividade na Noruega	1,000	0,823
Largura do Casco	1,000	0,834
Possui Posicionamento dinâmico	1,000	0,854
Profundidade de perfuração	1,000	0,669
Tipo de sonda	1,000	0,947
Top drive	1,000	0,663
Volume água de perfuração	1,000	0,402
Capacidade de EWT	1,000	0,716
Profundidade lâmina d'água	1,000	0,777

Fonte: o autor

Os dados expostos na Tabela 29 atestam que os valores associados a cada variável são suficientemente altos, mas vale uma ressalva. Para as variáveis volume de água de perfuração, descarga de fluídos zero, capacidade de realizar dupla atividade e se a sonda é especulativa os valores foram medianos e ficaram abaixo de 0,600.

Os autovalores representam o total de variância que pode ser explicado por cada um dos componentes. Foi determinado como premissa, que apenas valores maiores que 1 serão extraídos.

Na Tabela 30 estão todos os autovalores com seus valores iniciais, o percentual de variância que explicam e seu percentual acumulado.

Tabela 30 - Autovalores iniciais

Componentes	Autovalores iniciais		
	Total	% da variância	% cumulativo
1	7,798	29,992	29,992
2	3,158	12,145	42,137
3	3,013	11,588	53,724
4	1,829	7,033	60,757
5	1,593	6,126	66,883
6	1,260	4,847	71,730
7	1,176	4,523	76,253
8	0,982	3,777	80,029
9	0,829	3,190	83,219
10	0,748	2,878	86,097
11	0,643	2,475	88,572
12	0,547	2,104	90,676
13	0,508	1,956	92,631
14	0,394	1,517	94,148
15	0,317	1,218	95,366
16	0,278	1,070	96,436
17	0,267	1,025	97,462
18	0,208	0,799	98,261
19	0,153	0,589	98,850
20	0,128	0,493	99,342
21	0,073	0,280	99,622
22	0,048	0,186	99,808
23	0,029	0,113	99,921
24	0,018	0,068	99,989
25	0,003	0,011	100,000
26	0,000	0,000	100,000

Fonte: o autor

Como visto na Tabela 30, os componentes de 1 a 7 apresentam valores maiores que 1 e seu percentual de variância acumulada é 76,253%, próximo a 80%, o valor tido como referência. Assim esses 7 componentes foram extraídos.

Na Tabela 31 estão apresentados os componentes selecionados para extração, com seus valores de variância explicada total, percentual e o percentual cumulativo e seus valores após a rotação fatorial, mediante uso do método Varimax.

Tabela 31 - Total da variância explicada

Componentes	Soma dos quadrados da extração			Soma dos quadrados rotacionados		
	Total	% da variância	% cumulativo	Total	% da variância	% cumulativo
1	7,798	29,992	29,992	6,919	26,611	26,611
2	3,158	12,145	42,137	3,400	13,078	39,689
3	3,013	11,588	53,724	2,498	9,608	49,297
4	1,829	7,033	60,757	2,179	8,382	57,679
5	1,593	6,126	66,883	1,913	7,357	65,036
6	1,260	4,847	71,730	1,507	5,797	70,833
7	1,176	4,523	76,253	1,409	5,420	76,253

Fonte: o autor

Como pode ser visto, os componentes se tornam mais equilibrados em relação à quantidade de variância explicada após a rotação. Por último serão analisadas as matrizes componentes, onde os componentes ou fatores são associados às variáveis originais.

A Tabela 32 mostra essa associação antes da rotação fatorial.

Tabela 32 - Matriz componente

	Componentes						
	1	2	3	4	5	6	7
Tipo de sonda	-0,917	-0,265	0,164	-0,045	-0,058	0,029	0,063
Comprimento do casco	0,913	0,282	-0,146	0,022	0,052	-0,006	-0,052
Largura do Casco	-0,887	-0,151	0,127	0,012	-0,038	0,063	0,047
Capacidade de carga variável da sonda	0,878	0,363	-0,176	0,104	0,059	0,029	-0,025
Carga máxima suportada pela sonda	0,859	0,314	-0,156	0,146	0,102	0,079	0,048
Fabricada no Oriente	0,858	-0,244	0,104	0,179	0,052	-0,076	-0,120
Fabricado no sudeste asiático	-0,788	0,379	-0,052	0,063	0,247	0,036	-0,182
Capacidade de EWT	0,594	0,091	0,039	-0,016	0,073	0,584	-0,080
Volume água de perfuração	0,502	-0,126	0,117	0,014	0,273	-0,112	0,184
Dupla atividade	0,438	0,198	0,270	0,088	0,085	0,243	0,406
Carga máxima suportada pela torre	-0,013	0,897	0,218	0,019	-0,072	0,013	-0,015
Profundidade de perfuração	0,222	0,720	-0,027	0,200	0,246	0,004	-0,010
Profundidade lâmina d'água	0,212	0,685	-0,231	0,408	-0,007	0,079	0,190
Capacidade do alojamento	0,536	0,663	-0,117	0,108	0,029	-0,088	-0,057
Habilitado para atividade na Noruega	0,015	-0,605	0,535	0,167	0,159	0,107	0,327
Adaptado à ambiente hostil	0,094	-0,075	0,886	0,020	0,016	0,078	0,160
Adaptado ao mar do norte	-0,223	0,049	0,833	0,026	0,119	0,136	0,036
Descarga zero	-0,281	0,004	0,533	0,252	0,120	-0,166	-0,175
Possui posicionamento dinâmico	0,133	0,102	0,064	0,903	-0,062	0,039	0,018
Fabricada no mar Cáspio	-0,099	-0,171	-0,128	-0,879	0,032	0,073	0,022
Fabricada no Oriente Médio	-0,159	-0,106	-0,092	0,081	-0,800	0,035	0,017
Top drive	0,091	-0,037	0,049	-0,110	0,790	0,121	-0,030
Especulativa	-0,127	0,036	0,113	0,406	0,582	-0,110	0,222
Estágio da Construção	-0,105	0,021	0,179	-0,080	0,032	0,842	0,162
Capacidade de água potável	0,423	0,261	0,178	-0,103	0,105	-0,504	0,222

Fonte: o autor

Os componentes se apresentam de forma mal distribuída, deixando assim confusa a sua interpretação, somado a isso existe uma participação excessiva dos componentes nas variáveis.

A Tabela 33 apresenta os valores após a rotação fatorial pelo método Varimax.

Tabela 33 - Matriz componente rotacionada

	Componentes						
	1	2	3	4	5	6	7
Tipo de sonda	0,970	-0,058	-0,037	0,015	0,034	-0,017	-0,022
Comprimento do casco	-0,959	0,025	0,149	-0,014	0,023	-0,026	0,051
Largura do Casco	0,956	-0,019	-0,142	0,036	0,010	0,039	-0,048
Capacidade de carga variável da sonda	0,943	0,024	-0,020	0,011	0,045	-0,087	0,021
Carga máxima suportada pela sonda	-0,877	-0,031	0,239	-0,004	0,043	-0,064	0,039
Fabricada no Oriente	0,762	-0,265	0,292	0,073	0,025	0,148	0,001
Fabricado no sudeste asiático	0,703	0,369	-0,334	-0,230	-0,158	0,006	-0,206
Capacidade de EWT	0,567	0,206	-0,114	0,170	0,418	-0,264	-0,257
Volume água de perfuração	0,457	0,411	0,169	0,072	0,291	-0,004	0,218
Dupla atividade	0,422	0,384	-0,140	0,070	-0,199	0,009	0,114
Carga máxima suportada pela torre	-0,238	0,847	-0,108	-0,160	-0,021	-0,084	0,069
Profundidade de perfuração	-0,052	0,758	0,212	-0,011	0,207	0,388	-0,128
Profundidade lâmina d'água	-0,278	0,594	0,407	0,116	0,180	0,303	-0,219
Capacidade do alojamento	0,286	-0,270	0,677	0,195	0,274	0,364	-0,017
Habilitado para atividade na Noruega	-0,538	-0,354	0,587	0,272	-0,099	-0,114	-0,104
Adaptado à ambiente hostil	0,520	-0,188	0,568	0,201	-0,015	0,013	0,005
Adaptado ao mar do norte	0,533	-0,297	0,552	-0,095	0,140	-0,149	0,220
Descarga zero	-0,003	0,326	0,459	0,075	-0,413	-0,279	0,156
Possui posicionamento dinâmico	-0,276	0,304	0,399	-0,116	-0,186	0,217	-0,276
Fabricada no mar Cáspio	0,274	0,143	0,498	-0,639	-0,062	-0,305	-0,079
Fabricada no Oriente Médio	-0,259	-0,126	-0,565	0,606	0,140	0,191	0,099
Top drive	0,117	0,312	0,082	0,579	-0,333	-0,299	-0,100
Especulativa	-0,232	-0,283	-0,185	-0,567	0,418	0,162	0,061
Estágio da Construção	-0,121	0,272	0,099	0,271	0,682	-0,375	-0,089
Capacidade de água potável	0,455	0,123	0,067	0,075	-0,275	0,471	0,271

Fonte: o autor

É importante associar os fatores as características gerais das sondas, chamadas de implícitas. São elas que garantem a fundamentação teórica para que os fatores sejam representativos. O componente 1 é muito correlacionado com características relacionadas ao casco, como comprimento do casco, largura, Tipo de sonda, assim, pode ser considerado o fator referente ao casco. O componente 2 está relacionado com

características de adaptação como Ambiente hostil ou Habilitado para a Noruega, assim, seria um componente relativo à adaptação ao meio ambiente em que irá funcionar. O componente 3 está associado aos equipamentos, sendo assim referente à esses. As outras relações são mais sutis, dada à diminuição dos valores.

Com os componentes devidamente rotacionados, será feita uma regressão múltipla tendo esses como variáveis independentes e o custo corrigido como variável dependente. O primeiro passo é fazer uma matriz de correlação entre as variáveis candidatas do modelo, ou seja, o custo corrigido e os 7 fatores.

A Tabela 34 apresenta esses valores.

Tabela 34 - Matriz de correlação entre o custo corrigido e componentes fatoriais

Variáveis	Custo corrigido	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5	Fator 6	Fator 7
Custo corrigido	1,000	0,646	0,328	0,069	0,155	0,072	0,190	0,115
Fator 1	0,646	1,000	0,000	0,000	0,000	0,000	0,000	0,000
Fator 2	0,328	0,000	1,000	0,000	0,000	0,000	0,000	0,000
Fator 3	0,069	0,000	0,000	1,000	0,000	0,000	0,000	0,000
Fator 4	0,155	0,000	0,000	0,000	1,000	0,000	0,000	0,000
Fator 5	0,072	0,000	0,000	0,000	0,000	1,000	0,000	0,000
Fator 6	0,190	0,000	0,000	0,000	0,000	0,000	1,000	0,000
Fator 7	0,115	0,000	0,000	0,000	0,000	0,000	0,000	1,000

Fonte: o autor

Como visto na Tabela 34 o pressuposto de não existir correlação entre os fatores foi atendido, sendo que apenas as correlações entre a variável dependente e os componentes apresentaram valores diferentes de 0.

As tabelas 35, 36 e 37 apresentam os resultados da regressão múltipla. A Tabela 35 expõe o sumário da regressão múltipla, a Tabela 36 a análise de variância e a Tabela 37 os coeficientes da regressão, seu erro padrão e seus valores, seus Betas e sua significância.

Tabela 35 - Sumário da regressão do custo corrigido pelos fatores

R	R ²	R ² ajustado	Erro padrão do estimador
0,765	0,585	0,565	4,260 E+7

Fonte: o autor

Tabela 36 - Análise da variância da regressão do custo corrigido pelos fatores

	Somas dos quadrados	Graus de liberdade	Média dos quadrados	F	Sig.
Regressão	2,171 E+17	4	5,427 E+16	29,901	0,000
Resíduo	1,543 E+17	85	1,815 E+15		
Total	3,714 E+17	89			

Fonte: o autor

Tabela 37 - Coeficientes da regressão do custo corrigido pelos fatores

Variáveis	Coeficientes	Erro padrão	Beta	t valor	Significância
Constante	3,146E+08	4,491 E+06		70,43	0,000
Componente 1	4,172 E+07	4,516 E+06	0,646	9,239	0,000
Componente 2	2,118 E+07	4,516 E+06	0,328	4,690	0,000
Componente 6	1,225 E+07	4,516 E+06	0,190	2,713	0,008
Componente 4	9,991 E+06	4,516 E+06	0,155	2,212	0,030

Fonte: o autor

A regressão múltipla com fatores utilizou o método *Stepwise* e assim selecionou os quatro componentes criados como variáveis independentes do modelo.

O valor de R encontrado foi de 0,765, menor que os 0,846 da regressão múltipla com as variáveis, no entanto, maior que 0,704 encontrado na regressão simples. O R² ajustado obteve o valor de 0,565, um valor intermediário quando se comparado com os outros dois já encontrados, 0,490 para a regressão simples e 0,694 para a múltipla.

A estatística F ficou em 29,901 e a significância em 0,000, menor que os 5% definidos, sendo assim, a regressão é válida.

Analisando os coeficientes individualmente, os valores encontrados foram 3,146E+08 para a constante, com t-valor de 70,430 e significância de 0,000, 4,172 E+7 para o componente 1, com t-valor de 9,239 e significância de 0,000, 2,118 E+07 para o componente 2, com t-valor de 4,690 e significância de 0,000, 1,225 E+07 para o componente 6, com t-valor de 2,713 e significância de 0,008, e 9,991 E+06 para o componente 4, com t-valor de 2,212 e significância de 0,030. Para esses valores, todos os coeficientes foram significativamente diferentes de zero.

Em relação aos Betas, nota-se que seus valores são os mesmos das correlações entre a variável dependente e os fatores. Isso ocorre pelo fato de eles serem totalmente independentes entre si.

Após a regressão múltipla envolvendo os fatores, serão realizados os estudos de diagnóstico. A primeira premissa a ser averiguada é a multicolinearidade.

A Tabela 38 apresenta os valores de Tolerância e de VIF para as variáveis da regressão.

Tabela 38 - Estatísticas de colinearidade da regressão múltipla dos fatores

Variáveis	Tolerância	VIF
Constante		
Componente 1	1,000	1,000
Componente 2	1,000	1,000
Componente 3	1,000	1,000
Componente 4	1,000	1,000

Fonte: o autor

Os valores encontrados para Tolerância e VIF foram iguais a 1, isso ocorreu por que a correlação entre os componentes é nula, logo, não existe o risco de existência de multicolinearidade.

Foram feitos os gráficos de probabilidade normal, o histograma dos resíduos, assim como, o gráfico de dispersão para análise de resíduos.

A Figura 42 testa a normalidade dos resíduos.

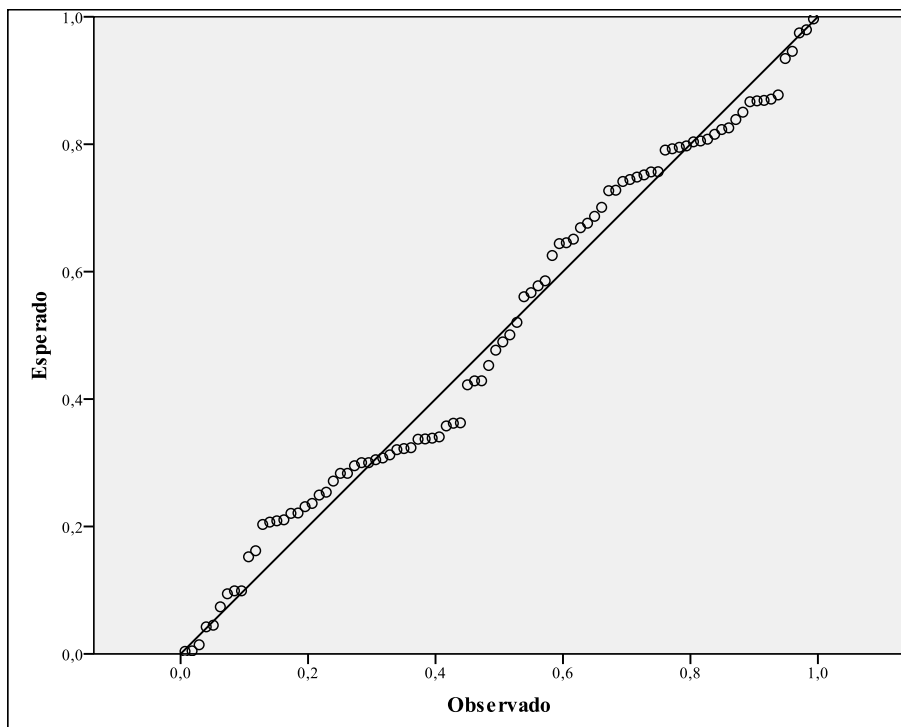


Figura 42 - Gráfico de probabilidade normal: resíduos padronizados da regressão múltipla dos fatores

Como mostrado da Figura 42, os valores estão distribuídos ao longo da diagonal, porém, com certos desvios que se apresentam de forma sistemática. Assim, não é

possível afirmar que os resíduos são representativos de uma distribuição normal, não satisfazendo a suposição de normalidade.

Em virtude dessa incerteza, foi feito o histograma dos resíduos padronizados, conforme apresentado na Figura 43.

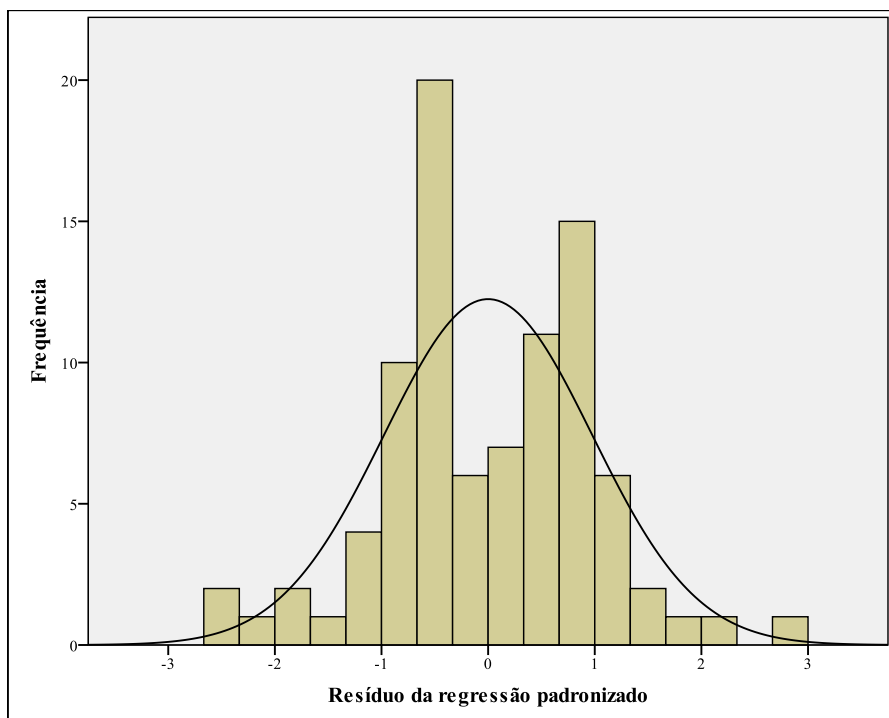


Figura 43 - Histograma dos resíduos da regressão após análise fatorial

Pelo histograma fica evidente que os dados não estão distribuídos em uma distribuição normal, o que prejudica a qualidade dos resultados obtidos. Essa característica é uma desvantagem clara de modelo, quando comparado com a regressão múltipla feita anteriormente.

A Figura 44 apresenta os resíduos em forma de gráfico de dispersão. No eixo das abscissas estão os valores previstos padronizados e no eixo das ordenadas os valores dos resíduos padronizados. Esse gráfico foi feito para se analisar as premissas de linearidade e homocedasticidade, além de se pesquisar a existência de valores anômalos.

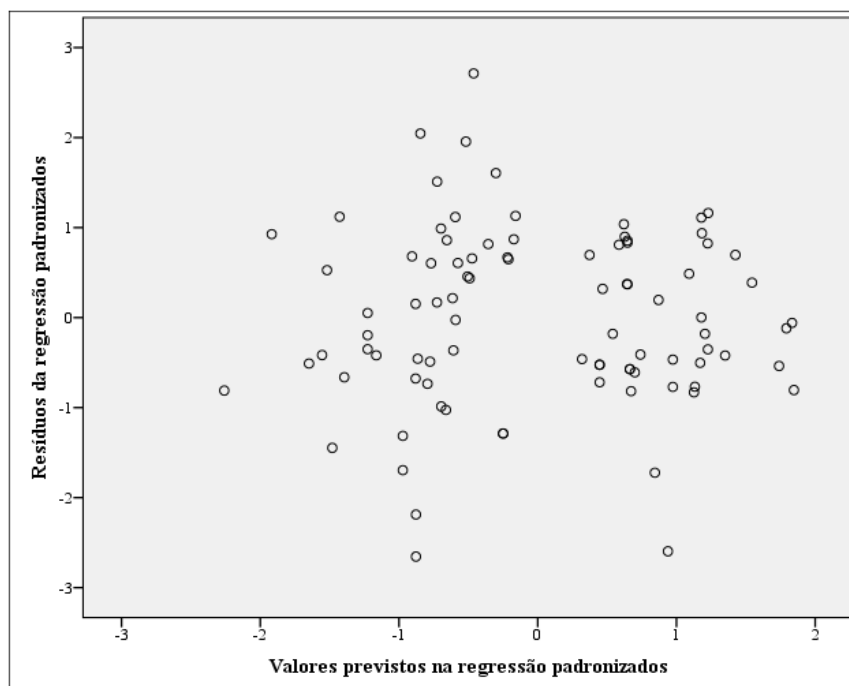


Figura 44 - Gráfico de dispersão dos resíduos padronizados para a regressão múltipla dos fatores

Os resíduos estão distribuídos em dois grupos, num exemplo de dependência por evento. Não há indícios de problemas quanto a linearidade, no entanto, a parte negativa do eixo horizontal aparenta ter uma dispersão maior que o lado positivo, caracterizando assim, a existência de heterocedasticidade.

Não é clara a existência de valores anômalos, é necessário trabalhar nos problemas das outras premissas antes dessa avaliação.

A regressão após a análise fatorial não se mostrou uma boa ferramenta nessa amostra, gerando resultados inferiores à regressão múltipla em relação ao R^2 ajustado, além de apresentar problemas na regressão que diminuem a credibilidade do modelo.

5.2.4. VALIDAÇÃO DO MODELO E COMPARAÇÃO DOS RESULTADOS

Depois de gerados os modelos será feita a validação dos resultados encontrados com 10 dados, retirados da amostra inicial para esse propósito. Os dados foram selecionados ao acaso, apenas evitando que fossem dados anômalos ou dados com baixa qualidade de informações.

Os valores reais e previstos e os resíduos para os dados que foram utilizados no ajuste do modelo se encontram no Apêndice.

Serão comparados os dados usados no modelo em relação à regressão múltipla e a regressão após a análise fatorial, em relação a média do erro absoluto e ao erro

máximo de previsão na amostra em termos percentuais. Erro absoluto, para evitar que os erros se compensem quando positivos e negativos e em percentual, para evitar problemas quanto a ordem de grandeza do preço.

A Tabela 39 apresenta os valores de erro médio e erro máximo para as duas regressões múltiplas.

Tabela 39 - Valores de erro médio e erro máximo para as regressões múltiplas

	Regressão múltipla	Regressão análise fatorial
Erro médio (%)	9,25	11,79
Erro máximo (%)	48,24	71,57

Fonte: o autor

O erro absoluto médio de previsão na amostra utilizada para ajustar o modelo foi de 9,25% para a regressão com as variáveis e de 11,79 para a regressão de fatores.

A diferença do erro máximo foi de 48,24% na regressão com variáveis e 71,57% para a regressão de fatores, uma diferença de mais de 20%.

Tendo apenas a regressão de variáveis como objeto de estudo, foi aferido o poder de previsibilidade do modelo e sua generalidade, ou seja, averiguada a capacidade de prever valores que não foram usados em sua construção.

A Tabela 40 apresenta as sondas que foram selecionadas para se fazer a validação, os valores do custo corrigido, o valor estimado pelo modelo, seus resíduos absolutos e percentuais.

Tabela 40 - Validação da regressão múltipla

Sondas	Custo corrigido (US\$)	Regressão múltipla	Resíduo (US\$)	Resíduo %
Bolshaya Medveditsa	301.566.649,64	297.926.880,00	-3.639.769,64	-1,20
La Muralla IV	330.978.397,79	279.930.160,00	-51.048.237,79	-15,42
Noble Bully I	288.983.555,76	305.661.991,39	16.678.435,64	5,77
Noble Bully II	335.540.423,55	305.661.991,39	-29.878.432,16	-8,90
Noble Jim Day	193.606.514,81	304.593.600,00	46.229.192,39	17,89
ODN I	275.547.295,12	316.422.700,00	40.875.404,88	14,83
ODN II	277.223.528,68	275.270.700,00	-1.952.828,68	-14,140
Petroserv Semi Tbn1	242.659.943,50	242.659.943,50	35.043.592,50	14,441
Scarabeo 8	304.624.258,89	304.624.258,89	-6.779.828,89	-2,226

Fonte: o autor

A média dos erros absolutos encontrados na amostra de validação foi de 9,48%, um valor muito próximo aos 9,25 encontrados na amostra de ajuste. O erro percentual

mais alto foi 17,89%, um valor aceitável dado o grande número de variáveis e características envolvidas em uma sonda marítima. Dessa forma, é possível afirmar que o modelo é generalizado dada a amostra.

6 CONCLUSÕES

O primeiro resultado encontrado foi feito a partir de uma regressão simples, seu objetivo era servir de base para comparação dos resultados seguintes. Foi feita a regressão do custo em relação à capacidade máxima de carga que uma sonda marítima pode suportar dentro dos limites de segurança. A correlação de Pearson, conhecida como linear, dentre as variáveis é de 0,704, a maior entre todas as analisadas, somado ao fato de ser altamente correlacionada com muitas das outras características que determinaram sua escolha. O modelo com apenas uma variável dependente obteve o valor de R^2 ajustado de 0,490, um valor aceitável para uma regressão simples, no entanto, os resíduos apresentaram um comportamento não aleatório e nem homogêneo quanto à dispersão.

O segundo resultado, o que encontrou a melhor posição do ponto de vista de maior adequação, foi alcançado através de uma regressão múltipla. Foram selecionadas entre as 22 variáveis aquelas que mais contribuiriam para o melhor ajuste do modelo. Métodos diferentes foram testados ,chegando-se ao melhor resultado com o *Stepwise*.

O modelo com melhor resultado foi obtido com 6 variáveis independentes, capacidade de teste de poço extensivo (EWT), se foi feita no sudeste asiático, se a sonda está habilitada para operar na Noruega, se a sonda está preparada para ambientes hostis, a capacidade máxima de carga suportada pela torre de perfuração e a capacidade do alojamento.

Os efeitos moderadores não foram usados no melhor resultado, assim, apesar de influenciar outras características, o tipo de sonda acaba sendo explicado por outras variáveis, ao ponto de nem mesmo ser incluído no modelo escolhido.

O valor de R^2 ajustado foi de 0,694 para o melhor modelo e a variável que mais influência o modelo é a capacidade do alojamento, com o coeficiente padronizado Beta, de 0,481, seguido pela habilitação de operar na Noruega, com coeficiente padronizado de 0,470.

A análise fatorial foi realizada com o objetivo de encontrar os componentes que em número reduzido representassem a maior parte da variância das variáveis independentes. O método dos componentes principais foi utilizado com os mesmos

dados de entrada que a regressão anterior para que os dois modelos pudessem ser comparados.

A análise fatorial encontrou 7 fatores que continham cerca de 76% de toda a variância encontrada nas 22 variáveis.

Uma regressão múltipla foi feita com esses componentes chegando ao resultado de R^2 ajustado de 0,565, melhor que da regressão simples, porém, pior que o da regressão múltipla com as variáveis originais. Além do pior ajuste, a regressão com fatores, apresentou ainda, problemas de não normalidade dos resíduos e de um padrão em sua distribuição, o que viola as premissas da regressão múltipla.

A regressão múltipla com variáveis, encontrou um erro médio absoluto de previsão de 9,25% para os dados usados no ajuste e de 9,48% para os dados usados na validação, um bom resultado em ambos os casos e ainda corroborando a eficiência do modelo em prever dados externos a amostra de ajuste.

Mesmo condensando as variâncias das diversas variáveis os fatores não necessariamente geram melhores previsões de variáveis dependentes. Isso poderia ter sido deduzido após a descoberta da baixa correlação entre eles e o custo corrigido da sonda.

Para trabalho futuros seria interessante utilizar outros métodos paramétricos de resolução, como redes neurais, algoritmos genéticos e outros tipos de metaheurísticas. Seguindo a vertente de análise multivariada, seria interessante o uso de modelos de equações estruturadas, pois eles poderiam validar ideias conjuntas de quais são os conceitos mais significativos para o custo de uma sonda marítima.

7 REFERÊNCIA BIBLIOGRÁFICA

AACE, International Recommended Practice, **Basis of Estimate**, 2007.

_____, **Cost Engineering Terminology**, 2009.

_____, **Cost Estimate Classification System**, 1998.

AGÊNCIA NACIONAL DE PETRÓLEO E GÁS NATURAL E BIOCMBUSTÍVEIS (ANP),2012. Disponível em: <<http://www.anp.gov.br>>. Acesso em: 30 jun. 2012, 21:00.

BUXTON, I. L., CROZIER, P, GUENOV, M.,”Concurrent conceptual design and cost estimating”, **Transactions of 13th International Cost Engineering**, Congress, London, 1994.

BUREAU OF OCEAN ENERGY MANAGEMENT, REGULATION AND ENFORCEMENT, 2012, disponível em <<http://www.boemre.gov/tarprojectcategories/structur.htm>>. Acesso em 10 jun. 2012, 22:00.

D’ALMEIDA, A. L., **Estruturação e Dimensionamento de Frota e Pessoal numa Empresa de Sondagem e Serviços Especiais em Petróleo**. Tese (Doutorado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.

FOUSSIER, P., **From Product Description to Cost: A Practical Approach**. 1 ed. London, Springer-Verlag, 2006.

DESCHAMPS, L. C.; TRUMBULE, J., **Ship Design & Construction**. The Society of Naval Architects and Marines Engineers. 1 ed. New Jersey, 2004.

FONSECA, M. M., **Arte Naval**. 5 ed. Serviço de Documentação Geral da Marinha, 1989.

HAIR, J. F., BLACK, W. C. BABIN, B. J., *et al.*, **Análise Multivariada de Dados**. 6 ed. Porto Alegre, Bookman, 2009.

HALLIBURTON, **Petroleum Well Construction**, 2007.

HOLLMANN, J. K., **The Engineer's Cost Handbook**, Marcel Decker.Inc, 1997.

INSTITUTO BRASILEIRO DE ENGENHARIA DE CUSTOS, **Orientação Técnica N.01/2009/IBEC**, 2009.

INTERNATIONAL SOCIETY OF PARAMETRIC ANALYSTS (ISPA), **Parametric Estimating Handbook**. 4 ed., 2008.

KAISER, M. J., SNYDER, B. F., “Labor, material needs estimated for US construction of jack ups”. **Oil & Gas Journal**. Louisiana, 2012.

LIMA, C. J. T., **Processo de Tomada de Decisão em Projetos de Exploração e Produção de Petróleo no Brasil: Uma Abordagem Utilizando Conjuntos**. Dissertação M.Sc., COPPE, Rio de Janeiro, RJ, Brasil, 2003.

LEAMON, G. R., **Petroleum Wells Costs**. Dissertação M.Sc., The University of New South Wales, Sydney, Austrália, 2006.

MEILS, C. J., Techniques for cost estimating in early program phases. **Engineering Costs and Production Economics**, California, 1988.

MILEHAM, A.R., CURRIE, G.C., MILES, A.W., *et al.*, “A Parametric Approach to Cost Estimation at the Conceptual Stage of Design”, **Journal of Engineering Design**, v. 4, n. 2, 1993.

MARINEINSIGHT, 2012. Disponível em < www.marineinsight.com >. Acesso em 10 abr. 2012, 20:00.

NASA, **Cost Estimating Handbook**. 2008, Disponível em <www.jsc.nasa.gov/bu2/NCEH/index.htm>. Acesso em 01 fev. 2012, 20:00.

_____, **JSC cost estimating and models web site**. 2010. Disponível em <<http://www.jsc.nasa.gov/bu2/>>, acessado em 01 fev. 2012, 20:00.

MILES, M.; HUBERMAN, M., **Qualitative data analysis**. Thousand Oaks, Sage Publications, 1994.

PETROBRAS, 2012. Disponível em: <<http://www.petrobras.com.br>>. Acesso em: 5 jun. 2012.

PETROINFONET, 2012. Disponível em: <<http://petroleoinfonet.blogspot.com.br>>. Acesso em 30 jun. 2012, 18:00.

PORTAL MARINHO, 2012. Disponível em: <<http://portalmaritimo.com>>. Acesso em 18 abr. 2012, 13:00.

PUGH, P., Working top-down: Cost estimating before development deigns. **Journal of Aerospace Engineering**, Part G, n. 206, pp. 143-151, 1992.

RIGZONE, 2012. Disponível em: < <http://www.rigzone.com>>. Acessado em 10 ago. 2012.

RIBAS, J. R.; VIEIRA, P. R. C., **Análise multivariada com o uso do SPSS**. 1 ed. Rio de Janeiro, Ciência Moderna, 2011.

STEWART, R., WYSKIDSA, R., JOHANNES, J., **Cost Estimator's Reference Manual**. 2 ed. Wiley Interscience, 1995.

TAYLOR, I. M., “Cost engineering - A feature based approach”. **85th Meeting of the AGARD Structures and Material Panel**. Aalborg (Denmark), pp. 1-9, 1997.

THOMAS, J. E., **Fundamentos de engenharia de Petróleo**. 2 ed. Rio de Janeiro, Interciência, 2004.

US BUREAU OF LABOR STATISTICS, 2011. Disponível em <http://www.bls.gov/>. Acessado em 10 jan. 2011, 15:30.

8 APÊNDICE

Em complemento ao resultado da pesquisa, neste apêndice serão desenvolvidas as análises das variáveis independentes quantitativas e qualitativas, por agregar, numa visão resumida, uma melhor compreensão dos dados assim como suas relações entre si, constituindo, assim, um suporte técnico para que as análises principais fossem realizadas.

A Tabela 41 apresenta as variáveis que foram utilizadas nos estudos.

Tabela 41 - Lista de variáveis independentes selecionadas

Variáveis Qualitativas	Variáveis Quantitativas
Tipo de Sonda	Profundidade da lâmina d'água (pés)
Se foi construída de forma especulativa	Profundidade de perfuração (pés)
Se apresenta posicionamento dinâmico	Carga máxima suportada pela sonda (ton)
Se é capaz de realizar dupla atividade	Capacidade de carga variável da sonda (ton)
Se está adaptado à ambiente hostil	Comprimento do casco (pés)
Se possui top drive	Largura do casco (pés)
Em que estágio da construção que a sonda se encontra	Capacidade do alojamento (pessoas)
Capacidade de teste extensivos de poços (EWT)	Capacidade de água de perfuração (BBL)
Se está adaptado ao mar do norte	Capacidade de água potável (BBL)
Se possui a capacidade de zero desperdício de fluidos	Carga máxima suportada pela torre (libras)
Se está habilitado para a atividade na Noruega	
Local de construção	

Fonte: o autor

8.1. ANÁLISE COM VARIÁVEL

8.1.1. ANÁLISE DAS VARIÁVEIS INDEPENDENTES QUANTITATIVAS

As variáveis quantitativas são de mais fácil compreensão. Serão apresentadas tabelas de dados com informações relevantes.

Na tabela 41 são apresentadas as estatísticas descritivas dessas variáveis.

Tabela 42 - Estatísticas descritivas variáveis independentes quantitativas

Variável	Média	Desvio-Padrão	Assimetria	Curtose
Profundidade da lâmina d'água	8.937,78	2.110,99	-2,12	4,74
Profundidade de perfuração	35.219,59	5.671,03	-1,50	4,13
Carga máxima da sonda	74.186,57	27.697,78	0,07	-1,66
Variação de carga suportável	14.264,42	6.988,95	0,08	-1,86
Comprimento do casco	522,99	219,86	0,16	-1,85
Largura do casco	199,42	64,31	0,75	-1,57
Capacidade do alojamento	175,94	27,53	-0,45	-0,76
Água de perfuração	16.257,33	3.693,44	0,90	2,97
Água potável	7.473,19	5.833,00	3,53	17,24

Fonte: o autor

Pelos dados apresentados na Tabela 42, é possível notar que aparentemente os dados não seguem uma distribuição normal em nenhum dos casos. Para corroborar isso será feito o teste de normalidade de Kolmogorov-Smirnov para todas as variáveis independentes quantitativas.

A Tabela 43 apresenta os resultados encontrados.

Tabela 43 - Teste de normalidade Kolmogorov-Smirnov para variáveis independentes quantitativas

Variável	Kolmogorov-Smirnov	Significância
Profundidade da lâmina d'água	0,337	0,000
Profundidade de perfuração	0,251	0,000
Carga máxima da sonda	0,220	0,000
Variação de carga suportável	0,243	0,000
Comprimento do casco	0,280	0,000
Largura do casco	0,286	0,000
Capacidade do alojamento	0,170	0,000
Água de perfuração	0,135	0,000
Água potável	0,189	0,000

Fonte: o autor

Com os resultados encontrados na Tabela 43, é notório que nenhuma das variáveis é significativamente uma normal. Isso é um ponto de atenção que deve ser considerado na validação e avaliação dos resultados encontrados.

Quatro variáveis quantitativas, que pela teoria das sondas seriam de grande relevância, foram estudadas mais a fundo, sendo feitos histogramas e gráficos Boxplot para um melhor entendimento das mesmas.

Foram escolhidas as características: profundidade da lâmina d'água, a profundidade máxima de perfuração, a carga máxima que a sonda pode suportar e a variação de carga suportável pela sonda.

A Figura 45 apresenta o histograma da profundidade da lâmina d'água.

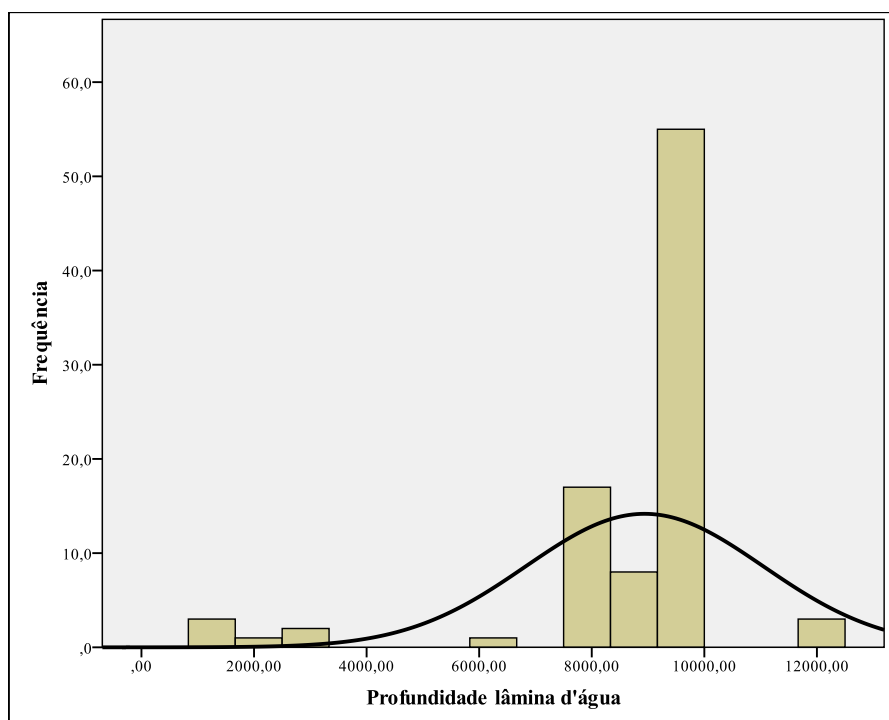


Figura 45 - histograma da profundidade da lâmina d'água

Pelo histograma é nítido que se não trata de uma distribuição normal, no entanto, nesse momento nenhuma transformação foi feita, essas só ocorrem caso seja necessário após a análise em relação a variável dependente. Os dados estão muito dispersos, e com uma concentração muito grande em 2 classes, além de apresentarem valores muito discrepantes, com grandes saltos de capacidade entre as diferentes classes. Uma retirada desses valores extremos poderia melhorar consideravelmente o comportamento da amostra.

A Figura 46 apresenta o Boxplot e analisa a existência de valores extremos na amostra da variável.

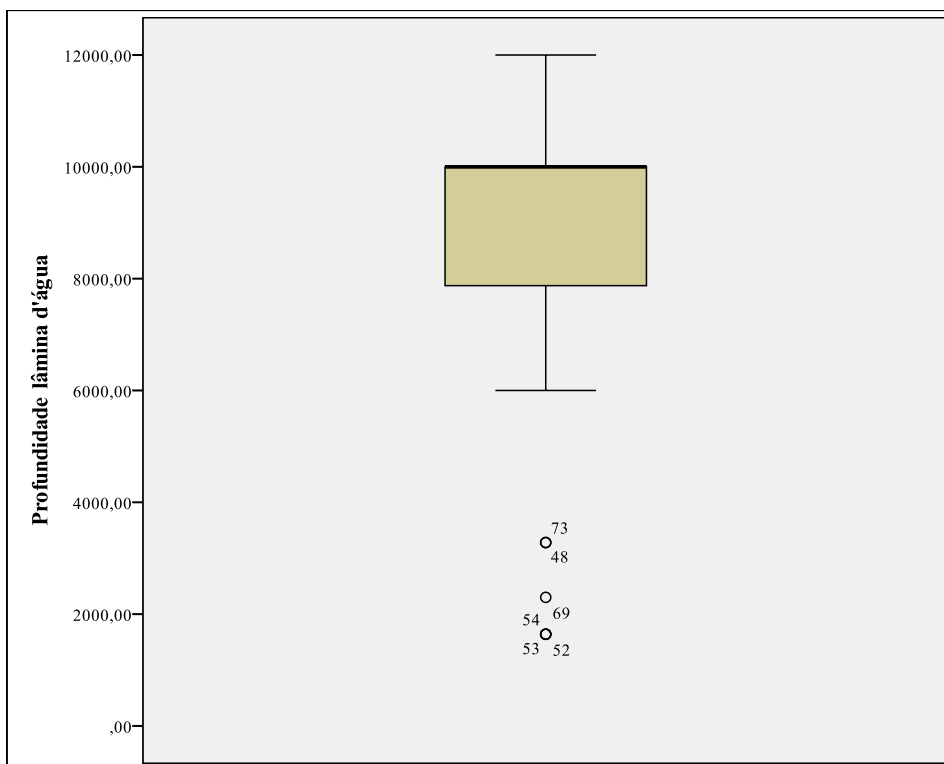


Figura 46 - Boxplot da profundidade da lâmina d'água

Como visto na Figura 46 existem quatro valores extremos, contudo assim como a questão da normalidade não será feito nenhum tratamento nesse momento.

A segunda variável a ser estudada é a carga máxima suportada pela sonda, do termo em inglês *displacement* é a carga máxima suportada pelo casco da sonda dentro dos limites de segurança. Nas Figuras 47 e 48 são apresentados respectivamente seu histograma e o Boxplot para analisar a distribuição das características assim como a existência de dados anômalos.

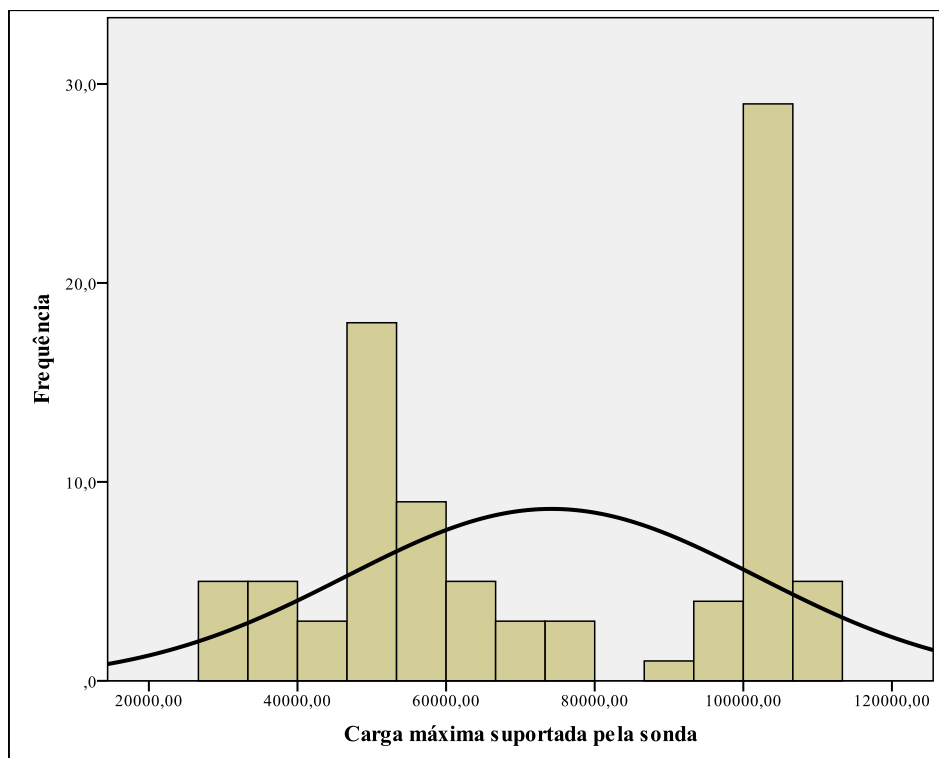


Figura 47 - Histograma carga máxima suportada pela sonda

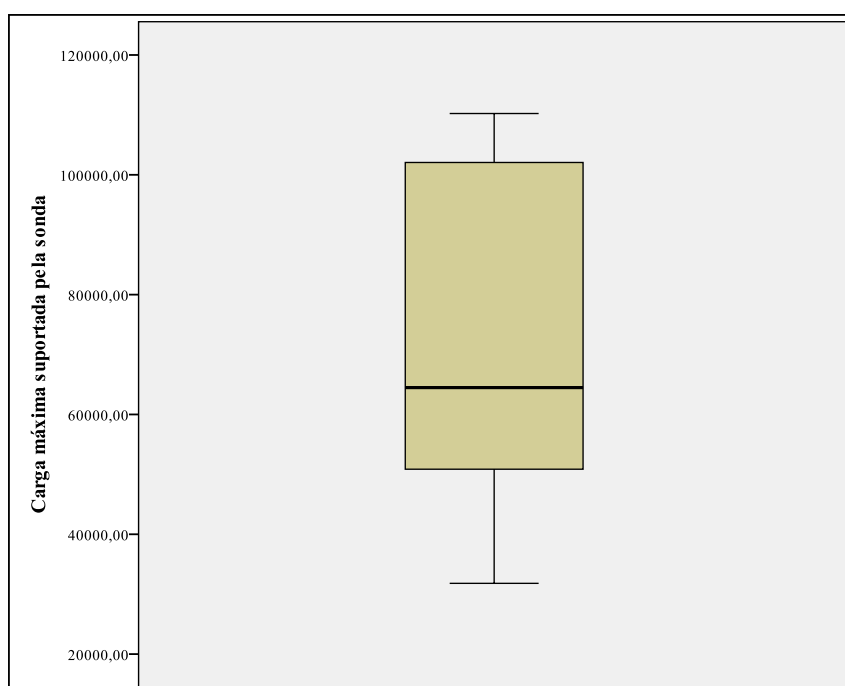


Figura 48 - Boxplot carga máxima suportada pela sonda

Como pode ser visto na Figura 47, a variável carga máxima não segue uma distribuição normal, corroborando o que já se constatou pelo teste de normalidade feito anteriormente. Os valores se concentram em duas faixas claras e não seguem um comportamento contínuo. Aparentemente podem ser divididos em dois grupos, o que

levanta a hipótese de que uma outra característica possa estar influenciando o seu comportamento.

Na Figura 48, é notada a inexistência de valores extremos, e nenhum tratamento é necessário nesse caso. No entanto o teste deve ser feito caso seja feita alguma transformação para buscar a normalidade dos dados.

A terceira variável quantitativa estudada é a variação de carga suportável, ou seja, é capacidade de carga variável que a sonda pode ser submetida. A Figura 49 apresenta seu histograma.

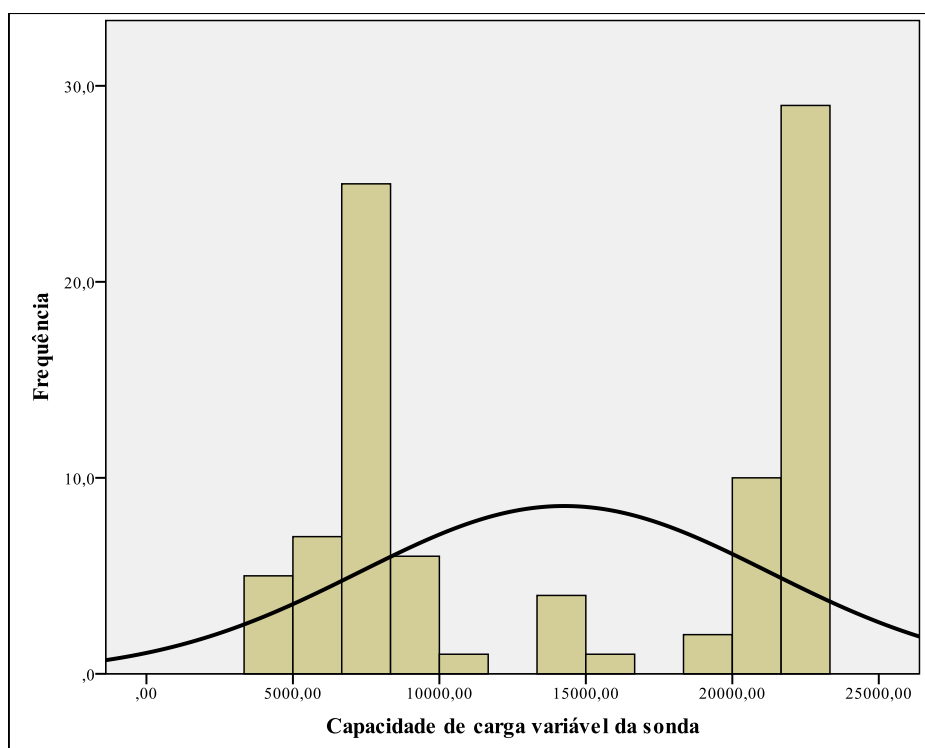


Figura 49 - Histograma da capacidade de carga variável da sonda

Como visto na Figura 49 a relação também não se aproxima de uma normal, existem dois grupos bem definidos e um pequeno grupo intermediário. A amplitude de valores é alta também, isto leva a crer que exista alguma característica qualitativa que influencia fortemente seu comportamento. Na Figura 50 é apresentado o Boxplot.

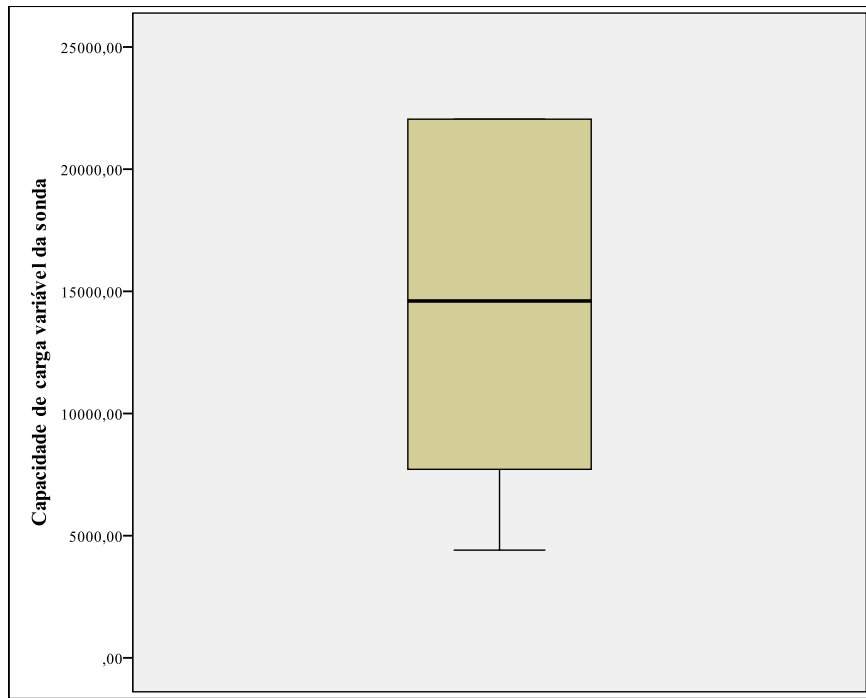


Figura 50 - Boxplot da capacidade de carga variável da sonda

Nenhum valor anômalo foi encontrado, no entanto esse teste deve ser feito caso haja uma divisão da amostra de acordo com outra característica.

Por último será analisada a capacidade de perfuração, em outras palavras, é a profundidade máxima que os equipamentos da sonda que realizam a perfuração podem alcançar. A Figura 51 apresenta seu histograma e a Figura 52 o Boxplot relativo aos seus valores.

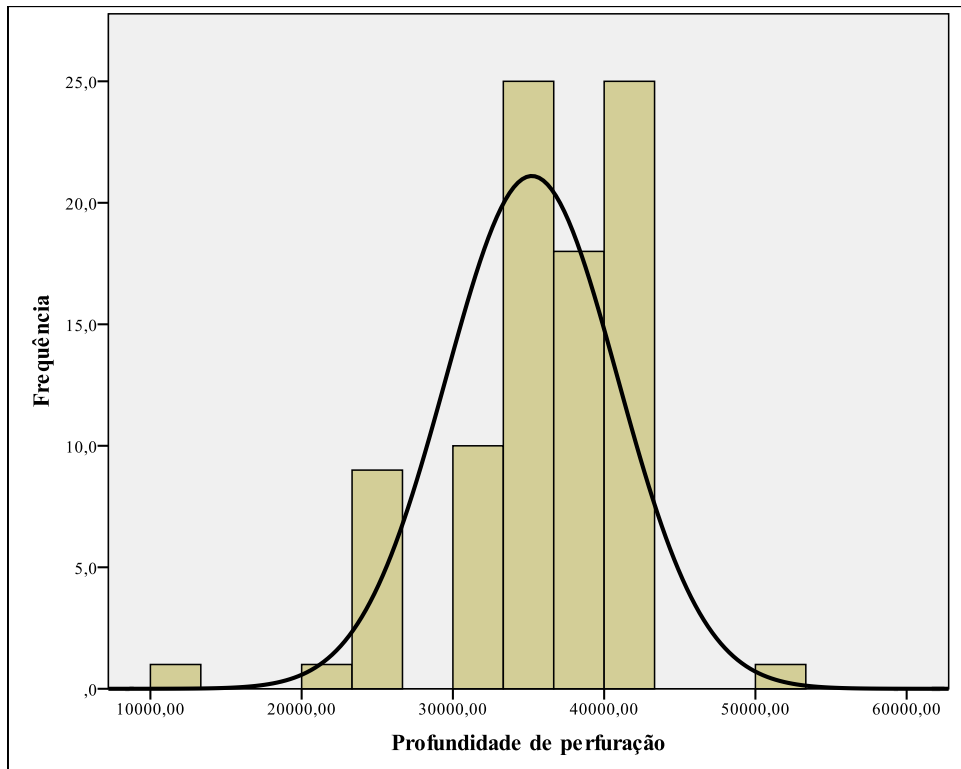


Figura 51 - Histograma da profundidade de perfuração

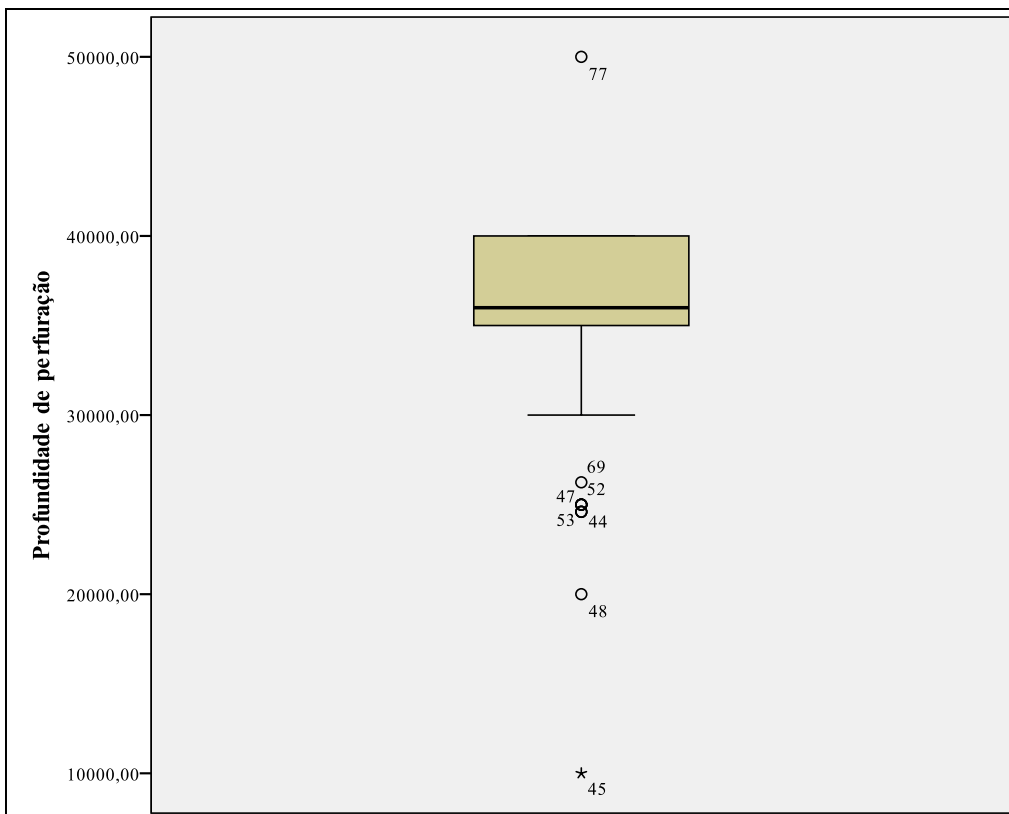


Figura 52 - Boxplot da profundidade de perfuração

Na Figura 51 reflete a não normalidade dos dados, no entanto é possível ver que a presença de dado anômalo, o que pode ser corroborado pela Figura 52. Com a retirada

desses é possível que a amostra tenha um comportamento mais interessante em relação a anormalidade.

8.1.2. ANÁLISE DAS VARIÁVEIS INDEPENDENTES QUALITATIVAS

Das 12 características qualitativas das sondas escolhidas para o estudo, podemos destacar oito delas para serem estudadas de forma individual, seu tipo, o estágio de construção da sonda, se foi feita de forma especulativa, ou seja, sem um cliente específico, região onde foi feita, a existência de posicionamento dinâmico, se foi feita para ambientes hostis, se possui um Top Drive e se pode realizar duas atividades ao mesmo tempo por possuir duas torres de perfuração.

As sondas selecionadas são de apenas dois tipos, navios-sondas e sonda semissubmersíveis. A Figura 53 apresenta um gráfico que determina sua proporção.

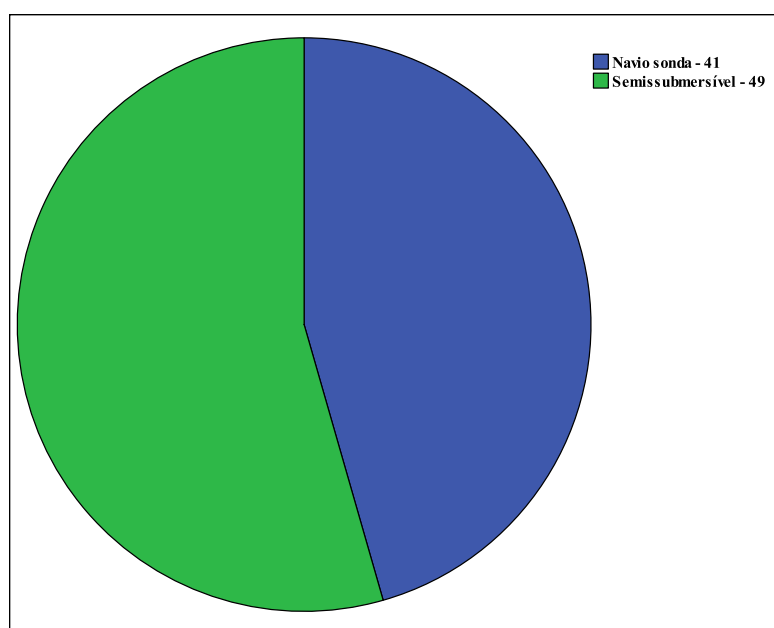


Figura 53 - Proporção entre os tipos de sondas

Das 90 observações, 49 são de sondas semissubmersíveis o que representa 54,4%, e 41 são de navios-sondas que representam 45,6%. Nesse quesito a amostra se mostrou bem equilibrada.

Em relação ao estado em que se encontra a construção da sonda, foram definidos três possíveis estágios: entregue, em construção e pedido feito.

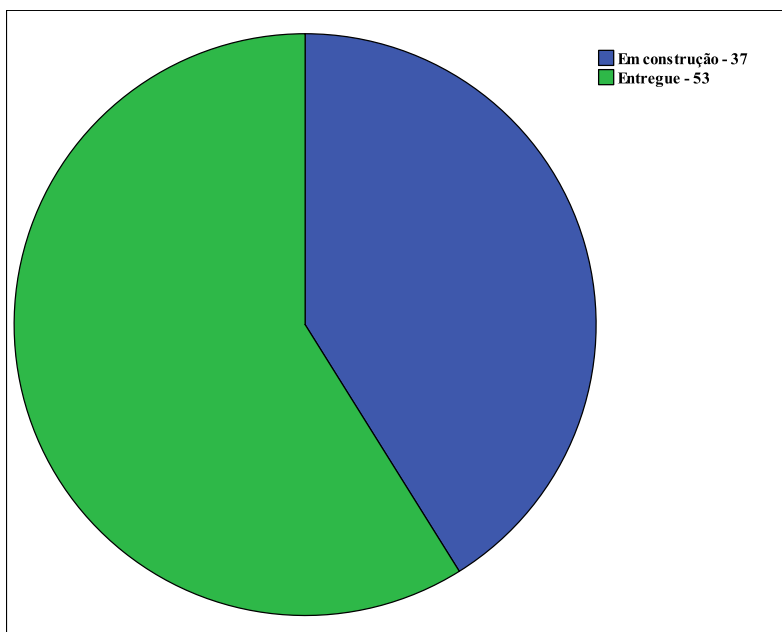


Figura 54 - Estado da construção das sondas

A Figura 54 mostra a proporção entre os diferentes estágios. Na amostra predomina o número de sondas já entregues com 53 casos dentre os 90 selecionados, gerando assim 58,9%, enquanto as em construção ou com o pedido já realizado mas sem que fosse iniciada a construção detêm 37 casos, ou seja, 41,1% deles.

Mais uma característica que pode ser de grande valia na construção do modelo é se a sonda foi feita de forma especulativa ou se já havia um contrato de utilização quando ela foi demandada. A Figura 55 mostra a proporção encontrada na amostra estudada.

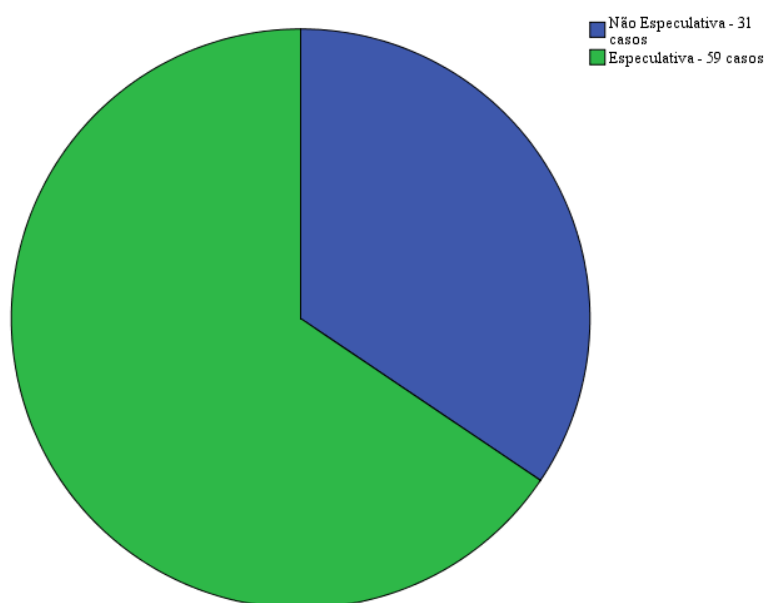


Figura 55 - Proporção entre sondas feitas de forma especulativa ou não

Como pode ser observado na Figura 55, a maioria das sondas foram feitas de forma especulativa, totalizando 59 casos, enquanto apenas 31 foram feitas já com um contrato firmado. Essa informação é importante, visto que reflete o otimismo do mercado de compradores de sondas que mesmo sem um cliente certo assumiram o risco de um investimento tão alto e tão arriscado.

A região do mundo onde foi contratada a construção da sonda é uma característica relevante, pois pode mostrar diferenças entre a produtividade e o custo de regiões e, por conseguinte países diferentes no tópico de construção de sondas. Na Figura 56 os mercados produtores foram divididos em cinco grupos, o Mar Cáspio (Caspian), Oriente (Far East), Oriente médio (Middle East), Noroeste europeu (NW Europe) e Sudeste Asiático (SE Ásia).

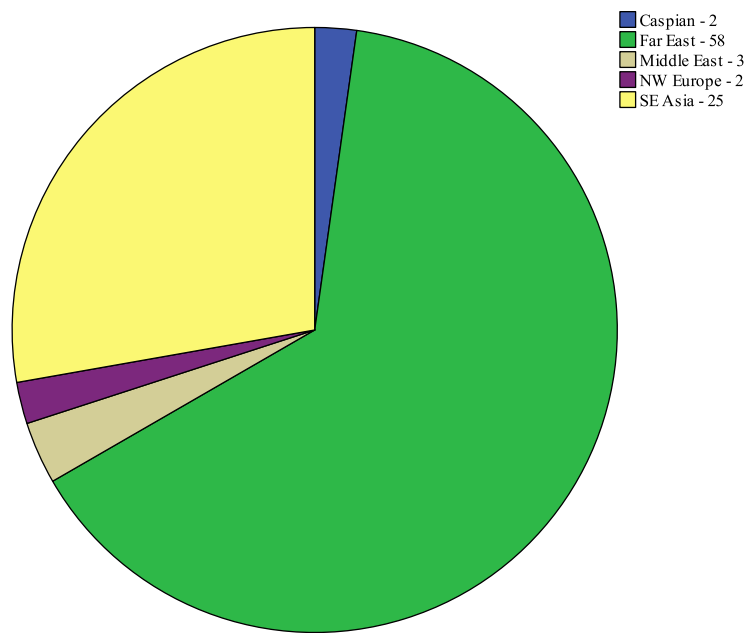


Figura 56 - Distribuição das sondas pelo local de construção

Pela Figura 56 é possível notar a maior concentração nas regiões Oriente e Sudeste Asiático. O oriente na amostra compreende os países Coreia do Sul e China, e obteve 58 das observações enquanto o Sudeste Asiático envolve apenas a Cingapura e obteve 25 casos. Como a amostra foi retirada de forma aleatória é possível ver o peso dessas duas regiões nesse mercado.

O posicionamento dinâmico é o substituto natural para âncoras quando se opera em lâminas d'água profundas. A Figura 57 apresenta a distribuição de sondas que possuem ou não esse equipamento.

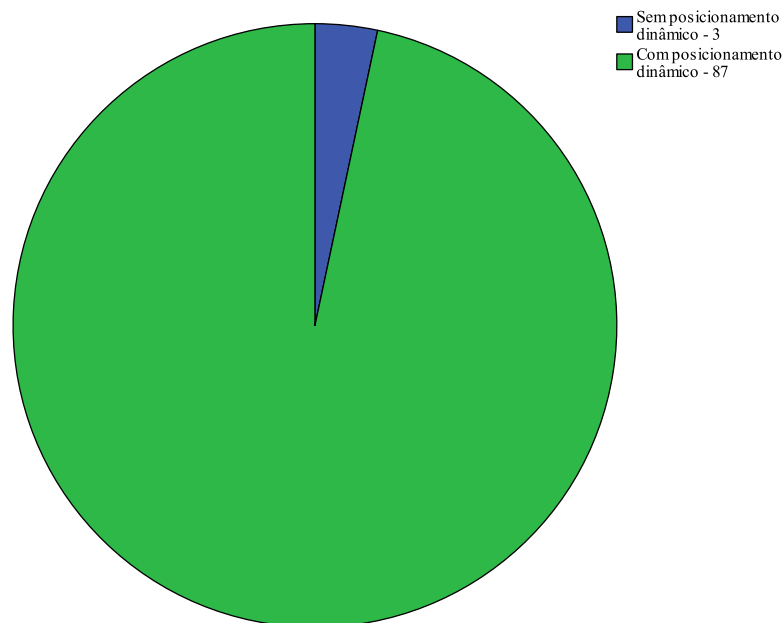


Figura 57 - Distribuição das sondas por possuir posicionamento dinâmico ou não

O resultado encontrado é de 87 sondas com posicionamento dinâmico e 3 sem, o que era esperado, pois a amostra de dados se refere a sondas destinadas a exploração em águas mais profundas. Essa concentração pode inviabilizar a utilização dessa variável no modelo final, pois não haverá significância na comparação entre os resultados encontrados com ou sem essa característica.

A próxima característica a ser analisada é quantas sondas foram preparadas para atuação em locais onde o meio ambiente é agressivo, como locais muito frios e com outros problemas climáticos e ambientais. A Figura 24 apresenta a composição entre as sondas que possuem ou não essa característica.

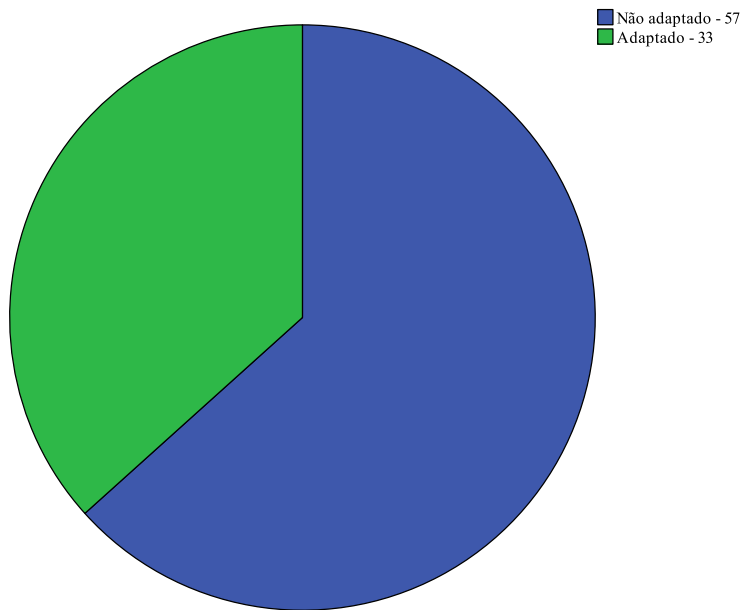


Figura 58 - Sondas adaptados a um ambiente hostil

Como mostrado na Figura 58, apesar de haver um certo equilíbrio, a maioria das sondas estudadas são não adaptadas a ambientes mais agressivos, no entanto essa equidade de observações pode ser útil ao se estudar a existência de uma diferença significativa entre o custo de uma sonda com ou sem esse artifício.

Outra característica importante na composição do custo de uma sonda, é a existência de um Top Drive, equipamento que substitui o *swivel*, o *Kelly* e a mesa rotativa, fazendo a conexão entre a parte fixa e a rotativa. Dá assim o torque à coluna e permite a entrada da mangueira de injeção de fluido de perfuração. Com o Top Drive, é possível que três seções de *drillpipe* sejam encaixadas na coluna de uma só vez, otimizando o processo de perfuração, uma vez que menos paralisações são necessárias.

A Figura 59 representa a quantidade de sondas que possuem Top Drive.

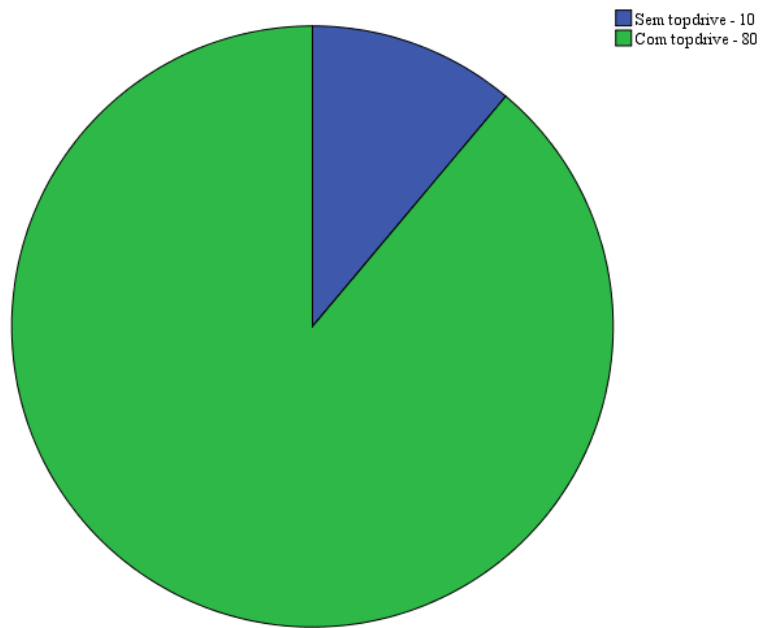


Figura 59 - Sondas com a presença de Top Drive

Na amostra selecionada, há uma quantidade significativamente maior de sondas que possuem Top Drive em relação às aquelas que não apresentam o equipamento. Dessa forma, apesar de ser uma característica relevante no aspecto conceitual para uma diferenciação de preço, existe a possibilidade de essa não ser captada pelos testes estatísticos.

Por último será analisada a presença de dupla atividade. A dupla atividade representa a presença de duas torres de perfuração, o que permite a sonda fazer operações como conexões de dutos em uma enquanto perfura com a outra, acelerando o processo e, assim, reduzindo os custos operacionais.

Na Figura 60 estão representadas as quantidades de sondas da amostra com a presença de duas torres.

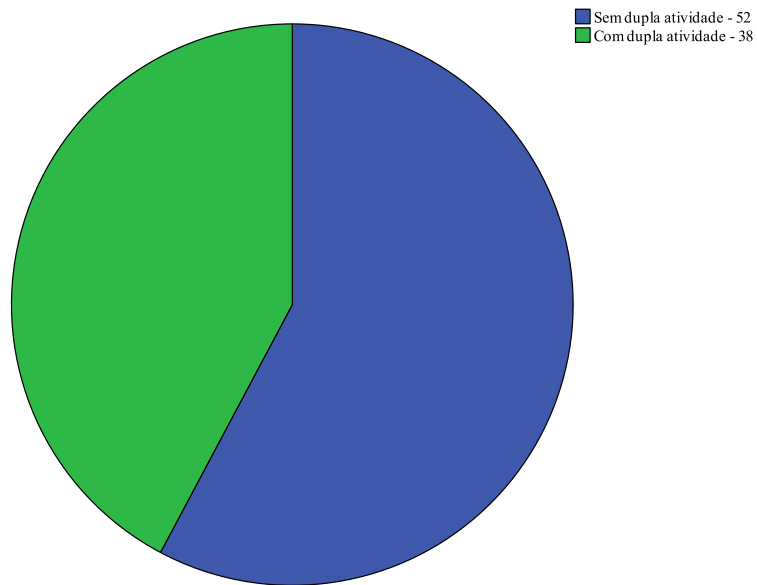


Figura 60 - Sondas com dupla atividade

Apesar de apresentar uma maior quantidade de sondas sem dupla atividade, as quantidades ficaram equilibradas, assim os estudos estatísticos tem maior chance de captar uma possível diferença entre o custo de uma sonda com ou sem essa característica.

8.2. ANÁLISE COM DUAS VARIÁVEIS

O gráfico 61 apresenta o Gráfico de dispersão de variáveis independentes quantitativas entre si.

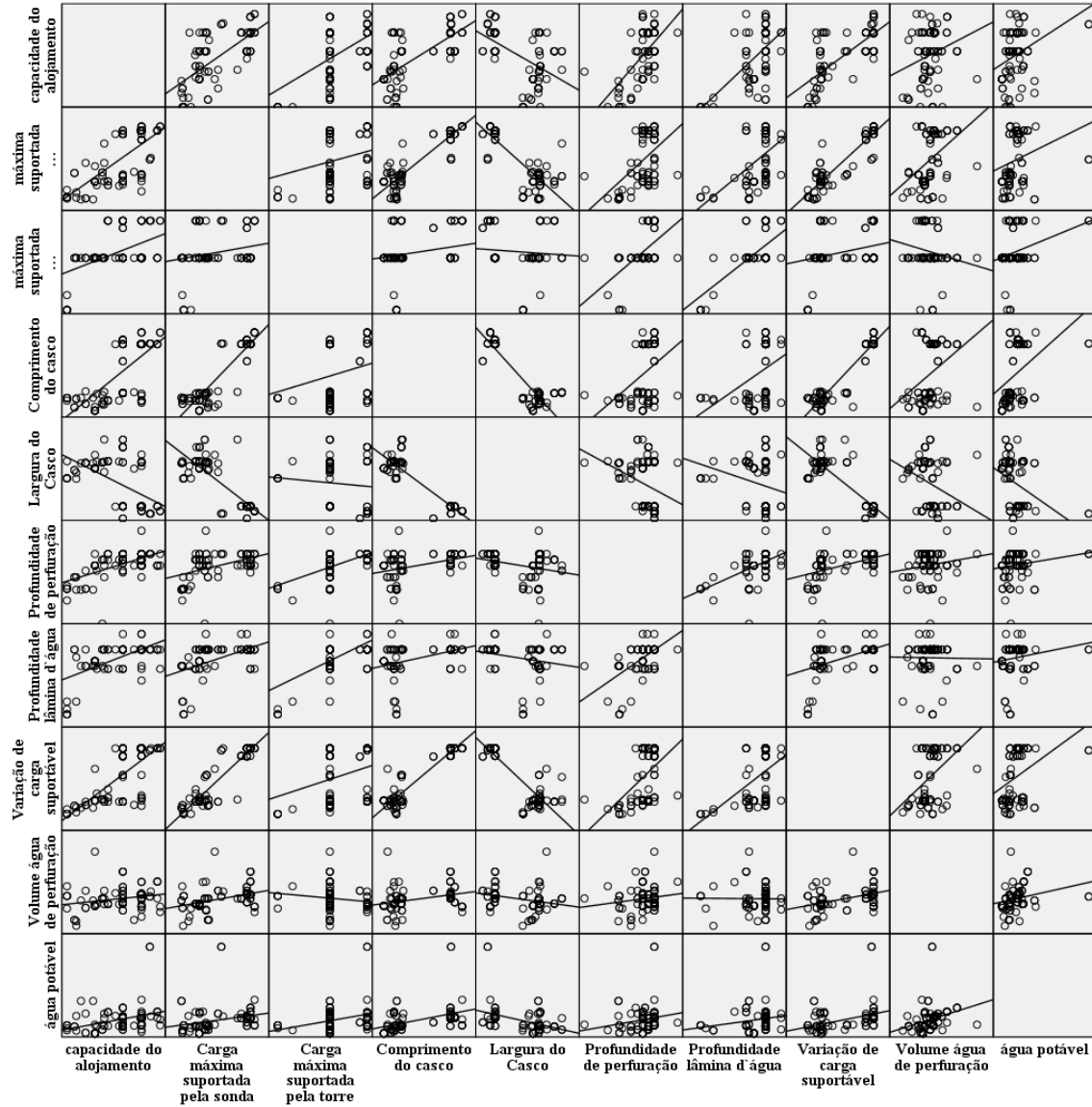


Figura 61 - Gráfico de dispersão variáveis independentes quantitativas

Assim é possível se analisar todas as relações lineares existentes entre essas variáveis.

8.3. ANÁLISE DAS REGRESSÕES MÚLTIPLAS

A Tabela 44 apresenta os valores encontrado para as previsões feita a partir das regressões múltiplas tanto para a que utilizou variáveis quanto para que utilizou os fatores encontrados na análise fatorial.

Tabela 44 - Resultado da previsão das regressões múltiplas

	Regressão Múltipla com Variáveis				Regressão Múltipla com Fatores		
	Custo corrigido	Valor previsto	Resíduo	%	Valor previsto	Resíduo	%
1	\$3,626E+08	\$3,356E+08	\$2,704E+07	7,46%	\$3,330E+08	\$2,961E+07	8,17%
2	\$3,775E+08	\$3,917E+08	-\$1,411E+07	-3,74%	\$4,005E+08	-\$2,291E+07	-6,07%
3	\$3,780E+08	\$3,594E+08	\$1,859E+07	4,92%	\$3,435E+08	\$3,448E+07	9,12%
4	\$2,828E+08	\$3,170E+08	-\$3,416E+07	-12,08%	\$3,562E+08	-\$7,341E+07	-25,96%
5	\$3,060E+08	\$3,356E+08	-\$2,960E+07	-9,67%	\$3,366E+08	-\$3,062E+07	-10,01%
6	\$3,130E+08	\$3,594E+08	-\$4,643E+07	-14,83%	\$3,478E+08	-\$3,483E+07	-11,13%
7	\$3,838E+08	\$3,356E+08	\$4,825E+07	12,57%	\$3,455E+08	\$3,828E+07	9,97%
8	\$3,659E+08	\$3,356E+08	\$3,034E+07	8,29%	\$3,575E+08	\$8,361E+06	2,28%
9	\$3,513E+08	\$3,356E+08	\$1,570E+07	4,47%	\$3,377E+08	\$1,361E+07	3,87%
10	\$3,230E+08	\$3,587E+08	-\$3,574E+07	-11,07%	\$3,474E+08	-\$2,440E+07	-7,56%
11	\$3,230E+08	\$3,587E+08	-\$3,574E+07	-11,07%	\$3,474E+08	-\$2,440E+07	-7,56%
12	\$4,074E+08	\$4,029E+08	\$4,412E+06	1,08%	\$3,908E+08	\$1,653E+07	4,06%
13	\$4,129E+08	\$3,471E+08	\$6,584E+07	15,94%	\$3,730E+08	\$3,993E+07	9,67%
14	\$4,102E+08	\$3,787E+08	\$3,149E+07	7,68%	\$3,750E+08	\$3,512E+07	8,56%
15	\$3,715E+08	\$3,917E+08	-\$2,019E+07	-5,44%	\$4,058E+08	-\$3,429E+07	-9,23%
16	\$4,026E+08	\$4,142E+08	-\$1,162E+07	-2,89%	\$4,051E+08	-\$2,489E+06	-0,62%
17	\$3,980E+08	\$3,917E+08	\$6,341E+06	1,59%	\$4,031E+08	-\$5,064E+06	-1,27%
18	\$3,633E+08	\$3,711E+08	-\$7,763E+06	-2,14%	\$3,813E+08	-\$1,793E+07	-4,93%
19	\$3,894E+08	\$3,344E+08	\$5,498E+07	14,12%	\$3,452E+08	\$4,421E+07	11,35%
20	\$2,503E+08	\$3,170E+08	-\$6,670E+07	-26,65%	\$3,609E+08	-\$1,106E+08	-44,19%
21	\$3,336E+08	\$3,130E+08	\$2,060E+07	6,17%	\$3,511E+08	-\$1,750E+07	-5,25%
22	\$3,336E+08	\$3,130E+08	\$2,060E+07	6,17%	\$3,413E+08	-\$7,660E+06	-2,30%
23	\$3,622E+08	\$3,547E+08	\$7,474E+06	2,06%	\$3,463E+08	\$1,588E+07	4,38%
24	\$3,829E+08	\$3,547E+08	\$2,817E+07	7,36%	\$3,465E+08	\$3,642E+07	9,51%
25	\$3,622E+08	\$3,547E+08	\$7,474E+06	2,06%	\$3,465E+08	\$1,572E+07	4,34%
26	\$3,819E+08	\$3,547E+08	\$2,721E+07	7,12%	\$3,465E+08	\$3,546E+07	9,28%
27	\$3,232E+08	\$3,356E+08	-\$1,239E+07	-3,83%	\$3,491E+08	-\$2,595E+07	-8,03%
28	\$3,143E+08	\$3,356E+08	-\$2,130E+07	-6,78%	\$3,366E+08	-\$2,232E+07	-7,10%
29	\$3,143E+08	\$3,356E+08	-\$2,130E+07	-6,78%	\$3,366E+08	-\$2,232E+07	-7,10%
30	\$4,146E+08	\$3,775E+08	\$3,704E+07	8,93%	\$3,849E+08	\$2,966E+07	7,15%
31	\$3,892E+08	\$3,561E+08	\$3,307E+07	8,50%	\$3,684E+08	\$2,074E+07	5,33%
32	\$3,378E+08	\$3,787E+08	-\$4,092E+07	-12,12%	\$3,705E+08	-\$3,274E+07	-9,69%
33	\$3,298E+08	\$3,787E+08	-\$4,885E+07	-14,81%	\$3,626E+08	-\$3,282E+07	-9,95%
34	\$3,428E+08	\$3,787E+08	-\$3,589E+07	-10,47%	\$3,626E+08	-\$1,986E+07	-5,79%
35	\$3,729E+08	\$4,038E+08	-\$3,089E+07	-8,28%	\$3,728E+08	\$8,413E+04	0,02%

36	\$3,348E+08	\$4,038E+08	-\$6,898E+07	-20,60%	\$3,702E+08	-\$3,538E+07	-10,57%
37	\$4,201E+08	\$4,038E+08	\$1,634E+07	3,89%	\$3,728E+08	\$4,732E+07	11,26%
38	\$4,248E+08	\$3,787E+08	\$4,608E+07	10,85%	\$3,753E+08	\$4,950E+07	11,65%
39	\$3,509E+08	\$3,561E+08	-\$5,173E+06	-1,47%	\$3,724E+08	-\$2,142E+07	-6,10%
40	\$3,664E+08	\$3,561E+08	\$1,030E+07	2,81%	\$3,741E+08	-\$7,672E+06	-2,09%
41	\$3,601E+08	\$3,561E+08	\$3,956E+06	1,10%	\$3,751E+08	-\$1,504E+07	-4,18%
42	\$3,222E+08	\$3,381E+08	-\$1,585E+07	-4,92%	\$2,801E+08	\$4,215E+07	13,08%
43	\$3,119E+08	\$2,904E+08	\$2,150E+07	6,89%	\$2,861E+08	\$2,582E+07	8,28%
44	\$2,776E+08	\$2,679E+08	\$9,688E+06	3,49%	\$2,711E+08	\$6,499E+06	2,34%
45	\$2,554E+08	\$2,548E+08	\$5,838E+05	0,23%	\$2,762E+08	-\$2,084E+07	-8,16%
46	\$3,188E+08	\$3,155E+08	\$3,292E+06	1,03%	\$2,822E+08	\$3,664E+07	11,49%
47	\$1,797E+08	\$2,118E+08	-\$3,202E+07	-17,82%	\$2,414E+08	-\$6,168E+07	-34,32%
48	\$1,684E+08	\$1,989E+08	-\$3,051E+07	-18,11%	\$2,030E+08	-\$3,453E+07	-20,50%
49	\$3,431E+08	\$3,371E+08	\$5,978E+06	1,74%	\$2,787E+08	\$6,433E+07	18,75%
50	\$2,918E+08	\$2,907E+08	\$1,051E+06	0,36%	\$2,440E+08	\$4,774E+07	16,36%
51	\$3,089E+08	\$2,631E+08	\$4,584E+07	14,84%	\$2,903E+08	\$1,862E+07	6,03%
52	\$2,201E+08	\$2,280E+08	-\$7,985E+06	-3,63%	\$2,378E+08	-\$1,773E+07	-8,06%
53	\$2,174E+08	\$2,281E+08	-\$1,066E+07	-4,90%	\$2,457E+08	-\$2,823E+07	-12,98%
54	\$2,114E+08	\$2,280E+08	-\$1,667E+07	-7,89%	\$2,331E+08	-\$2,176E+07	-10,29%
55	\$3,317E+08	\$3,358E+08	-\$4,115E+06	-1,24%	\$3,041E+08	\$2,761E+07	8,32%
56	\$3,323E+08	\$3,358E+08	-\$3,536E+06	-1,06%	\$3,038E+08	\$2,852E+07	8,58%
57	\$2,859E+08	\$3,356E+08	-\$4,970E+07	-17,39%	\$2,787E+08	\$7,188E+06	2,51%
58	\$3,681E+08	\$2,907E+08	\$7,734E+07	21,01%	\$2,996E+08	\$6,841E+07	18,59%
59	\$1,581E+08	\$2,343E+08	-\$7,625E+07	-48,24%	\$2,712E+08	-\$1,131E+08	-71,57%
60	\$1,780E+08	\$2,343E+08	-\$5,634E+07	-31,66%	\$2,712E+08	-\$9,322E+07	-52,38%
61	\$1,943E+08	\$2,343E+08	-\$4,001E+07	-20,59%	\$2,665E+08	-\$7,223E+07	-37,17%
62	\$2,105E+08	\$2,343E+08	-\$2,379E+07	-11,30%	\$2,665E+08	-\$5,600E+07	-26,60%
63	\$2,390E+08	\$2,343E+08	\$4,702E+06	1,97%	\$2,540E+08	-\$1,497E+07	-6,26%
64	\$2,456E+08	\$2,343E+08	\$1,133E+07	4,61%	\$2,540E+08	-\$8,346E+06	-3,40%
65	\$2,562E+08	\$2,343E+08	\$2,185E+07	8,53%	\$2,540E+08	\$2,175E+06	0,85%
66	\$2,381E+08	\$2,456E+08	-\$7,446E+06	-3,13%	\$2,802E+08	-\$4,204E+07	-17,65%
67	\$2,381E+08	\$2,636E+08	-\$2,550E+07	-10,71%	\$2,819E+08	-\$4,372E+07	-18,36%
68	\$3,023E+08	\$2,814E+08	\$2,088E+07	6,91%	\$2,766E+08	\$2,572E+07	8,51%
69	\$2,620E+08	\$2,281E+08	\$3,389E+07	12,94%	\$2,395E+08	\$2,246E+07	8,57%
70	\$2,422E+08	\$2,679E+08	-\$2,564E+07	-10,58%	\$2,711E+08	-\$2,883E+07	-11,90%
71	\$2,474E+08	\$2,682E+08	-\$2,079E+07	-8,41%	\$3,022E+08	-\$5,486E+07	-22,18%
72	\$2,474E+08	\$2,682E+08	-\$2,079E+07	-8,41%	\$3,022E+08	-\$5,486E+07	-22,18%
73	\$2,593E+08	\$2,566E+08	\$2,751E+06	1,06%	\$2,199E+08	\$3,949E+07	15,23%
74	\$3,431E+08	\$3,037E+08	\$3,939E+07	11,48%	\$3,060E+08	\$3,707E+07	10,81%
75	\$3,090E+08	\$2,631E+08	\$4,595E+07	14,87%	\$2,896E+08	\$1,945E+07	6,29%
76	\$2,934E+08	\$2,631E+08	\$3,035E+07	10,34%	\$2,842E+08	\$9,210E+06	3,14%
77	\$3,548E+08	\$3,356E+08	\$1,924E+07	5,42%	\$3,067E+08	\$4,815E+07	13,57%
78	\$2,440E+08	\$2,483E+08	-\$4,321E+06	-1,77%	\$2,753E+08	-\$3,133E+07	-12,84%
79	\$3,107E+08	\$2,889E+08	\$2,178E+07	7,01%	\$3,303E+08	-\$1,966E+07	-6,33%
80	\$2,988E+08	\$2,631E+08	\$3,568E+07	11,94%	\$2,698E+08	\$2,898E+07	9,70%
81	\$2,392E+08	\$2,622E+08	-\$2,305E+07	-9,64%	\$2,571E+08	-\$1,788E+07	-7,47%

82	\$2,523E+08	\$2,769E+08	-\$2,457E+07	-9,74%	\$2,718E+08	-\$1,949E+07	-7,72%
83	\$3,317E+08	\$3,607E+08	-\$2,893E+07	-8,72%	\$2,969E+08	\$3,478E+07	10,49%
84	\$3,599E+08	\$3,145E+08	\$4,539E+07	12,61%	\$2,728E+08	\$8,715E+07	24,21%
85	\$3,723E+08	\$3,020E+08	\$7,027E+07	18,88%	\$2,890E+08	\$8,329E+07	22,37%
86	\$3,192E+08	\$3,607E+08	-\$4,149E+07	-13,00%	\$2,912E+08	\$2,801E+07	8,78%
87	\$3,328E+08	\$3,145E+08	\$1,826E+07	5,49%	\$2,852E+08	\$4,760E+07	14,30%
88	\$4,073E+08	\$3,020E+08	\$1,053E+08	25,86%	\$2,917E+08	\$1,156E+08	28,38%
89	\$2,842E+08	\$3,145E+08	-\$3,033E+07	-10,67%	\$2,853E+08	-\$1,104E+06	-0,39%
90	\$2,690E+08	\$3,145E+08	-\$4,552E+07	-16,92%	\$2,845E+08	-\$1,552E+07	-5,77%

GLOSSÁRIO

A

Addition Date: Período de atraso na construção da sonda.

ANOVA: Análise da variância.

Auto Pipe Handling:

B

Base Oil: Quantidade mínima de óleo como base.

Bilinearidade: Equação cujo primeiro membro é linear em relação a duas variáveis diferentes ou a dois grupos diferentes de variáveis.

Blow out preventer: sistema de válvulas de alta pressão com a finalidade principal de segurança. No caso de um kick não controlado, as válvulas se fecham por acionamento da plataforma para que não ocorra um blow out. Além disso, serve para possibilitar a perfuração com lama, mantendo-a no anular do poço e para fazer o controle de influxos indesejáveis no poço, dentre outras funções.

Blow out preventer (BOP) Diameter: Diâmetro do *blow out preventer* a bordo da sonda.

BOP Mfr: Tipo de *blow out preventer* a bordo da sonda.

BOP WP Max: Pressão máxima suportada pela gaveta cega do *blow out preventer*.

Boxplot: É um gráfico em forma de caixa construído com base em 5 números, mínimo, primeiro quartil (Q1), mediana, terceiro quartil (Q3) e valor máximo. Para o teste de valores atípicos, usa-se a medida interquartil (AIQ), $Q3-Q1$ e usa-se $Q1-1,5 * AIQ$ e $Q3 + 1,5*AIQ$ como limites. Os valores que ultrapassarem esses valores serão considerados valores atípicos.

Bulk Mud: matéria prima para fazer lama.

Build Country: País onde o casco da sonda foi construído.

Build Cost: Valor monetário estimado para obra de construção da sonda.

Build Cost Remark: Custo acima do estimado para a construção da sonda.

Build Location: Local ou país onde o casco foi construído.

Build Region: Região do mundo onde o casco da sonda foi construído.

Build Yard: Estaleiro em que o casco da sonda foi construído.

Bulk Cement: Volume máximo de cimento que a sonda é capaz de armazenar.

C

Coefficiente Beta: Expressa o valor padronizado dos coeficientes da regressão, permite comparar os coeficientes quanto a sua relevância, mesmo quando as variáveis possuem diferente ordem de grandeza.

Commodity: mercadoria, no entanto, em português ela apresenta a ideia de bens pouco diferenciáveis, com mercado consumidor mundial. Nesse trabalho é um sinônimo para petróleo.

Comunalidade: é a proporção da variância de uma variável observada que é explicada pelos componentes extraídos, seu valor pode variar de 0 a 1.

Construction Status: Estágio de construção em que a sonda se encontra.

Contractor: Empresa que solicitou o projeto e sua construção.

Current Market Category:

D

Dados Anômalos: É um dado que por si só altera de forma significativa a relação entre duas variáveis.

Delivery Date: Data em que a sonda foi de fato entregue.

Derrick:

Derrick Mfr: Tipo de guincho de carga que a sonda é equipada, usado para sustentar a coluna de perfuração.

Derrick Height: Altura entre o guincho e a mesa rotativa.

Derrick Capacity: Capacidade máxima de carga suportada pelo guincho.

Design Remark: Ajustes no projeto da sonda após o início das operações da construção.

Design Water Depth: Lâmina d'água máxima na qual a sonda tem capacidade de operar.

Displacement: Deslocamento. Carga máxima suportada pelo casco da sonda dentro dos limites de segurança. É medido em toneladas.

Dynamic Positioning (DP): Posicionamento dinâmico. Consiste em um sistema de turbinas ao redor da plataforma, que são acionadas quando a deriva é excessiva.

Dynamic Positioning Class: Tipo de posicionamento dinâmico.

Draft: Profundidade que o navio ou plataforma fica submerso dentro do mar. É medido em pés.

Drilling Depth: Profundidade máxima que os equipamentos de superfície permitem realizar a perfuração.

Dual Activity: Presença de duas torres de perfuração. Assim a sonda tem a possibilidade de fazer dupla operação, conexões de dutos em uma, enquanto perfura com a outra, acelerando o processo, e assim, reduzindo os custos.

E

Engine Mfr: Tipo do motor instalado na sonda.

Engine Qty: Quantidade de motores que a sonda possui.

Engine Model: Modelo de motor que a sonda possui.

Engine HP: Potência máxima desenvolvida pelos motores da sonda.

Extended well testing (EWT) Capable: É a capacidade da sonda de avaliar o volume e as características do reservatório.

F

G

Generator Mfr: Tipo de gerador de energia que a sonda possui, geralmente à diesel.

Generator Qty: Quantidade de geradores que a sonda possui.

Generator Model: Modelo do gerador que a sonda possui.

Generator KW: Tensão máxima disponibilizada nos terminais do gerador.

H

Hull Breadth: Largura do casco.

Hull Depth: Altura total do casco.

Harsh Environment: Capacidade da sonda de suportar ambientes de mar severo.

Heterocedasticidade: É a existência de variâncias diferentes dentro de uma mesma amostra de dados.

Hull Length: Comprimento do casco.

J

L

Linearidade: É o grau de variação apresentado por uma variável dependente que esta associada a uma variável independente.

Liquid Mud: Volume máximo de fluído de perfuração que a sonda é capaz de armazenar.

M

Mudpump Mfr: Fabricante de bombas de fluído de perfuração que a sonda é equipada.

Mudpump Qty: Quantidade de bombas de fluído de perfuração a bordo na sonda.

Mudpump Type: Modelo das bombas de lama que a sonda possui.

Mudpump HP: Potência máxima desenvolvida pelas bombas de fluído de perfuração.

Mudpump Pressure: Acréscimo de pressão sofrido pelo fluído ao passar pelas bombas.

N

Normalidade: Propriedade de um conjunto de dados ser bem ajustado por uma distribuição normal.

North Sea Capable: Capacidade da sonda de operar no Mar do Norte, uma vez que submete as sondas a temperaturas muito baixas e a um estado de mar mais violento.

Norwegian SUT: Capacidade de operar na Noruega. Estar de acordo com sua regulamentação da Noruega.

O

Original Delivery Date: Data programada para entrega da sonda;

Order Date: Data em que foi ordenada o início da construção da sonda.

P

Portable Water: Volume máximo de água potável suportado pela sonda.

Q

Quarters Capacity: Capacidade de alojamento de pessoal da sonda no que se refere a acomodações.

R

Registry: País onde a sonda está registrada.

Resíduos: diferença entre valores observados e os valores previstos para a variável dependente.

Rig Design: Empresa que projetou a sonda.

Rig Name: Nome da sonda.

Rig Type: Tipo de sonda. Apenas dois tipos foram estudadas neste trabalho, sondas semissubmersíveis e navios sonda.

Rig Water Depth: Lâmina d'água máxima na qual a sonda tem capacidade de operar. Tem como unidade de medida pés.

Risers: São tubos que fazem a ligação entre o poço de petróleo marítimo e a plataforma ou navio.

S

Speculative Build: Construir sem contrato de alocação predefinido.

T

Top Drive: Equipamento único que substitui o Swivel, o Kelly e a mesa rotativa, fazendo a conexão entre a parte fixa e a rotativa, dando torque à coluna e permitindo a entrada da mangueira de injeção de fluido de perfuração. Com o Top Drive ganha-se no comprimento do Kelly, o que permite que três seções de *drillpipe* sejam encaixadas na coluna por vez, otimizando o processo de perfuração, com menos paralisações são necessárias.

Top Drive Mfr: Tipo de Top Drive;

Total Mudpumps: Número total de bombas utilizadas para pressurizar o fluido de perfuração por dentro da coluna.

V

Variable Load: Variação de carga a que a sonda pode ser submetida.

Variável Dependente: Variável que está sujeita ao efeito de outra.

Variável independente: É a variável que independe da ação de outra.

Variável Estatística: Também chamada de Variate. Combinação linear, com pesos empiricamente determinados, aplicada a um conjunto específico de variáveis. Modelo estatístico criado para explicar a relação entre variáveis dependentes e independentes.

Variáveis dicotômicas (*dummy*): Variáveis qualitativas em que só há duas respostas possíveis do tipo existência ou ausência.

Y

Year Hull Built: Ano do término efetivo da construção do casco ou da estimativa de finalização.

Year in Service: Ano em que a sonda entrou em operação.

Z

Zero Discharge: É a capacidade da sonda de reutilizar a água.