



Anderson Luiz dos Santos Ferreira

**Comparação de diferentes técnicas para detecção e
tratamento de *outliers* na determinação de fatores de
medidores**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo Programa de Pós-
Graduação em Metrologia (Área de concentração:
Metrologia para Qualidade e Inovação) da PUC-Rio.

Orientador: Prof. Elcio Cruz de Oliveira

Rio de Janeiro

Julho de 2017



Anderson Luiz dos Santos Ferreira

**Comparação de diferentes técnicas para detecção e
tratamento de *outliers* na determinação de fatores de
medidores**

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo Programa de Pós-
Graduação em Metrologia (Área de concentração:
Metrologia para Qualidade e Inovação) da PUC-Rio.
Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Elcio Cruz de Oliveira

Orientador

Programa de Pós-Graduação em Metrologia (PósMQI/PUC-Rio)

Prof. Koffi Djima Amouzou

Faculdades São José

Dra. Maria Helena Farias

INMETRO

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 12 de Julho de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Anderson Luiz dos Santos Ferreira

Formado em Ciências Contábeis pela Faculdade Pedro Leopoldo, em 2012.

Ficha Catalográfica

Ferreira, Anderson Luiz dos Santos

Comparação de diferentes técnicas para detecção e tratamento de *outliers* na determinação de fatores de medidores / Anderson Luiz dos Santos Ferreira; orientador: Elcio Cruz de Oliveira. – 2017.

86 f. : il. color. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Centro Técnico Científico, Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), 2017.

Inclui bibliografia

1. Metrologia – Teses. 2. Detecção e tratamento de outliers. 3. Fator de medidor. 4. Medidor do tipo turbina. 5. Dados não paramétricos. I. Oliveira, Elcio Cruz de. II. Pontifícia Universidade Católica do Rio de Janeiro. Centro Técnico Científico. Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação). III. Título.

CDD: 389.1

Agradecimentos

Ao meu orientador, Professor Elcio Cruz de Oliveira, pelo apoio, paciência e pela enorme motivação concedida a mim durante o curso.

À CAPES e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

À PUC-Rio e ao PósMQI, especialmente aos professores e funcionários deste, por todo o apoio dado durante o mestrado e pelos conhecimentos disponibilizados.

A professora Maria Fátima Ludovico de Almeida por suas excelentes aulas que me possibilitaram conhecer novas metodologias e agregaram fortemente ao meu conhecimento.

À minha família, em especial minha mãe Ana, por todo amor e carinho recebido ao longo da minha vida.

Aos colegas que conheci durante o Mestrado, agradeço pela ajuda e pela troca de conhecimento.

Aos meus amigos de longa data, Denise, Gilsilene, Nataly, Regina e Tatiane, que mesmo distantes sempre estiveram do meu lado.

A Deus por iluminar meus caminhos e por me dar equilíbrio nos momentos mais difíceis.

Resumo

Ferreira, Anderson Luiz dos Santos; Oliveira, Elcio Cruz de. **Comparação de diferentes técnicas para detecção e tratamento de *outliers* na determinação de fatores de medidores.** Rio de Janeiro, 2017. 86p. Dissertação de Mestrado - Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), Pontifícia Universidade Católica do Rio de Janeiro.

O **objetivo** da dissertação é analisar o comportamento das diferentes metodologias utilizadas para detecção e tratamento de *outliers* na determinação de fatores de prova de medidores do tipo turbina. A **motivação** desse trabalho é evitar tomadas de decisão equivocadas decorrentes de tratamento inadequado de *outliers*, comprometendo a confiabilidade na medição e consequentemente seu faturamento. A prova do fator de medidor pode ser considerada como um parâmetro de calibração, expressando a razão entre o volume de referência e o volume total de líquido que passa através do medidor. A Norma internacional recomenda o teste de Dixon para o tratamento de *outliers* para um conjunto de prova do fator de medidor. No entanto, a literatura é explícita quanto à avaliação do comportamento dos dados, a priori. A **metodologia** avalia se o comportamento dos dados do conjunto de prova do fator de medidor é Gaussiano, em seguida se comparam diferentes abordagens paramétricas e não paramétricas para a detecção e tratamento de *outliers* aplicados às provas dos fatores de medidores do tipo turbina para a transferência de custódia de gás liquefeito de petróleo. Posteriormente, este efeito é avaliado em relação ao número de *outliers* e como este manuseio afeta os critérios da amplitude variável para a incerteza expandida na prova do fator de medidor médio. Os **resultados** mostram que diferentes fatores de medidores médios podem ser alcançados para cada teste paramétrico e não paramétrico; de qualquer forma, **conclui-se** que estatisticamente não é observada diferença significativa entre eles.

Palavras-chave

Metrologia; Detecção e tratamento de *outliers*; Fator de medidor; Medidor do tipo turbina; Dados não paramétricos.

Abstract

Ferreira, Anderson Luiz dos Santos; Oliveira, Elcio Cruz de (Advisor). **Comparison of different approaches for detection and treatment of outliers in meter factors determination.** Rio de Janeiro, 2017. 86p. Dissertação de Mestrado - Programa Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), Pontifícia Universidade Católica do Rio de Janeiro.

The objective of the dissertation is to analyze the behavior of the different methodologies used for the detection and treatment of outliers in the determination of meter proving factors of turbine type meters. The motivation of this work is to avoid mistaken decision-making as a result of inadequate treatment of outliers, compromising reliability in measurement and consequently its billing. A meter proving factor can be considered as a calibration parameter, by expressing the ratio the reference volume and the gross volume of liquid passed through a meter. The international guideline recommends Dixon's test for outliers to a meter proving factor set. However, the literature is explicit regarding the evaluation of data behavior, a priori. The methodology evaluates if the behavior of the meter proving factor set is Gaussian, then different parametric and nonparametric approaches for detection and treating outliers applied to turbine meter proving factors for custody transfer of liquefied petroleum gas are compared. Afterwards, this effect is evaluated in relation to the number of outliers and how this handling affects the variable range criteria for expanded uncertainty in average meter proving factor. The results show that different average meter factors can be reached for each nonparametric and parametric test; anyway, it is concluded that no statistically significant difference between them is noticed.

Keywords

Metrology; Detection and treatment of outliers; Meter factor; turbine meter; nonparametric data.

Sumário

1. Introdução	14
1.1. Definição do problema de pesquisa	15
1.2. Motivação	16
1.3. Objetivos: geral e específicos	16
1.4. Estrutura da dissertação	17
2. A medição dinâmica de derivados líquidos de petróleo	18
2.1. O GLP e a matriz energética brasileira	18
2.2. Medidores do tipo turbina	20
2.3. Revisão de conceitos estatísticos	22
2.3.1. Valor Verdadeiro	22
2.3.2. Incerteza de medição	22
2.3.3. Nível de Confiança	23
2.3.4. Erros Sistemáticos	23
2.3.5. Erros aleatórios	24
2.3.6. Exatidão e precisão	24
2.3.7. Repetibilidade	25
2.3.8. Precisão Intermediária	25
2.3.9. Reprodutibilidade	25
2.4. Controle Estatístico	25
2.5. API MPMS 13.2 de 2016: O fator do medidor	26
2.6. API MPMS 4.8 de 2013: Operação dos sistemas de prova	27
2.7. O API MPMS e o problema	29
3. Controle metrológico do Fator de Medidor	30
3.1. Normalidade dos dados: teste de Shapiro-Wilk	30

3.2. Testes paramétricos para o tratamento de <i>outliers</i>	31
3.2.1. Teste de Dixon (Massart et al., 1997)	32
3.2.2. Critério de Chauvenet (Chauvenet, 1960)	33
3.2.3. Teste de Grubbs (Grubbs, 1969)	34
3.2.4. Testes não paramétricos para tratamento de <i>outliers</i>	37
3.2.5. Desvio absoluto da mediana (Miller, J.N; Miller, J.C., 2005)	37
3.2.6. Intervalo interquartilico (Montgomery; Runger, 2003)	38
3.3. Critério da amplitude variável para a incerteza no MF médio	39
3.4. Teste de hipótese	40
4. Comparação de diferentes técnicas para o tratamento de <i>outliers</i> em fatores de medidores	42
4.1. Dados experimentais	42
4.2. Resultados e Discussão	42
4.2.1. Exemplo numérico B-1 do API MPMS 13.2, Apêndice B	43
4.2.2. Conjunto de dados de provas da medição de GLP por medidores do tipo turbina em dois terminais brasileiros	50
4.2.3. Conjunto de dados de provas da medição de GLP por medidores do tipo turbina em duas refinarias brasileiras	55
5. Conclusões e recomendações	64
6. Referências bibliográficas	66
Anexo A – Artigo publicado na Revista Flow Measurement and Instrumentation - 2016	70
Anexo B – Tabelas para testes estatísticos	77

Lista de figuras

Figura 1 – Produção e consumo de GLP (bilhões de litros) – Brasil 1970-2030	19
Figura 2 – Medidor de vazão do tipo turbina	21
Figura 3 – Representação esquemática do BoxPlot (vertical)	39

Lista de tabelas

Tabela 1 – Sequência inicial dos fatores de medidores	43
Tabela 2 – Ordem ascendente para teste de <i>outlier</i>	43
Tabela 3 – Fatores de medidores	44
Tabela 4 – Fatores dos medidores após a rejeição do primeiro <i>outlier</i>	45
Tabela 5 – Fatores dos medidores após a rejeição do segundo <i>outlier</i>	46
Tabela 6 – Comparação de resultados provenientes de diferentes abordagens do exemplo numérico, API MPMS 13.2	50
Tabela 7 – Fatores de medidores de dois terminais brasileiros de armazenamento de combustível	51
Tabela 8 – Fatores dos medidores em ordem crescente - Terminal 1 e 2	51
Tabela 9 – Fatores de medidores de duas refinarias brasileiras	55
Tabela 10 – Fatores dos medidores em ordem crescente - Refinaria 1 e 2	56
Tabela 11 – Fatores dos medidores após a rejeição do primeiro <i>outlier</i> , Refinaria 1	57
Tabela 12 – Comparação das diferentes abordagens dos fatores de medidores da Refinaria 1	62
Tabela 13 – Teste de Shapiro-Wilk: Coeficiente $a_n(W)$ para cálculo de W	77
Tabela 14 – Valores críticos da estatística W de Shapiro-Wilk	80
Tabela 15 – Valores críticos para o teste de Dixon	82
Tabela 16 – Critério de Chauvenet para rejeição de valor medido	83
Tabela 17 – Valores críticos para o teste de Grubbs	84
Tabela 18 – Tabela t-Student	85
Tabela 19 – Fator de conversão para estimar o desvio padrão para n pontos a partir da amplitude	86

Lista de Siglas e Abreviaturas

Símbolo	Descrição
API	<i>American Petroleum Institute</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal do Nível Superior
GLP	Gás liquefeito de petróleo
GUM	Guia para a Expressão da Incerteza de Medição
IQR	Intervalo interquartilico
ISO	<i>International Organization for Standardization</i>
MAD	Desvio absoluto da mediana
MF	<i>Meter Factor</i>
MPMS	<i>Manual of Petroleum Measurement Standards</i>
PUC-Rio	Pontifícia Universidade Católica do Rio de Janeiro
PósMQI	Programa de Pós-Graduação em Metrologia, Qualidade e Inovação
VIM	Vocabulário Internacional de Metrologia – Conceitos Fundamentais e Gerais e Termos Associados
VIML	Vocabulário Internacional de Termos de Metrologia Legal

“Eu sei de muito pouco. Mas tenho a meu favor tudo
o que não sei...” (Clarice Lispector).

1. Introdução

As primeiras tentativas de como se tratar *outliers* em uma amostra eram arcaicas; isto é, sem rigor técnico científico pois os pesquisadores, a princípio, imaginavam que a melhor prática era a eliminação destes valores dentre os dados analisados a partir de suas impressões visuais. Nas últimas décadas, essa prática tem sido minimizada; visto que, normas e padrões internacionais sugerem diversas técnicas para a devida avaliação e tratamento dos valores discrepantes na amostra estudada.

A definição de *outlier* é abordada por diversos autores:

“Uma observação afastada, ou "*outlier*", é aquela que parece desviar-se significativamente de outros membros da amostra em que ocorre” (Grubbs, 1969, p.1).

“No caso mais simples, um *outlier* é uma observação que parece ser inconsistente com o resto de um determinado conjunto de dados” (ISO, 2010, p.11).

Outlier é: “Membro de um conjunto de valores que é inconsistente com os outros membros desse conjunto” (ISO, 2015, p.3).

A identificação de *outliers* proporciona conhecimento útil, suficiente, significativo e pode ser aplicado em distintas áreas de trabalho (Kamble; Doke, 2017).

Muitos métodos estatísticos e estatísticas de teste sumárias são sensíveis à presença de *outliers*. Por exemplo, a média e o desvio padrão de uma amostra são naturalmente influenciados pela presença de um único valor discrepante que pode levar a conclusões inválidas (ISO, 2010).

Uma base consistente para se estimar o tamanho e significância de erros é uma prática metrológica essencial a fim de evitar desgastes entre as partes envolvidas. Não há mais espaços para dúvidas que a estimativa e controle de erros podem auxiliar e evitar disputas e dúvidas correlatas à confiabilidade das operações nos sistemas de medição.

Nas mais diversas áreas, conjuntos de dados são testados por diferentes abordagens estatísticas para o tratamento de *outliers*, com a finalidade de minimizar os erros sistemáticos e aleatórios, principalmente na indústria de petróleo, devido ao alto valor agregado de suas commodities.

1.1.

Definição do problema de pesquisa

A indústria petrolífera durante o processo de medição de seus derivados líquidos, especialmente o GLP (gás liquefeito de petróleo), é extremamente criteriosa, tendo como propósito garantir a confiabilidade metrológica em toda cadeia logística, além de evitar disputas comerciais na relação estabelecida entre o fornecedor e cliente.

O GLP é o produto que mais pode afetar os fatores de medidores de uma turbina, devido a uma propriedade físico química característica quando comparada aos demais derivados líquidos de petróleo: equilíbrio entre as fases líquido-vapor, conhecido como “ponto de bolha”. A pressão de equilíbrio de bolha para o GLP é maior que zero, pois há uma vaporização natural quando este produto está submetido à pressão atmosférica.

A fim de efetivamente mitigar estes litígios, no Brasil, o medidor de vazão do tipo turbina é o mais usado para a medição de derivados líquidos de petróleo, devido ao seu excelente desempenho metrológico.

Como principal parâmetro de qualidade associado ao desempenho dos medidores do tipo turbina, destaca-se a importância do controle estatístico dos seus fatores de medição, uma vez que a presença de erros sistemáticos associados a este parâmetro pode indicar a necessidade de calibração do medidor.

A norma *American Petroleum Institute (API) Manual of Petroleum Measurement Standards* (MPMS) 13.2 de 1994, reafirmada em 2016, padrão internacional para a avaliação estatística de fatores de medidores utilizados na indústria de petróleo, sugere que o tratamento de *outliers* seja feito pelo teste de Dixon e que outras abordagens são desnecessárias independentemente do comportamento do conjunto de dados. Esta análise se faz inevitável; visto que, os testes estatísticos para tratamento de *outliers* podem ser elencados entre os que apresentam distribuição normal e os demais. A definição do problema da pesquisa

vai sendo respondida ao longo da dissertação, se permeando na seguinte questão: Como selecionar a melhor técnica para o tratamento de *outliers* nos fatores de medidores do tipo turbina para derivados líquidos de petróleo?

Para tal, os fatores de medidores provenientes da medição por medidores do tipo turbina para GLP, derivado de petróleo mais susceptível a estas questões, será o campo de aplicação para esta temática.

1.2. Motivação

A proposta deste trabalho é avaliar se o conjunto de provas de fatores de medidores segue uma distribuição Gaussiana ou normal e, em seguida, comparar as diferentes abordagens paramétricas e não paramétricas para a detecção e tratamento de valores *outliers* aplicados a medidores do tipo turbina para a transferência de custódia de gás liquefeito de petróleo.

Os resultados desta dissertação poderão trazer os seguintes benefícios para as partes interessadas:

- Minimização de erros aleatórios e correção dos erros sistemáticos dos fatores de medidores do tipo turbina na indústria de petróleo;
- Garantir o aumento da confiabilidade nas transações comerciais;
- Propor subsídios para a revisão, atualização e ampliação dos métodos estatísticos para tratamento de *outliers* descritos pelo API MPMS 13.2.

Enfim, a motivação deste trabalho é mostrar que *outliers* (valores aberrantes, dispersos, espúrios ou discrepantes) na construção da base estatística para acompanhamento do fator do medidor de uma turbina podem mascarar a variabilidade e consequentemente a incerteza do fator do medidor, além de desviá-lo do seu valor mais provável. Este comportamento pode comprometer o faturamento do produto a ser comercializado por este sistema de medição.

1.3. Objetivos: geral e específicos

O objetivo geral deste trabalho é avaliar a potencialidade das diferentes metodologias (paramétricas e não paramétricas) para o tratamento de *outliers* no

levantamento da base estatística dos fatores de medidores do tipo turbina para derivados líquidos de petróleo.

Esta dissertação tem os seguintes objetivos específicos:

- Recalcular o exemplo numérico B-1 do API MPMS 13.2 para o tratamento de *outliers* pelo teste de Dixon e avaliar sua robustez;
- Identificar na literatura testes paramétricos e não paramétricos para tratamento de *outliers*, avaliar se os mesmos oferecem melhorias à norma API MPMS 13.2 e se são estatisticamente compatíveis quando aplicados à base estatística dos fatores de medidores do tipo turbina para derivados líquidos de petróleo;
- Propor recomendações para a revisão do Apêndice B do guia API MPMS 13.2.

1.4.

Estrutura da dissertação

Este trabalho está dividido em cinco capítulos, o capítulo 1 é a introdução da dissertação.

O capítulo 2 descreve a medição dinâmica de derivados líquidos de petróleo, inicialmente mostrando a importância do GLP na matriz energética brasileira. Na sequência fala sobre a função dos medidores do tipo turbina, os tipos e as vantagens da utilização desses medidores. Por último, faz uma revisão de conceitos estatísticos (baseado no API MPMS 13.1, 2016 e no VIM, 2012) e explana sobre as normas API MPMS 13.2 e 4.8.

O capítulo 3 disserta sobre o controle metrológico do fator de medidor, trazendo as seguintes metodologias a serem utilizadas nos dados da pesquisa: teste de normalidade, testes paramétricos e não paramétricos, cálculo da incerteza expandida e teste de hipótese para comparação de medidas de posição.

O capítulo 4 revisa o exemplo numérico B-1 do API MPMS 13.2, Apêndice B e realiza o tratamento de dados reais de duas refinarias e dois terminais brasileiros; por fim, discute a comparação de diferentes técnicas para o tratamento de *outliers* em fatores de medidores.

O último capítulo discute os resultados da pesquisa, apresenta as conclusões e faz recomendações para trabalhos futuros.

2.

A medição dinâmica de derivados líquidos de petróleo

2.1.

O GLP e a matriz energética brasileira

O uso de derivados líquidos de petróleo, como fonte de energia, continua sendo a base da economia mundial. Sendo assim, a confiabilidade da medição na cadeia logística desses produtos é extremamente importante.

Historicamente, as medições de transferência de custódia de gás liquefeito de petróleo (GLP) têm sido o alvo de muitas discórdias quando comparadas às medições dos demais derivados líquidos de petróleo. Este comportamento pode ser atribuído ao fato deste derivado se apresentar em fase líquida tendendo naturalmente a entrar em equilíbrio com sua fase vapor, em ambiente fechado. Devido a esta susceptibilidade, se observa na prática, que este é o produto que mais ocasiona dispersão nos fatores de medidores do tipo turbina, o ratificando como objeto de estudo desta dissertação.

Apesar de não ser um combustível renovável, o GLP é um combustível alternativo, seguro, com poucas emissões atmosféricas, de baixo custo e benefícios sociais consideráveis. Estudos recentes em muitos mercados globais indicam que o GLP pode ser uma excelente alternativa de combustível para o setor dos transportes rodoviários (Aboul-Gheit et al., 2015).

Considerando que o petróleo e seus derivados são responsáveis pela produção de energia da maior parcela utilizada pela população em todo mundo, vale ressaltar a importância e o crescimento do GLP, nesse contexto.

No Brasil, o gás liquefeito de petróleo tem uma contribuição extremamente importante na matriz energética brasileira, Figura 1. Observa-se um aumento gradativo na produção e também no consumo de GLP, no período de 1970 a 2030.

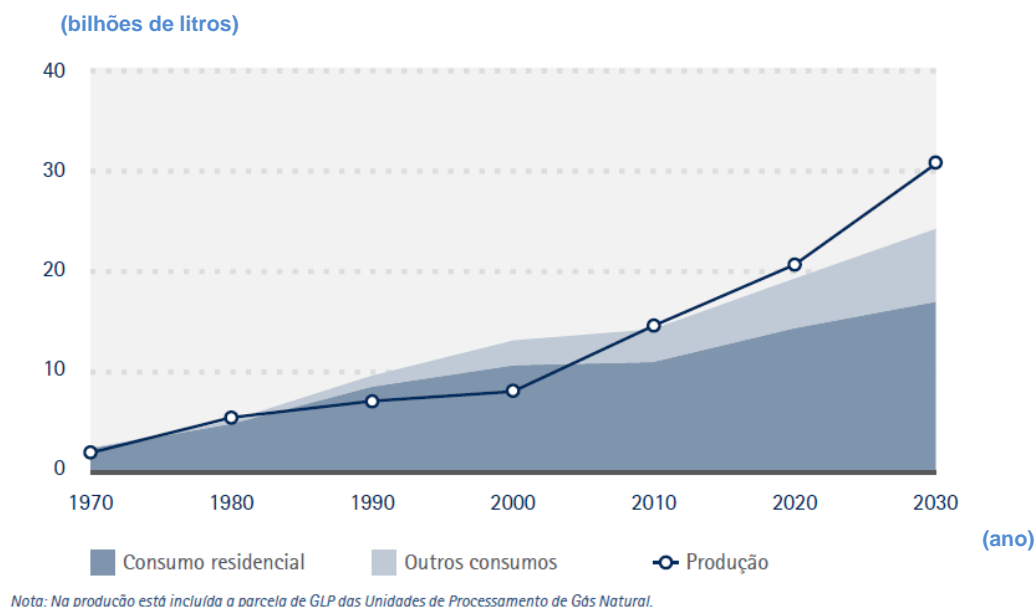


Figura 1 – Produção e consumo de GLP (bilhões de litros) – Brasil 1970-2030

Fonte: Plano Nacional de Energia 2030

A matriz energética brasileira sofre frequentes modificações. O GLP que é mais conhecido pelos brasileiros como o gás de cozinha, tem as seguintes aplicações: residencial, comercial / serviços, industrial e na agricultura entre outros exemplos (Liquigás, 2016).

A utilização do GLP nas residências brasileiras é habitual e seu uso vem crescendo a cada ano. Embora o gás liquefeito de petróleo seja mais utilizado pelos usuários brasileiros para o preparo de alimentos, ele tem inúmeras outras serventias, como: aquecimento de ambientes, churrasqueiras e etc.

O gás liquefeito de petróleo é amplamente utilizado na maior parte das empresas, sendo considerado um recurso energético multifuncional, podendo ser utilizado no cozimento de alimentos e aquecimento da água. Mais especificamente, o GLP pode ter as seguintes aplicações:

- Indústria de vidro (moldagem do material, solda e acabamento);
- Indústria automotiva (secagem da tinta na pintura);
- Indústria de papel e celulose (secagem do papel);
- Siderúrgicas (aquecimento de fornos) e etc.

O uso do gás liquefeito de petróleo na agricultura tem as seguintes finalidades:

- Secagem de grãos;

- Torrefação de grãos;
- Horticultura;
- Avicultura.

O sistema de medição equipado com medidores do tipo turbina é uma excelente opção para a medição dinâmica de GLP, especialmente por se tratar de um fluido limpo, estável, com baixa viscosidade e vazão geralmente medido à alta velocidade.

2.2. Medidores do tipo turbina

Os medidores do tipo turbina possuem a função de medir a vazão de fluídos (líquidos e gases). No mercado, há variados tipos de turbinas. Esses modelos se diferenciam basicamente no modo que os eixos do rotor são posicionados em referência ao fluxo e na forma que a medição de suas rotações é realizada (Metrosul, 2004).

Na indústria de petróleo, se encontram medidores do tipo turbina que são exclusivamente mecânicos e outros que possuem itens eletrônicos auxiliares mais modernos. Vale ressaltar ainda, que os medidores do tipo turbina são utilizados também para fazer a calibração de outros medidores; quando os mesmos exercem essa função, são considerados como um padrão para calibração (Metrosul, 2004).

O conceito de calibração de acordo com o Vocabulário Internacional de Termos de Metrologia Legal (2016, p. 7) é o seguinte:

Operação que estabelece, sob condições especificadas, numa primeira etapa, uma relação entre os valores e as incertezas de medição fornecidos por padrões e as indicações correspondentes com as incertezas associadas; numa segunda etapa, utiliza esta informação para estabelecer uma relação visando a obtenção dum resultado de medição a partir duma indicação.

Nota 1 Uma calibração pode ser expressa por meio duma declaração, uma função de calibração, um diagrama de calibração, uma curva de calibração ou uma tabela de calibração. Em alguns casos, pode consistir duma correção aditiva ou multiplicativa da indicação com uma incerteza de medição associada.

Nota 2 Convém não confundir a calibração com o ajuste dum sistema de medição, frequentemente denominado de maneira imprópria de “autocalibração”, nem com a verificação da calibração.

Nota 3 Frequentemente, apenas a primeira etapa na definição acima é entendida como sendo calibração.

O medidor do tipo turbina para derivados líquidos de petróleo é amplamente utilizado em razão do seu ótimo comportamento, baseado em suas medições que

são executadas com altíssima confiabilidade, linearidade e repetibilidade (Ribeiro, 2003).

O princípio básico de funcionamento da turbina é o seguinte: a vazão do fluido a ser medida impulsiona o rotor da turbina e o faz girar numa velocidade angular definida. A rotação das pás da turbina é diretamente proporcional a vazão do fluido. Através da detecção mecânica ou eletrônica da passagem das lâminas do rotor da turbina pode se inferir o valor da vazão. Há a geração de pulsos com frequência linearmente proporcional a velocidade do fluido e como consequência, diretamente proporcional a vazão (Ribeiro, 2003, p. 236).

Nesse trabalho, destaca-se o medidor tipo turbina convencional, Figura 2. O medidor convencional mais utilizado é o que tem as seguintes particularidades: o seu rotor é usado com eixo longitudinal à vazão, junto com bitola integral, e possui um diâmetro com valor aproximado ou igual ao tamanho da tubulação (Ribeiro, 2003).



Figura 2 – Medidor de vazão do tipo turbina

Fonte: Smartins & Pires Comércio e Representação LTDA

De acordo com o “manual de padrões para a medição de petróleo” do Instituto Americano de Petróleo – API MPMS 5.1 de 2005 reafirmado em 2016, os medidores do tipo turbina têm as seguintes vantagens em relação a outros tipos de medidores:

- Ampla faixa de vazão para líquidos de baixa viscosidade;
- O tamanho e peso são menores;
- A vida útil é maior;
- Ampla faixa de temperatura e pressão.

2.3. Revisão de conceitos estatísticos

A seguir, são destacados alguns conceitos estatísticos utilizados ao longo de todo o trabalho e que são extremamente importantes para se ter um melhor entendimento dos procedimentos utilizados na medição dinâmica de GLP por um medidor do tipo turbina.

Os conceitos utilizados nesta dissertação são baseados no jargão da indústria de petróleo (API MPMS 13.1, 2016) e no Vocabulário Internacional de Metrologia – Conceitos fundamentais e gerais e termos associados (VIM, 2012).

2.3.1. Valor Verdadeiro

De acordo com o VIM (2012, p. 19), valor verdadeiro é o “Valor duma grandeza compatível com a definição da grandeza”.

Coloquialmente, valor verdadeiro é considerado por muitos como valor correto de uma variável.

Normalmente, o **valor verdadeiro** nem sempre pode ser determinado, mas pode ser obtida a sua melhor estimativa, esperança matemática. Mediante um estudo estatístico dos diversos erros envolvidos, é viável usar os valores analisados para se obter uma estimativa do valor verdadeiro e seu nível de confiança. Em todo o conjunto de medidas com distribuição normal ou Gaussiana, a melhor estimativa do valor verdadeiro é a média aritmética, após se tratar os *outliers* e se corrigir os erros sistemáticos (API MPMS 13.1, 2016).

2.3.2. Incerteza de medição

Incerteza de Medição é o “Parâmetro não negativo que caracteriza a dispersão dos **valores** atribuídos a um **mensurando**, com base nas informações utilizadas” (VIM, 2012, p.24).

Um determinado resultado se torna mais útil quando acompanhado de uma declaração da sua confiabilidade. Os cálculos estatísticos fornecidos pelo API MPMS 13.1 e que foram utilizadas na parte metodológica desta pesquisa fornecem um intervalo dentro do qual se pode esperar encontrar o valor mais

provável de uma variável com um grau de confiança estabelecido. A expressão estatística utilizada para um intervalo com essa característica é o intervalo de confiança que esta pesquisa, de modo simplificado, o chama de incerteza de medição, pois o desvio padrão é a única fonte de incerteza, avaliação do tipo A (API MPMS 13.1, 2016).

2.3.3. Nível de Confiança

Não é uma prática comum, se estabelecer limites absolutos para a faixa de incerteza. É mais prático estabelecer uma indicação do nível de confiança que deve ser dado a um intervalo estimado de incerteza. O nível de confiança indica qual a probabilidade de que o valor verdadeiro da grandeza medida esteja incluso no intervalo demarcado. A prática mais comum é a utilização de um nível de confiança estatística de 95 % (ou 95,45 %). Este nível implica que existe uma probabilidade de 95 % (19 eventos em 20) de que o valor mais provável esteja dentro do intervalo definido. O nível de 95 % é recomendado para todas as aplicações comerciais na medição da indústria do petróleo. Em certas condições limitadas, pode ser necessário um nível de confiança diferente (API MPMS 13.1, 2016).

2.3.4. Erros Sistemáticos

Erro sistemático é: “Componente do **erro de medição** que, em **medições** repetidas, permanece constante ou varia de maneira previsível” (VIM, 2012, p.21).

Os erros sistemáticos podem causar variações nos resultados, distanciando a variável medida do seu valor mais provável. Sendo assim, a análise dos resultados de uma determinada amostra deve ser feita após a identificação e correção dos erros sistemáticos.

2.3.5. Erros aleatórios

Erro aleatório é: “Componente do **erro de medição** que, em **medições** repetidas, varia de maneira imprevisível” (VIM, 2012, p.22).

Somente após a identificação de todos os erros sistemáticos, se deve calcular o valor médio corrigido e o seu intervalo de incerteza. O aumento do número de medições reduz os erros aleatórios, pois a amostra representa melhor a população, e melhora a confiabilidade da estimativa do valor médio (API MPMS 13.1, 2016).

2.3.6. Exatidão e precisão

Exatidão é o “grau de concordância entre um valor medido e um valor verdadeiro dum mensurando” (VIM, 2012, p.20).

Precisão é o “grau de concordância entre indicações ou valores medidos, obtidos por medições repetidas, no mesmo objeto ou em objetos similares, sob condições especificadas” (VIM, 2012, p.21).

Um conjunto de medições sujeitas a erros sistemáticos menores apresenta menor tendência de medição, e conseqüentemente tem um conjunto de medidas mais exato. Todas as medições sujeitas a erros aleatórios menores compõem um conjunto de medições mais preciso. Dentro destas definições estatísticas bastante restritas, medições precisas não são necessariamente exatas, uma vez que podem apresentar baixa variabilidade, mas estarem próximas a um ponto que não é o valor verdadeiro (API MPMS 13.1, 2016).

Por outro lado, é possível se ter um conjunto de medições que são precisas quando estão sendo consideradas em grupo onde estão amplamente distribuídas, i.e., muita variabilidade, mas baixa tendência na média, pois são de confiabilidade duvidosa quando vistas individualmente (API MPMS 13.1, 2016).

Esta distinção é importante porque qualquer declaração sobre a confiabilidade deve considerar tanto os erros sistemáticos quanto os erros aleatórios. Na prática, a exatidão na medição não pode existir sem precisão, de modo que se devem fazer todos os esforços para satisfazer os critérios de confiabilidade mencionados anteriormente (API MPMS 13.1, 2016).

2.3.7. Repetibilidade

Repetibilidade é a “Condição de medição num conjunto de condições, as quais incluem o mesmo procedimento de medição, os mesmos operadores, o mesmo sistema de medição, as mesmas condições de operação e o mesmo local, assim como medições repetidas no mesmo objeto ou em objetos similares durante um curto período de tempo” (VIM, 2012, p.22).

2.3.8. Precisão Intermediária

“Condição de medição num conjunto de condições, as quais compreendem o mesmo procedimento de medição, o mesmo local e medições repetidas no mesmo objeto ou em objetos similares, ao longo dum período extenso de tempo, mas pode incluir outras condições submetidas a mudanças” (VIM, 2012, p. 23).

2.3.9. Reprodutibilidade

Reprodutibilidade é a “Condição de medição num conjunto de condições, as quais incluem diferentes locais, diferentes operadores, diferentes sistemas de medição e medições repetidas no mesmo objeto ou em objetos similares” (VIM, 2012, p.23).

2.4. Controle Estatístico

O uso adequado de técnicas estatísticas exige que o processo de medição esteja controlado. Quando não se tem controle estatístico sobre o mesmo, o valor verdadeiro da grandeza a ser medida e a incerteza de medição desses valores medidos podem não ser representativos e sem valia. Em um determinado processo de medição que está sendo controlado estatisticamente (as medições devem ser repetidas sob condições especificadas) se pode verificar: uma distribuição normal dos resultados individuais em torno de um valor médio constante (API MPMS 13.1, 2016).

Um controle estatístico rígido durante o processo de medição é uma meta ousada de ser atingida. É fundamental determinar quais serão as metodologias utilizadas para monitorar o controle estatístico durante todo o procedimento de medição. Alguns procedimentos são indispensáveis para que se tenha um controle estatístico consistente de acordo com o API MPMS 13.1:

- Deve-se definir claramente todo o procedimento de medição e suas instruções devem ser seguidas minuciosamente;
- Deve-se dispor de procedimentos específicos para monitorar e manter o equipamento;
- Devem ser incorporados meios para identificar e remover o mau funcionamento de equipamentos e erros humanos; i.e., erros grosseiros.

A aplicação destes conceitos e controles minimiza erros sistemáticos e aleatórios, razão pela qual tem sido cada vez mais importante e praticado, face às crescentes exigências da indústria de petróleo. Uma base consistente para se estimar o tamanho e significância de erros é uma prática metrológica essencial a fim de aumentar a cumplicidade entre os atores da negociação. Não há mais espaços para dúvidas que a estimativa e controle de erros podem auxiliar e evitar disputas e dúvidas correlatas à exatidão das operações nos sistemas de medição.

2.5.

API MPMS 13.2 de 2016: O fator do medidor

Uma grande variedade de projetos, equipamentos e condições operacionais estão envolvidos na operação de sistemas de prova para os fatores de medidores. Assim sendo, não é trivial, se estabelecer procedimentos padronizados para manutenção, calibração e atividades envolvendo provas para todas as instalações.

Os fatores dos medidores são habitualmente controlados para identificar tendências ou desvios rápidos, atuando como indicadores que sinalizam a necessidade de realização de manutenção e/ou calibração no equipamento de medição.

É essencial o estabelecimento de técnicas para registrar, observar e monitorar as mudanças nos fatores de medidores de modo que as incertezas de medição, mesmo levando em conta somente as contribuições do tipo A, reflitam todas as operações de medição.

A resposta do medidor de vazão do tipo turbina depende principalmente de mudanças na vazão, condição mecânica do medidor, propriedades físico-químicas do fluido, contaminantes e impurezas de fluxo. Estes parâmetros podem alterar os números de pulsos para cada unidade volumétrica de líquido que passa através do medidor de vazão da turbina, isto é, o fator de medidor (Saboochi et al., 2015).

Fator de medidor (MF) é um número adimensional obtido pela divisão entre o volume de líquido que passa através do medidor e provador (padrão de calibração), ambos nas condições de referência (Iraque, 2007).

Seguem alguns exemplos de parâmetros que podem afetar o desempenho do medidor (API MPMS 5.3, 2009, p.3):

- Vazão do fluido;
- Viscosidade do fluido;
- Temperatura do fluido;
- Massa específica do fluido;
- Pressão do líquido que flui;
- Qualidade de limpeza e lubricidade do fluido;
- Condições do provador;
- Material estranho alojado no medidor ou no elemento condicionador de fluxo;
- Alterações nas folgas mecânicas ou na geometria da pá devido a desgaste ou danos;
- Mudanças na tubulação, válvulas ou posições de válvulas que afetam o perfil do fluido ou redemoinho.

2.6.

API MPMS 4.8 de 2013: Operação dos sistemas de prova

Um medidor é considerado como provado, quando uma medida materializada de volume, denominada de provador, é comparada contra um medidor. O volume totalizado pelo medidor e pelo provador são então submetidos a uma série de cálculos utilizando os fatores de correção para converter volumes para a mesma condição de referência, estabelecendo um fator de medidor. Esta operação é também conhecida como prova do medidor (API MPMS 4.8, 2007).

Os medidores que são utilizados para a medição de transferência de custódia exigem que as provas sejam periódicas. As provas são responsáveis por verificar a exatidão e a repetibilidade ou precisão intermediária, estabelecendo fatores de medidores válidos.

O medidor quando está sendo operado constantemente deve ser provado periodicamente para comprovar sua exatidão. O fator de medidor previamente determinado pode deixar de ser aplicável devido às mudanças nas características do fluido, condições de operação e seu desgaste. Segundo o API MPMS 4.8, as razões específicas para provar medidores incluem:

- Minimizar o impacto financeiro de potenciais alterações de exatidão não detectadas no fator do medidor;
- Caso existam requisitos contratuais, tais como a manutenção programada do medidor com base no rendimento e tempo decorrido;
- Os componentes mecânicos ou elétricos do contador tenham sido abertos, alterados, reparados, removidos, trocados ou reprogramados;
- Ocorreram alterações nas condições de funcionamento, tais como tipo de fluido, massa específica, viscosidade, temperatura, pressão ou vazão.

De acordo com o API MPMS 4.8, a frequência necessária para provar os medidores pode variar de várias vezes ao dia, duas vezes por ano ou até mais, dependendo das razões listadas anteriormente. Outras razões para que sejam feitas as provas dos medidores são:

- Valor do derivado a ser comercializado;
- Custo / benefício para provar;
- Histórico de prova do medidor;
- Estabilidade do sistema de medição;
- Variações dos sistemas operacionais.

No Brasil, a frequência real de comprovação depende do contrato, dos procedimentos estabelecidos pelo operador do sistema de medição e dos regulamentos sancionados pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Em geral, a frequência de comprovação para um novo sistema começa com intervalos curtos e pode ser estendida a intervalos mais longos à medida que aumenta a confiança no sistema. O operador deve especificar o

intervalo de tempo ou o volume medido, após esses procedimentos o medidor deve ser provado novamente.

O medidor deve ser operado dentro de sua curva de desempenho, e o provador deve ser operado dentro de suas limitações de vazão. O medidor ao ser provado para se estabelecer o seu fator, o deve ser feito o mais próximo possível das mesmas condições sob as quais opera normalmente, i.e., em condições de fluxo, viscosidade, temperatura, massa específica e pressão semelhante às que existem durante a operação pretendida. O desempenho do medidor depende destas condições do processo. Portanto, durante a prova é essencial que a vazão seja mantida dentro da faixa de vazão rotineira de operação do medidor.

Critérios de aceitação bem definidos são usualmente utilizados para avaliar o fator do medidor. Os direcionadores (principais envolvidos) nas medições de transferência de custódia na indústria de petróleo estabelecem um consenso em relação a um número mínimo de corridas de provas que concordam dentro de um intervalo máximo entre consecutivos fatores de um medidor (API MPMS 13.2, 2016). O MF é parâmetro que é muito útil para avaliar a confiabilidade do sistema de medição completo, o medidor e os sistemas de provas. Os limites fixos são baseados na experiência do operador, contudo os métodos estatísticos podem ser usados para decidir se a variabilidade da medição é adequada ou não.

2.7.

O API MPMS e o problema

Para evidenciar o controle estatístico na avaliação do fator do medidor em um sistema de prova, a norma API MPMS 13.1 exemplifica o tratamento de *outliers*, exclusivamente pelo teste de Dixon, sempre utilizando a média aritmética como única medida de posição central, sem uma prévia avaliação a respeito da normalidade do conjunto de dados (Oliveira, 2014).

3. Controle metrológico do Fator de Medidor

Este capítulo descreve a metodologia utilizada, que está dividida em quatro partes. As duas primeiras partes são revisões do procedimento de Shapiro-Wilk, para avaliar se os dados seguem uma distribuição normal e dos testes paramétricos para detecção e tratamento de *outliers*, respectivamente. A terceira parte é uma introdução aos testes não paramétricos para detecção e tratamento de *outliers* e a última parte descreve um critério de aceitação baseado na incerteza de medição para avaliar uma medida central do fator do medidor.

A norma API MPMS 13.2 (2016), Apêndice B, recomenda o teste de Dixon para se avaliar a existência e tratamento de valores aberrantes e faz a seguinte consideração: “Testes alternativos para *outliers* não são esperados duplicar os resultados exatos fornecidos por estes procedimentos; no entanto, métodos computacionais alternativos devem alcançar a mesma finalidade pretendida pelos testes para *outliers* deste apêndice”.

Neste trabalho, se recomenda, primeiramente, que o comportamento normal dos dados seja avaliado. Em seguida, além deste teste proposto pelo API, teste de Dixon, outras abordagens paramétricas e não paramétricas são discutidas.

3.1. Normalidade dos dados: teste de Shapiro-Wilk

Muitos testes estatísticos são baseados na hipótese de que os dados obtidos seguem uma distribuição normal ou Gaussiana. Nem sempre isso acontece. Alguns conjuntos de dados podem não apresentar uma distribuição normal, sendo necessários alguns procedimentos, ou testes para verificar este desvio da normalidade. Além do mais, quando se considera que uma distribuição é normal, podem ser feitas previsões sobre muitos resultados individuais; por isso é necessário estar seguro que os dados estão normalmente distribuídos.

O teste de Shapiro-Wilk, apresentado em 1965, tem como finalidade calcular uma estatística de teste W e consequentemente analisar se uma determinada amostra aleatória segue uma distribuição Gaussiana.

Neste trabalho, se usa o teste de Shapiro-Wilk (Madansky, 1988), por ser considerado como uma boa alternativa para se verificar se um conjunto de dados pode ser tratado como distribuição normal ou Gaussiana. Este teste tem como limitação o número de dados, n : $3 < n \leq 50$.

Para aplicar o teste:

- Inicialmente, os valores são rearranjados em ordem crescente;
- As subtrações são formadas: $(X_{(n+1)-i} - X_i)$;
- O índice i varia 1 a $n/2$ ou de 1 a $\frac{(n+1)}{2}$, conforme n seja par ou ímpar, respectivamente;
- Os produtos são formados: $a_i(X_{(n+1)-i} - X_i)$;
- Os coeficientes a_i são tabelados;
- A soma é calculada: $SW = \sum a_i(X_{(n+1)-i} - X_i)$;
- A soma quadrática é calculada: $SQT = \sum (X_i - \bar{X})^2$, ou $(n-1)S^2$, onde o valor de \bar{X} é a média aritmética dos valores;
- A razão é formada: $W_{calculado} = \frac{SW^2}{SQT}$;
- Compara-se o valor de $W_{calculado}$ com o $W_{tabelado}$. Quando $W_{calculado} > W_{tabelado}$, se aceita que os valores estão distribuídos de acordo com certa função de distribuição normal.

3.2.

Testes paramétricos para o tratamento de *outliers*

Conforme foi comentado anteriormente, é essencial que a normalidade dos dados seja testada antes que se decida pelo teste mais adequado a ser utilizado para o tratamento de *outliers*.

Estudos científicos provenientes de medições de vazão da indústria petrolífera têm inserido testes para tratamento de valores aberrantes em suas abordagens, sempre considerando o comportamento dos dados como Gaussiano,

sem testá-los a priori (Amaral et al., 2013; Ficco, 2014; Pipa et al., 2014; Reader-Harris; Sattary; Spearman, 1990, 1995).

Devido à falta de padronização na literatura, este trabalho utiliza os seguintes termos com o mesmo sentido: *outliers*, valores aberrantes, dispersos, espúrios e discrepantes.

3.2.1. Teste de Dixon (Massart et al., 1997)

Este é o teste recomendado pela API MPMS 13.2. O teste de Dixon – também conhecido como teste-Q – avalia dados considerados suspeitos de pertencerem a uma mesma população. O valor Q de Dixon é definido como a relação entre a diferença existente entre o valor suspeito e o valor mais próximo a este e a diferença entre o maior e o menor valor do conjunto de medidas.

O Guia API MPMS 13.1, Apêndice B (2016), também contém informações sobre o teste de Dixon para o tratamento de outliers que podem ser utilizadas em um conjunto de provas de medição. Os dados considerados como outliers podem aumentar, erroneamente, o desvio padrão, o intervalo de confiança e a incerteza de medição de um conjunto de dados, neste caso específico, de provas de um medidor.

A estatística de teste se baseia na comparação do valor Q de Dixon calculado com o valor Q de Dixon crítico (tabelado), a partir do seguinte procedimento:

- Ordenar os dados amostrais em ordem crescente;
- Calcular o valor de Q;
- Comparar com o valor crítico (tabelado) de Dixon.

A rejeição de dados deve ser procedida com extrema cautela a fim de se evitar uma superestimativa do sistema de medição. Dados considerados como outliers ou podem ser, na verdade, um indicativo de erros grosseiros ou erros sistemáticos que precisam ser detectados e eliminados, quando possível.

O valor de Q calculado pela Eq. (1), Eq. (2) ou Eq. (3), depende do tamanho da amostra. Este valor é comparado ao valor crítico Q para um nível de confiança desejado. Caso este valor não seja maior do que o valor crítico, o valor suspeito é mantido; caso contrário, o valor é rejeitado.

Considerando um conjunto de n dados x_i ($i = 1, 2, \dots, n$) dispostos em ordem crescente, o teste estatístico, para $3 \leq n \leq 7$, Eq. (1) (dependendo se x_1 ou x_n é o valor suspeito):

$$Q_{3 \leftrightarrow 7} = \frac{x_2 - x_1}{x_n - x_1} \quad \text{ou} \quad \frac{x_n - x_{n-1}}{x_n - x_1} \quad (1)$$

Para conjunto de dados com $8 \leq n \leq 12$, Eq. (2):

$$Q_{8 \leftrightarrow 12} = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad \text{ou} \quad \frac{x_n - x_{n-1}}{x_n - x_2} \quad (2)$$

Para conjunto de dados com $13 \leq n \leq 40$, Eq. (3):

$$Q_{13 \leftrightarrow 40} = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad \text{ou} \quad \frac{x_n - x_{n-2}}{x_n - x_3} \quad (3)$$

Uma limitação para essa abordagem é quando se tem dois resultados suspeitos na parte baixa ou alta dos dados da amostra e quando há dois resultados suspeitos, um em cada extremo do conjunto de dados (Miller, J.N; Miller, J.C., 2005).

Outro inconveniente do teste de Dixon é o fato de que o aumento do número n de medidas, também aumente a probabilidade de ocorrência de grandes lacunas no seu conjunto de dados. Por exemplo, para duas mil medidas, a probabilidade de se detectar um desvio maior que 3,29 é grande e não há sentido descartar a medida uma vez que a probabilidade de um valor da distribuição normal padrão maior que 3,29 seja de cerca de 0,0005, isto é, $P(X > 3,29) \approx 0,0005$, onde $X \sim N(0,1)$. O critério de Chauvenet elimina esse problema.

3.2.2. Critério de Chauvenet (Chauvenet, 1960)

O teste Chauvenet é utilizado para estabelecer se existe algum valor aberrante presente na amostra estudada, levando-se em consideração que estes dados amostrais são referentes a medições que seguem uma distribuição normal.

Para que o critério de Chauvenet seja usado com o objetivo de eliminar valores discrepantes, deve-se proceder da seguinte forma:

- Primeiramente, calcular a média aritmética e desvio padrão do conjunto de dados;
- Detectar qual é o menor e o maior valor medido;

- Calcular a diferença de cada valor extremo em relação a média aritmética $|d_i| = |(x_i - \bar{X})|$;
- Determinar o valor da Equação 4, considerando \bar{X} como média aritmética e $s(X)$ como desvio padrão:

$$r = \frac{|x_i - \bar{X}|}{s(X)} \quad (4)$$

- Estabelecer qual é o valor tabelado para os graus de liberdade correspondentes;
- Com base neste critério, uma medida deve ser rejeitada se $|d_i| = |(x_i - \bar{X})| > d_{ch}$, onde d_{ch} é o limite para rejeição de Chauvenet, definido por: $p_o = \int_{-\infty}^{-d_{ch}} G(\eta) d\eta + \int_{+d_{ch}}^{+\infty} G(\eta) d\eta = \int_{-d_{ch}}^{+d_{ch}} G(\eta) d\eta = \frac{1}{2n}$, onde $G(\eta)$ é a função Gaussiana. Em outras palavras, uma medida pode ser excluída se a probabilidade de se obter o desvio específico da média for inferior a $1 / (2n)$.

Este critério estabelece que uma medida x_i deve ser descartada se o valor de r calculado pela Equação (4) é maior que o valor crítico para aqueles graus de liberdade.

Este procedimento deve ser realizado até que todos os valores considerados aberrantes sejam eliminados.

3.2.3. Teste de Grubbs (Grubbs, 1969)

O teste de Grubbs é primeiramente realizado para verificar a existência de um valor aberrante em cada extremidade do conjunto de dados. Se nesta primeira análise, um dos dois valores for considerado aberrante, ele é rejeitado, retirado do conjunto e um novo teste, verificando a existência de um valor aberrante em cada extremidade do conjunto, é realizado e assim sucessivamente; caso contrário, se nesta primeira análise, ambos os valores forem aceitos como não aberrantes, o teste é encerrado e utilizamos o conjunto de dados remanescentes para análise. Se na segunda análise os dois resultados de uma das extremidades forem considerados como aberrantes, eles devem ser rejeitados, retirados do conjunto e novo teste verificando a existência de dois valores aberrantes em cada

extremidade do conjunto é realizado e assim sucessivamente, até que ambos os valores sejam aceitos como não aberrantes.

3.2.3.1.

Um valor discrepante observado

Dado um conjunto de dados g_i com $i = 1, 2, \dots, p$, dispostos em ordem crescente, a determinação, pelo teste de Grubbs, quanto ao (maior ou menor) valor observado ser um valor discrepante, usa-se o seguinte teste estatístico, Eq. (5), (Grubbs, 1969):

$$G_C = \frac{|g_i - \bar{g}|}{s} \quad (5)$$

Onde:

- g_i é o valor suspeito;
- \bar{g} é o valor médio;
- s é o desvio padrão, Eq. (6):

$$s = \sqrt{\left[\sum_{i=1}^p (g_i - \bar{g})^2 / (p-1) \right]} \quad (6)$$

O valor de G calculado (G_c) é comparado com um valor crítico, para um nível de significância escolhido. Um valor aberrante é detectado se $G_{\text{calculado}} > G_{\text{tabelado}}$.

3.2.3.2.

Dois valores discrepantes observados

Dado um conjunto de resultados g_i , para $i = 1, 2, \dots, p$, dispostos em ordem crescente, para determinar se dois (maiores ou menores) valores são considerados como discrepantes, usando o teste de Grubbs, se deve calcular as seguintes estatísticas de teste:

3.2.3.2.1.

Dupla alta

Para testar se dois maiores valores podem ser considerados como discrepantes, calcular a relação entre as diferenças quadráticas (G), Eq. (7):

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} \quad (7)$$

Onde:

- $S_o^2 = \sum_{i=1}^p (g_i - \bar{g})^2$ é a diferença quadrática no conjunto com todos os valores;
- $S_{p-1,p}^2 = \sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2$ é a diferença quadrática no conjunto, sem os dois maiores valores;
- $\bar{g}_{p-1,p} = \sum_{i=1}^{p-2} \frac{g_i}{(p-2)}$ é a média do conjunto, sem os dois maiores valores.

O valor de G calculado é comparado com um valor crítico, para um nível de significância escolhido. Os valores são considerados aberrantes se $G_{tabelado} > G_{calculado}$.

3.2.3.2.2. Dupla baixa

Para os dois menores valores, Eq. (8):

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} \quad (8)$$

Onde:

- $S_o^2 = \sum_{i=1}^p (g_i - \bar{g})^2$ é a diferença quadrática no conjunto com todos os valores;
- $S_{1,2}^2 = \sum_{i=3}^p (g_i - \bar{g}_{1,2})^2$ é a diferença quadrática no conjunto, sem os dois menores valores.

O valor de G calculado é comparado com um valor crítico, para um nível de significância escolhido. Os valores são considerados aberrantes se $G_{tabelado} > G_{calculado}$.

Deve-se fazer a seguinte observação para os exemplos mencionados acima: (3.2.3.2.1 dupla alta e 3.2.3.2.2 dupla baixa) quando o terceiro valor da

extremidade alta ou baixa for igual ao segundo valor dessa mesma extremidade, o mesmo passa a ser avaliado pelo teste de Grubbs conforme os valores suspeitos.

3.2.4.

Testes não paramétricos para tratamento de *outliers*

Alguns conjuntos de dados não atendem ao teorema do limite central e não podem ser tratados como distribuições Gaussianas ou normais. Nestes casos, os testes que não fazem suposições sobre a forma da distribuição a partir da qual os dados tomados são chamados testes não-paramétricos ou sem distribuição particular. Para esses conjuntos de dados com comportamento que se afasta da normalidade, as abordagens paramétricas, discutidas em 3.2, não são recomendadas.

3.2.5.

Desvio absoluto da mediana (Miller, J.N; Miller, J.C., 2005)

O desvio absoluto da mediana, MAD, é considerado como uma medida robusta de dispersão estatística univariada. Além disso, MAD é a estatística mais recomendada para se tratar *outliers* do que o desvio padrão, quando os dados não apresentam comportamento normal. Nos cálculos de desvio padrão, as distâncias em relação à média são quadráticas, então grandes desvios são mais ponderados e, assim, valores aberrantes podem fortemente influenciá-la. No MAD, os desvios de um pequeno número de valores aberrantes são irrelevantes.

O valor mediano, utilizado para o cálculo do MAD em detrimento à média, separa um número igual de valores maiores dos de valores menores no conjunto de dados, na população ou uma distribuição de probabilidade, a partir da metade inferior.

Uma abordagem mais lógica para uma análise robusta, focada em dados com o comportamento não normal pode ser baseada no conceito de uma função de distância. A estimativa robusta da variância de x_i valores pode ser obtida a partir da Eq. (9):

$$\text{MAD} = \text{mediana}|x_i - \text{mediana}(x_i)| \quad (9)$$

Para o resultado suspeito (menor ou maior valor de um conjunto de dados), x_0 , se $|x_0 - \text{median}(x_i)|/MAD$ proporciona resultados superiores a 5, fica rejeitada a hipótese de nulidade, sendo este valor suspeito considerado como um *outlier*.

3.2.6.

Intervalo interquartilício (Montgomery; Runger, 2003)

Aqui, o desvio padrão, geralmente usado em testes paramétricos, é substituído pelo intervalo interquartilício (IQR).

Nesta abordagem, se leva em consideração a regra dos quartis, considerando como pontos fora da curva (*outliers*) aqueles menores ou maiores que 1,5 vezes a faixa interquartilícia (diferença entre o terceiro e primeiro quartil) a partir do primeiro e terceiro quartis, respectivamente.

O método mais prático é utilizar o princípio da mediana para os três quartis, isto é, “três medianas” são calculadas em uma mesma série, a partir do seguinte procedimento:

- Ordene os valores de forma crescente ou decrescente;
- Obtenha a mediana, que é o segundo quartil;
- Obtenha o primeiro quartil, procedendo como se fosse calcular a mediana dos dados iguais ou menores do que a mediana já calculada, Eq. (10):

$$Q_1 - 1,5 \times (Q_3 - Q_1) \quad (10)$$

- Obtenha o terceiro quartil, procedendo como se fosse calcular a mediana dos dados iguais ou maiores do que a mediana já calculada, Eq. (11):

$$Q_3 + 1,5 \times (Q_3 - Q_1). \quad (11)$$

Onde Q_1 e Q_3 são as medianas para os quartis 1 e 3, respectivamente. Os valores considerados como discrepantes são aqueles maiores do que $Q_3 + 1,5 \times (Q_3 - Q_1)$ e menores do que $Q_1 - 1,5 \times (Q_3 - Q_1)$, Figura 3.

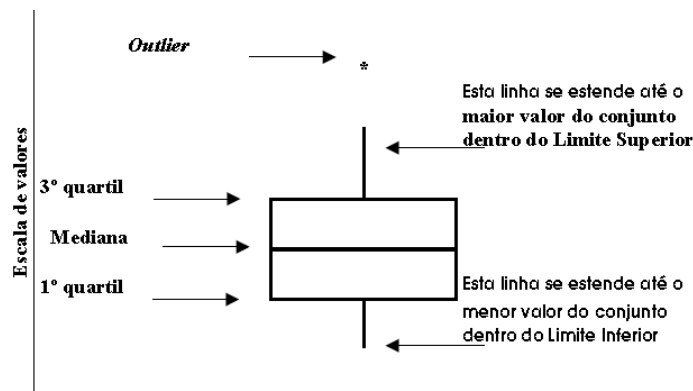


Figura 3 – Representação esquemática do BoxPlot (vertical)

Fonte: REIS, E. A.; REIS, I.A, 2002

3.3.

Critério da amplitude variável para a incerteza no MF médio

Após o tratamento de *outliers*, a incerteza expandida, considerando apenas a fonte aleatória (avaliação do tipo A), do valor médio ou mediana de um conjunto de corridas de prova de um medidor pode ser estimada de acordo com o Capítulo 13.1 do API MPMS de 2016, Equação (12):

$$a(\overline{MF}) = \frac{t \times w}{\sqrt{n} \times d_2} \quad (12)$$

Onde:

- $a(\overline{MF})$ é a incerteza expandida do valor médio ou mediana de um Fator do Medidor (MF), calculado a partir de um conjunto de corridas de prova de um medidor;
- t é o valor da estatística t-Student para um nível de confiança de 95 %, $n-1$ graus de liberdade;
- w é a amplitude de valores (máximo menos mínimo) no conjunto de provas do medidor;
- n é o número de corridas de prova do medidor;
- d_2 é o fator de conversão para estimar o desvio padrão para n pontos a partir da amplitude.

O Guia API MPMS 13.1 (2016) assume que os erros aleatórios provenientes do conjunto de dados dos fatores dos medidores seguem uma distribuição Gaussiana. No entanto, nem sempre isso ocorre. Quando a real distribuição de erros aleatórios é tal que a dispersão dos dados é menor do que seria caso se a considerasse como uma distribuição Gaussiana, os intervalos usados para estimar

os valores das variáveis de processo podem ser um pouco mais longos do que o esperado. Caso contrário, quando os intervalos são muito curtos ou são deslocados para longe do valor médio real da variável de processo a ser estimada, se obtêm intervalos com os verdadeiros parâmetros do processo menores do que os esperados (NIST/SEMATECH *e-Handbook of Statistical Methods*, acesso em 21/12/2015).

Para um número típico de corridas de prova do medidor ($n = 5$), por exemplo, a amplitude máxima deve concordar em 0,0005, a fim de não superestimar a incerteza expandida. A incerteza expandida (somente considerando-se fonte aleatória) da média deste conjunto de teste pode ser estimada a partir da Equação (12): $a(\overline{MF}) = \frac{t \times w}{\sqrt{n} \times d_2} = \frac{2,770 \times 0,0005}{\sqrt{5} \times 2,326} = 0,00027$

Assim sendo, este trabalho se baseia neste parâmetro fixo, 0,00027, como critério de aceitação, por estabelecer uma relação entre a amplitude máxima entre os valores das corridas de prova e o número de conjunto de provas.

3.4. Teste de hipótese

O teste de hipótese é uma técnica fundamentada na observação de uma amostragem, cuja finalidade é analisar parâmetros que ainda são desconhecidos na população, tendo por base a teoria das probabilidades.

No entanto, é possível comparar estatisticamente a abordagem recomendada pelo API MPMS 13.2 (independente do comportamento dos dados), teste de Dixon, os demais testes paramétricos e os testes não paramétricos (MAD ou IQR). Não há diferença significativa entre eles se o valor absoluto da diferença entre as medidas, $|MF_1 - MF_2|$, for menor ou igual à raiz quadrada da soma dos quadrados das incertezas expandidas (para um mesmo fator de abrangência) dos valores centrais (média ou mediana), $\sqrt{a(\overline{MF})_1^2 + a(\overline{MF})_2^2}$ (ISO Guide 35, 2006).

Neste trabalho, os resultados dos diferentes testes após o tratamento de *outliers* foram comparados por meio do seguinte teste de hipótese, Eq. (13):

$$|MF_1 - MF_2| \leq \sqrt{a(\overline{MF})_1^2 + a(\overline{MF})_2^2} \quad (13)$$

Esta avaliação se faz necessária, pois testes diferentes para tratamento de *outliers* podem ser estatisticamente semelhantes.

4.

Comparação de diferentes técnicas para o tratamento de *outliers* em fatores de medidores

4.1.

Dados experimentais

Os dados tratados nesta dissertação são provenientes:

- Revisão do exemplo numérico B-1 do API MPMS 13.2 (2016), Apêndice B;
- Conjunto de provas dos fatores de medidores do tipo turbina para medição de GLP em dois terminais brasileiros de armazenamento de combustível, coletados de novembro de 2008 a fevereiro de 2009;
- Conjunto de provas dos fatores de medidores do tipo turbina para medição de GLP em duas refinarias brasileiras, coletados de março de 2014 a dezembro de 2014.

Este trabalho utiliza dados experimentais reais de fatores de medidores oriundos de sistema de medição dinâmica de GLP; coletados em termos de precisão intermediária, variando operador e tempo, com um número mínimo de dez corridas de provas válidas, após o tratamento de *outliers*.

4.2.

Resultados e Discussão

O conjunto de provas dos fatores de medidores são avaliados quanto ao seu comportamento Gaussiano e são tratados por diferentes abordagens para a avaliação de *outliers*. Em seguida, o número de *outliers*, a medida de posição central (média ou mediana) de um conjunto de corridas de provas do medidor e suas incertezas expandidas são comparadas.

Em alguns casos, a fim de melhorar a sensibilidade na avaliação dos resultados, se utiliza um algoritmo significativo a mais para a média e desvio padrão do que o conjunto original de dados.

4.2.1.

Exemplo numérico B-1 do API MPMS 13.2, Apêndice B

As tabelas 1 e 2, cópia das Tabelas B-1 e B-2 da Norma API MPMS 13.2, mostram o teste de Dixon para o tratamento de *outliers*.

Tabela 1 – Sequência inicial dos fatores de medidores

Sequência dos fatores de medidores	Cálculo dos fatores de medidores
1	1,0004
2	1,0006
3	1,0005
4	1,0007
5	1,0000
6	1,0004
7	1,0009
8	1,0005
9	1,0003
10	1,0008
11	1,0006
12	1,0007
13	1,0007
14	1,0015
15	1,0009

Tabela 2 – Ordem ascendente para teste de *outlier*

Ordem ascendente	Sequência inicial de corrida	Valor
1	5	1,0000
2	9	1,0003
3	1	1,0004
4	6	1,0004
5	8	1,0005
6	3	1,0005
7	11	1,0005
8	15	1,0006
9	2	1,0006
10	12	1,0006
11	4	1,0007
12	13	1,0007
13	10	1,0008
14	7	1,0009
15	14	1,0015

É possível se observar que na Tabela 1 (correspondente à Tabela B-1, Apêndice B do API MPMS 13.2), os valores apresentados para os fatores de medidores dos números de sequência iniciais 11, 12 e 15 são diferentes daqueles apresentados em ordem crescente, Tabela 2 (correspondente à Tabela B-2 da Norma API MPMS 13.2).

Os 15 fatores de medidores foram colocados em sequência de aquisição e, posteriormente, em ordem crescente, tabela 3. Em seguida, foi feita a revisão do exemplo numérico B-1 do API MPMS 13.2, Apêndice B.

Tabela 3 – Fatores de medidores

Sequência MF	MF	Sequência inicial de corrida conforme tabela 1	MF
1	1,0004	5	1,0000
2	1,0006	9	1,0003
3	1,0005	1	1,0004
4	1,0007	6	1,0004
5	1,0000	8	1,0005
6	1,0004	3	1,0005
7	1,0009	2	1,0006
8	1,0005	11	1,0006
9	1,0003	4	1,0007
10	1,0008	12	1,0007
11	1,0006	13	1,0007
12	1,0007	10	1,0008
13	1,0007	7	1,0009
14	1,0015	15	1,0009
15	1,0009	14	1,0015

Testes paramétricos para o tratamento de *outliers* são aplicados, pois com base no teste de Shapiro-Wilk, os dados seguem uma distribuição normal:

$$W_{\text{calculado}} = 0,934 > W_{\text{tabelado}} = 0,881.$$

4.2.1.1.

Teste de Dixon

Os valores extremos, em negrito na Tabela 3, são testados pelo teste de Dixon. Para um conjunto de dados com $13 \leq n \leq 40$, Eq. (3):

$$Q_1 = \frac{x_3 - x_1}{x_{15-2} - x_1} = \frac{1,0004 - 1,0000}{1,0009 - 1,0000} = 0,444$$

$$Q_{15} = \frac{x_{15} - x_{15-2}}{x_{15} - x_3} = \frac{1,0015 - 1,0009}{1,0015 - 1,0004} = 0,546$$

O valor crítico de Q (P = 0,05 para um teste bicaudal) para um tamanho de amostra 15 é 0,565. Os valores suspeitos (1,0000 e 1,0015) são mantidos.

4.2.1.2. Critério de Chauvenet

O valor médio e o desvio padrão dos dados da Tabela 3 foram calculados, respectivamente, 1,00063 e 0,00034. Os valores extremos, em negrito na Tabela 3, são testados usando a Eq. (4):

$$r_1 = \frac{|1,0000 - 1,00063|}{0,00034} = 1,88$$

$$r_{15} = \frac{|1,0015 - 1,00063|}{0,00034} = 2,57$$

O valor crítico do critério de Chauvenet (P = 0,05 para um teste bicaudal) para um tamanho de amostra 15 é de $2,13 < 2,57$. O valor suspeito 1,0015 é rejeitado.

Um novo valor médio e o desvio padrão foram calculados, respectivamente, 1,00057 e 0,00025 e os valores extremos, em negrito na Tabela 4, são novamente tratados pela Eq. (4):

Tabela 4 – Fatores dos medidores após a rejeição do primeiro outlier

1,0000	1,0005	1,0007
1,0003	1,0006	1,0008
1,0004	1,0006	1,0009
1,0004	1,0007	1,0009
1,0005	1,0007	

$$r_1 = \frac{|1,0000 - 1,00057|}{0,00025} = 2,32$$

$$r_{14} = \frac{|1,0009 - 1,00057|}{0,00025} = 1,33$$

O valor crítico do critério de Chauvenet ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 14 é $2,11 < 2,32$. O valor suspeito 1,0000 é assim rejeitado.

Mais uma vez, o valor médio e o desvio padrão dos dados da Tabela 5 são calculados, respectivamente, 1,00062 e 0,00019. Os valores extremos, em negrito na Tabela 5, são testados usando a Eq. (4):

Tabela 5 – Fatores dos medidores após a rejeição do segundo *outlier*

	1,0005	1,0007
1,0003	1,0006	1,0008
1,0004	1,0006	1,0009
1,0004	1,0007	1,0009
1,0005	1,0007	

$$r_1 = \frac{|1,0003 - 1,00062|}{0,00019} = 1,65$$

$$r_{13} = \frac{|1,0009 - 1,00062|}{0,00019} = 1,49$$

O valor crítico do critério de Chauvenet ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 13 é 2,08. Os valores suspeitos (1,0003 e 1,0009) são mantidos.

4.2.1.3. Teste de Grubbs

Um valor discrepante observado:

Similarmente, ao critério de Chauvenet, o valor médio e o desvio padrão dos dados da Tabela 3 são calculados, respectivamente, 1,00063 e 0,00034, para a detecção de um *outlier*, a Eq. (5).

$$G_1 = \frac{|1,0000 - 1,00063|}{0,00034} = 1,877$$

$$G_{15} = \frac{|1,0015 - 1,00063|}{0,00034} = 2,569$$

O valor crítico de G ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 15 é $2,549 < 2,569$. O valor suspeito 1,0015 é rejeitado.

Um novo valor médio e o desvio padrão dos dados da Tabela 4 são calculados, respectivamente, 1,00057 e 0,00025. Os valores extremos, em negrito na Tabela 4, são testados usando a Eq. (5):

$$G_1 = \frac{|1,0000 - 1,00057|}{0,00025} = 2,320$$

$$G_{14} = \frac{|1,0009 - 1,00057|}{0,00025} = 1,334$$

Para o tamanho de amostra 14, o valor crítico de G é 2,507, ($P = 0,05$ para um teste bicaudal). Uma vez que os valores calculados de G não excedem 2,507, as medições suspeitas devem ser mantidas.

Dois valores discrepantes observados:

Para a detecção de dois *outliers*, as variâncias dos conjuntos de dados são comparadas, Equações (7) e (8), respectivamente, a fim de se avaliar, valores superiores, G_{alta} ($G_{14} = 1,0009$ e $G_{15} = 1,0015$) e os valores inferiores, G_{baixa} ($G_1 = 1,0000$ e $G_2 = 1,0003$):

Dupla alta:

$$\bar{g} = \frac{\begin{pmatrix} 1,0000 + 1,0003 + 1,0004 + 1,0004 + 1,0005 \\ 1,0005 + 1,0006 + 1,0006 + 1,0007 + 1,0007 \\ 1,0007 + 1,0008 + 1,0009 + 1,0009 + 1,0015 \end{pmatrix}}{15} = 1,0006$$

$$S_o^2 = \sum_{i=1}^p (g_i - \bar{g})^2$$

$$S_o^2 = \frac{\left[\begin{aligned} &(1,0000 - 1,0006)^2 + (1,0003 - 1,0006)^2 + (1,0004 - 1,0006)^2 + \\ &(1,0004 - 1,0006)^2 + (1,0005 - 1,0006)^2 + (1,0005 - 1,0006)^2 + \\ &(1,0006 - 1,0006)^2 + (1,0006 - 1,0006)^2 + (1,0007 - 1,0006)^2 + \\ &(1,0007 - 1,0006)^2 + (1,0007 - 1,0006)^2 + (1,0008 - 1,0006)^2 + \\ &(1,0009 - 1,0006)^2 + (1,0009 - 1,0006)^2 + (1,0015 - 1,0006)^2 \end{aligned} \right]}{12} = 0,000001593$$

$$\bar{g}_{14,15} = \frac{\left(\begin{aligned} &1,0000 + 1,0003 + 1,0004 + 1,0004 + 1,0005 + \\ &1,0005 + 1,0006 + 1,0006 + 1,0007 + 1,0007 + \\ &1,0007 + 1,0008 \end{aligned} \right)}{12} = 1,0005$$

$$S_{14,15}^2 = \sum_{i=3}^p (g_i - \bar{g}_{14,15})^2$$

$$S_{14,15}^2 = \frac{\left[\begin{aligned} &(1,0000 - 1,0005)^2 + (1,0003 - 1,0005)^2 + (1,0004 - 1,0005)^2 + \\ &(1,0004 - 1,0005)^2 + (1,0005 - 1,0005)^2 + (1,0005 - 1,0005)^2 + \\ &(1,0006 - 1,0005)^2 + (1,0006 - 1,0005)^2 + (1,0007 - 1,0005)^2 + \\ &(1,0007 - 1,0005)^2 + (1,0007 - 1,0005)^2 + (1,0008 - 1,0005)^2 \end{aligned} \right]}{12} = 0,0000005367$$

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = \frac{0,0000005367}{0,000001593} = 0,3368$$

O valor crítico de G (P = 0,05 para um teste bicaudal) para um tamanho de amostra 15 é 0,3367. Os valores suspeitos ($G_{14} = 1,0009$ e $G_{15} = 1,0015$) não são considerados como valores discrepantes, pois G_{alta} é maior do que o valor crítico.

Dupla baixa:

$$\bar{g} = \frac{\left(\begin{aligned} &1,0000 + 1,0003 + 1,0004 + 1,0004 + 1,0005 + \\ &1,0005 + 1,0006 + 1,0006 + 1,0007 + 1,0007 + \\ &1,0007 + 1,0008 + 1,0009 + 1,0009 + 1,0015 \end{aligned} \right)}{15} = 1,0006$$

$$S_o^2 = \sum_{i=1}^p (g_i - \bar{g})^2$$

$$S_o^2 = \left[\begin{aligned} &(1,0000 - 1,0006)^2 + (1,0003 - 1,0006)^2 + (1,0004 - 1,0006)^2 + \\ &(1,0004 - 1,0006)^2 + (1,0005 - 1,0006)^2 + (1,0005 - 1,0006)^2 + \\ &(1,0006 - 1,0006)^2 + (1,0006 - 1,0006)^2 + (1,0007 - 1,0006)^2 + \\ &(1,0007 - 1,0006)^2 + (1,0007 - 1,0006)^2 + (1,0008 - 1,0006)^2 + \\ &(1,0009 - 1,0006)^2 + (1,0009 - 1,0006)^2 + (1,0015 - 1,0006)^2 \end{aligned} \right] = 0,000001593$$

$$\bar{g}_{1,2} = \frac{\left(\begin{aligned} &1,0004 + 1,0004 + 1,0005 + 1,0005 + 1,0006 + \\ &1,0006 + 1,0007 + 1,0007 + 1,0007 + 1,0008 + \\ &1,0009 + 1,0009 + 1,0015 \end{aligned} \right)}{13} = 1,0007$$

$$S_{1,2}^2 = \sum_{i=3}^p (g_i - \bar{g}_{1,2})^2$$

$$S_{1,2}^2 = \left[\begin{aligned} &(1,0004 - 1,0007)^2 + (1,0004 - 1,0007)^2 + (1,0005 - 1,0007)^2 + \\ &(1,0005 - 1,0007)^2 + (1,0006 - 1,0007)^2 + (1,0006 - 1,0007)^2 + \\ &(1,0007 - 1,0007)^2 + (1,0007 - 1,0007)^2 + (1,0007 - 1,0007)^2 + \\ &(1,0008 - 1,0007)^2 + (1,0009 - 1,0007)^2 + (1,0009 - 1,0007)^2 + \\ &(1,0015 - 1,0007)^2 \end{aligned} \right] = 0,000001009$$

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = \frac{0,000001009}{0,000001593} = 0,6334$$

O valor crítico de G (P = 0,05 para um teste bicaudal) para um tamanho de amostra 15 é 0,3367. Os valores suspeitos ($G_1 = 1,0000$ e $G_2 = 1,0003$) não são considerados como valores discrepantes, pois G_{baixa} é maior do que o valor crítico.

4.2.1.4. Comparação

Os resultados das três abordagens discutidos de 4.2.1.1. a 4.2.1.3 e da API MPMS 13.2 estão compilados na Tabela 6, com a inclusão da incerteza expandida

(tipo A), $a(\overline{MF})$, Equação (12). Os parâmetros MF e $a(\overline{MF})$ estão expressos com um dígito a mais, a fim de melhorar a sensibilidade para avaliação dos resultados.

Tabela 6 – Comparação de resultados provenientes de diferentes abordagens do exemplo numérico, API MPMS 13.2

Abordagem	Número de <i>outliers</i>	MF	$a(\overline{MF})$
API MPMS 13.2 (Dixon)	2	1,00062	0,00011
Dixon	0	1,00063	0,00024
Chauvenet	2	1,00062	0,00011
Grubbs	1	1,00057	0,00015

O exemplo B-1, Apêndice B do API MPMS 13.2, que utiliza o teste de Dixon, se equivoca nos seus cálculos e considerações, excluindo dois valores que não são *outliers*. Isso ocorre porque, na Tabela B-1, Apêndice B do API MPMS 13.2, os valores apresentados nos fatores de medidores dos números de sequência iniciais 11, 12 e 15 são diferentes daqueles apresentados em ordem crescente, Tabela B-2. Na transcrição dos dados da Tabela B-1 para a Tabela B-2, ocorreram alterações. Embora, o teste de Dixon tenha sido aplicado corretamente, os valores de MF tratados para *outliers* não representam os originais.

Abordagens diferentes podem excluir números diferentes de valores aberrantes, resultando fatores de medidores e incertezas expandidas médias diferentes.

Neste exemplo com dados simulados do API MPMS 13.2, todos os valores de incerteza expandida do conjunto de corridas de prova do medidor são menores que valor esperado, 0,00027 (API MPMS 13.1, 2016), até mesmo pelo ultrapassado teste de Dixon, 0,00024 (Miller, J.N; Miller, J.C., 2005).

4.2.2.

Conjunto de dados de provas da medição de GLP por medidores do tipo turbina em dois terminais brasileiros

Os fatores de medidores de dois terminais brasileiros são apresentados na Tabela 7.

Tabela 7 – Fatores de medidores de dois terminais brasileiros de armazenamento de combustível

Terminal 1		Terminal 2	
1,0011	1,0005	1,0001	0,9995
1,0010	0,9998	0,9989	1,0009
0,9999	0,9988	0,9995	0,9995
1,0000	0,9998	0,9996	1,0000
0,9995	1,0000	0,9997	1,0008
0,9997	1,0009	0,9994	1,0014
0,9999		1,0002	0,9990
		1,0010	

Tabela 8 – Fatores dos medidores em ordem crescente - Terminal 1 e 2

Terminal 1		Terminal 2	
0,9988	1,0000	0,9989	1,0000
0,9995	1,0005	0,9990	1,0001
0,9997	1,0009	0,9994	1,0002
0,9998	1,0010	0,9995	1,0008
0,9998	1,0011	0,9995	1,0009
0,9999		0,9995	1,0010
0,9999		0,9996	1,0014
1,0000		0,9997	

Testes paramétricos para o tratamento de *outliers* são aplicados, pois com base no teste de Shapiro-Wilk, os dados do terminal 1 e 2 seguem uma distribuição normal, respectivamente:

$$W_{\text{calculado}} = 0,924 > W_{\text{tabelado}} = 0,881 \text{ e } W_{\text{calculado}} = 0,931 > W_{\text{tabelado}} = 0,881.$$

4.2.2.1. Teste de Dixon

Os valores extremos no terminal 1 e 2, em negrito na Tabela 8, são testados pelo teste de Dixon. Para um conjunto de dados com $13 \leq n \leq 40$, Equação (3):

Terminal 1

$$Q_1 = \frac{x_3 - x_1}{x_{13-2} - x_1} = \frac{0,9997 - 0,9988}{1,0009 - 0,9988} = 0,4286$$

$$Q_{13} = \frac{x_{13} - x_{13-2}}{x_{13} - x_3} = \frac{1,0011 - 1,0009}{1,0011 - 0,9997} = 0,1429$$

O valor crítico de Q (P = 0,05 para um teste bicaudal) para um tamanho de amostra 13 é 0,611. Os valores suspeitos (0,9988 e 1,0011) são mantidos.

Terminal 2

$$Q_1 = \frac{x_3 - x_1}{x_{15-2} - x_1} = \frac{0,9994 - 0,9989}{1,0009 - 0,9989} = 0,2500$$

$$Q_{15} = \frac{x_{15} - x_{15-2}}{x_{15} - x_3} = \frac{1,0014 - 1,0009}{1,0014 - 0,9994} = 0,2500$$

O valor crítico de Q (P = 0,05 para um teste bicaudal) para um tamanho de amostra 15 é 0,565. Os valores suspeitos (0,9989 e 1,0014) são mantidos.

4.2.2.2. Critério de Chauvenet

Terminal 1

O valor médio e o desvio padrão dos dados da Tabela 8 foram calculados, respectivamente, 1,0001 e 0,0007. Os valores extremos, em negrito na Tabela 8, são testados usando a Eq. (4):

$$r_1 = \frac{|0,9988 - 1,0001|}{0,0007} = 1,95$$

$$r_{13} = \frac{|1,0011 - 1,0001|}{0,0007} = 1,58$$

O valor crítico do critério de Chauvenet (P = 0,05 para um teste bicaudal) para um tamanho de amostra 13 é de 2,06. Os valores suspeitos (0,9988 e 1,0011) são mantidos.

Terminal 2

O valor médio e o desvio padrão dos dados da Tabela 8 foram calculados, respectivamente, 1,0000 e 0,0008. Os valores extremos, em negrito na Tabela 8, são testados usando a Eq. (4):

$$r_1 = \frac{|0,9989 - 1,0000|}{0,0008} = 1,41$$

$$r_{15} = \frac{|1,0014 - 1,0000|}{0,0008} = 1,89$$

O valor crítico do critério de Chauvenet ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 15 é de 2,13. Os valores suspeitos (0,9989 e 1,0014) são mantidos.

4.2.2.3. Teste de Grubbs

Terminal 1

Um valor discrepante observado:

O valor médio e o desvio padrão dos dados da Tabela 8 são calculados, respectivamente, 1,0001 e 0,0007.

Para a detecção de um *outlier*, Eq. (5):

$$G_1 = \frac{|0,9988 - 1,0001|}{0,0007} = 1,95$$

$$G_{13} = \frac{|1,0011 - 1,0001|}{0,0007} = 1,58$$

O valor crítico de G ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 13 é 2,462. Os valores suspeitos (0,9988 e 1,0011) são mantidos.

Dois valores discrepantes observados:

Para dois valores discrepantes, as variâncias dos conjuntos de dados são comparadas, Equações (7) e (8), respectivamente, a fim de se avaliar, valores superiores, G_{alta} ($G_{12} = 1,0010$ e $G_{13} = 1,0011$) e os valores inferiores, G_{baixa} ($G_1 = 0,9988$ e $G_2 = 0,9995$):

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,5521$$

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,5593$$

O valor crítico (tabelado) para $n = 13$, com 95 % de confiança é 0,2836 menor do que G_{alta} e G_{baixa} que não são considerados como valores discrepantes e não devem ser excluídos.

Terminal 2

Um valor discrepante observado:

O valor médio e o desvio padrão dos dados da Tabela 8 são calculados, respectivamente, 1,0000 e 0,0008.

Para a detecção de um *outlier*, utilizamos a Eq. (5):

$$G_1 = \frac{|0,9989 - 1,0000|}{0,0008} = 1,41$$

$$G_{15} = \frac{|1,0014 - 1,0000|}{0,0008} = 1,89$$

O valor crítico de G ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 15 é 2,549. Os valores suspeitos (0,9989 e 1,0014) são mantidos.

Dois valores discrepantes observados:

Para dois valores discrepantes, as variâncias dos conjuntos de dados são comparadas, Equações (7) e (8), respectivamente, a fim de se avaliar, valores superiores, G_{alta} ($G_{14} = 1,0010$ e $G_{15} = 1,0014$) e os valores inferiores, G_{baixa} ($G_1 = 0,9989$ e $G_2 = 0,9990$):

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,552$$

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,702$$

O valor crítico (tabelado) para $n = 15$, com 95 % de confiança é 0,3367 menor do que G_{alta} e G_{baixa} que não são considerados como valores discrepantes e não devem ser excluídos.

4.2.2.4. Incerteza expandida

A incerteza expandida em relação à média, considerando apenas a fonte aleatória destes conjuntos de teste pode ser estimada a partir da Eq. (12):

- Terminal 1 : $a(\overline{MF}) = \frac{t(12) \times w}{\sqrt{n} \times d_2} = \frac{2,179 \times 0,0023}{\sqrt{13} \times 3,336} = 0,00042$
- Terminal 2 : $a(\overline{MF}) = \frac{t(14) \times w}{\sqrt{n} \times d_2} = \frac{2,145 \times 0,0025}{\sqrt{15} \times 3,472} = 0,00040$

Estes resultados são interessantes, pois embora *outliers* não tenham sido detectados, a incerteza expandida da média está acima do valor esperado, 0,00027.

4.2.3. Conjunto de dados de provas da medição de GLP por medidores do tipo turbina em duas refinarias brasileiras

Os fatores de medidores de duas refinarias brasileiras são apresentados na Tabela 9.

Tabela 9 – Fatores de medidores de duas refinarias brasileiras

Refinaria 1		Refinaria 2	
1,0007	0,9994	1,0007	1,0007
0,9977	0,9990	1,0002	1,0002
0,9992	0,9981	1,0000	0,9998
0,9996	0,9982	1,0000	0,9998
0,9988	0,9986	0,9995	1,0000
0,9988	0,9980		
0,9985	0,9982		
0,9980	0,9981		
0,9983	0,9983		
0,9984			

Tabela 10 – Fatores dos medidores em ordem crescente - Refinaria 1 e 2

Refinaria 1		Refinaria 2	
0,9977	0,9985	0,9995	1,0000
0,9980	0,9986	0,9998	1,0002
0,9980	0,9988	0,9998	1,0002
0,9981	0,9988	1,0000	1,0007
0,9981	0,9990	1,0000	1,0007
0,9982	0,9992		
0,9982	0,9994		
0,9983	0,9996		
0,9983	1,0007		
0,9984			

O conjunto de dados da refinaria 1 não apresenta distribuição normal, por Shapiro-Wilk, respectivamente: $W_{\text{calculado}} = 0,881 < W_{\text{tabelado}} = 0,901$. Como a norma recomendada, API MPMS 13.2, não faz este juízo de valor, esta dissertação se propõe a tratar estes fatores de medidores tanto pelos testes paramétricos quanto pelos não paramétricos para o tratamento de *outliers*, avaliando a compatibilidade destes testes estatísticos.

O conjunto de dados da refinaria 2 apresenta distribuição normal, por Shapiro-Wilk, respectivamente: $W_{\text{calculado}} = 0,916 > W_{\text{tabelado}} = 0,842$.

4.2.3.1. Teste de Dixon

Refinaria 1

Os valores extremos, em negrito na Tabela 10, são testados pelo teste de Dixon.

Para conjunto de dados com $13 \leq n \leq 40$, Eq. (3):

$$Q_1 = \frac{x_3 - x_1}{x_{19-2} - x_1} = \frac{0,9980 - 0,9977}{0,9994 - 0,9977} = 0,176$$

$$Q_{19} = \frac{x_{19} - x_{19-2}}{x_{19} - x_3} = \frac{1,0007 - 0,9994}{1,0007 - 0,9980} = 0,482$$

O valor crítico de Q ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 19 é 0,501, maior que Q_1 e Q_{19} . Assim, os valores suspeitos (0,9977 e 1,0007) são mantidos.

4.2.3.2. Critério de Chauvenet

O valor médio e o desvio padrão dos dados da Tabela 10 (**refinaria 1**) foram calculados, respectivamente, 0,9986 e 0,00072 e os valores extremos são tratados pela Equação (4):

$$r_1 = \frac{|0,9977 - 0,9986|}{0,00072} = 1,29$$

$$r_{19} = \frac{|1,0007 - 0,9986|}{0,00072} = 2,90$$

O valor crítico do critério de Chauvenet ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 19 é $2,22 < 2,90$. O valor suspeito 1,0007 é assim rejeitado.

Um novo valor médio e o desvio padrão dos dados da Tabela 11 (Refinaria 1) são calculados, respectivamente, 0,9985 e 0,000525. Os valores extremos, em negrito na Tabela 11, são testados usando a Eq. (4):

Tabela 11 – Fatores dos medidores após a rejeição do primeiro *outlier*, Refinaria 1

0,9977	0,9982	0,9985	0,9992
0,9980	0,9982	0,9986	0,9994
0,9980	0,9983	0,9988	0,9996
0,9981	0,9983	0,9988	
0,9981	0,9984	0,9990	

$$r_1 = \frac{|0,9977 - 0,9985|}{0,000525} = 1,55$$

$$r_{18} = \frac{|0,9996 - 0,9985|}{0,000525} = 2,08$$

O valor crítico do critério de Chauvenet ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 18 é 2,22. Agora, os valores suspeitos (0,9977 e 0,9996) são mantidos.

4.2.3.3. Teste de Grubbs

Um valor discrepante observado:

Similarmente, ao critério de Chauvenet, o valor médio e o desvio padrão dos dados da Tabela 10 (**refinaria 1**) foram calculados, respectivamente, 0,9986 e 0,00072, para a detecção de um *outlier*, a Eq. (5).

$$G_1 = \frac{|0,9977 - 0,9986|}{0,00072} = 1,294$$

$$G_{19} = \frac{|1,0007 - 0,9986|}{0,00072} = 2,898$$

O valor crítico de G ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 19 é $2,681 < 2,898$. O valor suspeito 1,0007 é assim rejeitado.

Um novo valor médio e o desvio padrão dos dados da Tabela 11 foram calculados, respectivamente, 0,9985 e 0,000525. Os valores extremos, em negrito na Tabela 11, são testados usando a Eq. (5):

$$G_1 = \frac{|0,9977 - 0,9985|}{0,000525} = 1,546$$

$$G_{18} = \frac{|0,9996 - 0,9985|}{0,000525} = 2,076$$

Para o tamanho de amostra 18, o valor crítico de G é 2,651, ($P = 0,05$ para um teste bicaudal). Uma vez que os valores calculados de G não excedem 2,651, os valores suspeitos devem ser mantidos.

Dois valores discrepantes observados:

Para a detecção de dois outliers, as variâncias dos conjuntos de dados são comparadas, Equações (7) e (8), respectivamente, a fim de se avaliar, valores superiores, G_{alta} ($G_{18} = 0,9996$ e $G_{19} = 1,0007$) e os valores inferiores, G_{baixa} ($G_1 = 0,9977$ e $G_2 = 0,9980$):

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,3713$$

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,7896$$

O valor crítico de G (P = 0,05 para um teste bicaudal) para um tamanho de amostra 19 é 0,4214. Os valores suspeitos ($G_{18} = 0,9996$ e $G_{19} = 1,0007$) são rejeitados, pois G_{alta} é menor do que o valor crítico.

Mais uma vez, para a detecção de dois *outliers*, as novas variâncias dos conjuntos de dados são comparadas, Equações (7) e (8), respectivamente, a fim de se avaliar os valores superiores, G_{alta} ($G_{16} = 0,9992$ e $G_{17} = 0,9994$) e os valores inferiores, G_{baixa} ($G_1 = 0,9977$ e $G_2 = 0,9980$):

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,5123$$

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,6639$$

O valor crítico de G (P = 0,05 para um teste bicaudal) para um tamanho de amostra 17 é 0,3822, menor do que G_{alta} e G_{baixa} que não são considerados como valores discrepantes e não devem ser excluídos.

4.2.3.4.

Desvio absoluto médio (MAD)

Os valores extremos, em negrito na Tabela 10 (**refinaria 1**), também são tratados pelo teste *MAD*, Eq. (9):

$$MAD = \text{mediana}(x_i) - \text{mediana}(x_i) = 0,0004$$

$$|x_0 - \text{mediana}(x_i)| / MAD = |0,9977 - 0,9984| / 0,0004 = 1,750$$

$$|x_0 - \text{mediana}(x_i)| / MAD = |1,0007 - 0,9984| / 0,0004 = 5,750$$

O valor suspeito 1,0007 proporciona resultado superior a 5, ficando rejeitada a hipótese de nulidade, sendo considerado como um *outlier*.

Após o tratamento (exclusão) deste *outlier*, o teste é processado novamente se utilizando os dados extremos, em negrito da tabela 11:

$$MAD = \text{mediana}(x_i) - \text{mediana}(x_i) = 0,0003$$

$$|x_0 - \text{mediana}(x_i)| / MAD = |0,9977 - 0,9984| / 0,0003 = 2,167$$

$$|x_0 - \text{mediana}(x_i)|/MAD = |0,9996 - 0,9984|/0,0003 = 4,167$$

Os valores extremos testados, 0,9977 e 0,9996, proporcionam resultados inferiores a 5, aceitando a hipótese de nulidade, não sendo considerados como *outliers*.

4.2.3.5.

Intervalo interquartilico (IQR)

Os valores extremos da **refinaria 1**, em negrito na Tabela 10 (0,9977 e 1,0007), são testados utilizando as equações (10) e (11), baseado nos quartis:

$$\text{Mediana} = Q_2 = 0,9984$$

$$Q_1 = 0,9981; Q_3 = 0,9990 \text{ e } Q_3 - Q_1 = 0,0009$$

$$Q_1 - 1,5 \times (Q_3 - Q_1) = 0,9968 \text{ e } Q_3 + 1,5 \times (Q_3 - Q_1) = 1,0004$$

Como o valor suspeito 1,0007 está fora do intervalo interquartilico (0,9968 a 1,0004), ele deve ser rejeitado e a estatística do teste deve ser repetida a partir dos dados da Tabela 11.

Os novos parâmetros são:

$$\text{Mediana} = Q_2 = 0,9984$$

$$Q_1 = 0,9981; Q_3 = 0,9988 \text{ e } Q_3 - Q_1 = 0,00075$$

$$Q_1 - 1,5 \times (Q_3 - Q_1) = 0,9970 \text{ e } Q_3 + 1,5 \times (Q_3 - Q_1) = 0,9999$$

Como os valores suspeitos (0,9977 e 0,9996) estão dentro do intervalo interquartilico (0,9970 a 0,9999), não são detectados mais *outliers*.

4.2.3.6.

Teste de Dixon

Refinaria 2

Os valores extremos, em negrito na Tabela 10, são testados pelo teste de Dixon.

Para conjunto de dados com $8 \leq n \leq 12$, Eq. (2):

$$Q_1 = \frac{x_2 - x_1}{x_{10-1} - x_1} = \frac{0,9998 - 0,9995}{1,0007 - 0,9995} = 0,25$$

$$Q_{10} = \frac{x_{10} - x_{10-1}}{x_{10} - x_2} = \frac{1,0007 - 1,0007}{1,0007 - 0,9998} = 0,00$$

O valor crítico de Q ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 10 é 0,530, maior que Q_1 e Q_{10} . Assim, os valores suspeitos (0,9995 e 1,0007) são mantidos.

4.2.3.7. Critério de Chauvenet

O valor médio e o desvio padrão dos dados da Tabela 10 (**refinaria 2**) foram calculados, respectivamente, 1,0001 e 0,0004 e os valores extremos são tratados pela Eq. (4):

$$r_1 = \frac{|0,9995 - 1,0001|}{0,0004} = 1,55$$

$$r_{10} = \frac{|1,0007 - 1,0001|}{0,0004} = 1,60$$

O valor crítico do critério de Chauvenet ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 10 é 1,96. Os valores suspeitos (0,9995 e 1,0007) são mantidos.

4.2.3.8. Teste de Grubbs

Um valor discrepante observado:

O valor médio e o desvio padrão dos dados da Tabela 10 (**refinaria 2**) são calculados, respectivamente, 1,0001 e 0,00038.

Para a detecção de um *outlier*, utilizamos a Eq. (5):

$$G_1 = \frac{|0,9995 - 1,0001|}{0,00038} = 1,55$$

$$G_{10} = \frac{|1,0007 - 1,0001|}{0,00038} = 1,60$$

O valor crítico de G ($P = 0,05$ para um teste bicaudal) para um tamanho de amostra 10 é 2,29. Os valores suspeitos (0,9995 e 1,0007) são mantidos.

Dois valores discrepantes observados:

Para dois valores discrepantes, as variâncias dos conjuntos de dados são comparadas, Equações (7) e (8), respectivamente, a fim de se avaliar, valores superiores, G_{alta} ($G_9 = 1,0007$ e $G_{10} = 1,0007$) e os valores inferiores, G_{baixa} ($G_1 = 0,9995$ e $G_2 = 0,9998$):

$$G_{alta} = \frac{S_{p-1,p}^2}{S_o^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,209$$

$$G_{baixa} = \frac{S_{1,2}^2}{S_o^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0,532$$

O valor crítico (tabelado) para $n = 10$, com 95 % de confiança é 0,1864 menor do que G_{alta} e G_{baixa} que não são considerados como valores discrepantes e não devem ser excluídos.

Não são encontrados *outliers* na refinaria 2 pelos testes de Dixon, Chauvenet e Grubbs.

4.2.3.9. Comparação

Os resultados dos fatores de medidores da refinaria 1 referentes as cinco abordagens discutidas (paramétricas e não paramétricas) de 4.2.3.1 a 4.2.3.5 estão compilados na Tabela 12, com a inclusão da incerteza expandida da média, $a(\overline{MF})$, baseada na Eq. (12). Os parâmetros MF e $a(\overline{MF})$ estão expressos com um dígito a mais, a fim de melhorar a sensibilidade para avaliação dos resultados.

Tabela 12 – Comparação das diferentes abordagens dos fatores de medidores da Refinaria 1

Abordagem	Número de <i>outliers</i>	MF	$a(\overline{MF})$
Dixon	0	0,99863	0,00034
Chauvenet	1	0,99851	0,00026
Grubbs	2	0,99845	0,00024
MAD	1	0,99835	0,00026
IQR	1	0,99835	0,00026

Mais uma vez, observa-se que abordagens diferentes podem excluir números diferentes de *outliers*, resultando fatores de medidores (média para abordagens paramétricas e mediana para não paramétricas) e incertezas de medição diferentes.

Neste conjunto de dados reais da refinaria 1, com exceção do teste de Dixon (0,00034), todos os valores de incerteza expandida da média do conjunto de corridas de prova do medidor estão dentro do valor esperado, 0,00027.

O teste de Grubbs, para este conjunto de dados, se mostra mais restritivo ao excluir dois valores; entretanto, vale a pena lembrar que, pelo fato de que a distribuição dos dados não corresponda ao comportamento normal, os testes mais adequados são MAD e IQR.

Com base na Eq. (13): $|0,99863 - 0,99835| \leq \sqrt{0,00034^2 + 0,00026^2}$, i.e., $|0,00028| \leq 0,00043$, não existem diferenças significativas entre os fatores de medidores calculados por estas diferentes abordagens.

5. Conclusões e recomendações

O presente trabalho avaliou metrologicamente a detecção e o tratamento de *outliers* por diferentes metodologias, na construção de conjuntos de dados para o levantamento da base estatística para o controle dos fatores dos medidores do tipo turbina, de acordo com o que foi esboçado no objetivo geral da dissertação.

Conforme foi proposto nos objetivos específicos, o trabalho faz as seguintes considerações:

Esta dissertação recomenda que a norma internacional API MPMS 13.2 reavalie o exemplo usado para ilustrar a aplicação do teste de teste de Dixon e antes de aplicar qualquer teste para *outliers*, verifique se a prova do fator de medidor segue um comportamento Gaussiano, ou seja, uma distribuição normal. Nota-se que nem sempre os conjuntos de dados dos fatores de medidores seguem uma distribuição normal. Portanto, este trabalho sugere a aplicação do teste de Shapiro-Wilk para avaliar o comportamento dos dados, antes da detecção e tratamento de *outliers*.

Dentro da seara dos testes paramétricos para o tratamento de *outliers*, o teste de Dixon mostra-se limitado em alguns casos, por não tratar dois valores aberrantes de uma única vez, como o teste de Grubbs o faz. Além disto, o teste de Chauvenet pode ser uma excelente alternativa por ser mais restritivo em algumas situações (para um *outlier*). Nas situações, onde o conjunto de dados se afasta da normalidade, os testes MAD e IQR são recomendados.

O exemplo B-1 do API MPMS 13.2, utilizou o teste de Dixon, como único teste para o tratamento de valores aberrantes. Esta norma internacional aplicou corretamente o teste de Dixon, embora tenham ocorrido alterações em três fatores de medição quando eles foram organizados em ordem crescente de magnitude, gerando dúvidas para o leitor em relação ao seu uso. Além disso, contrariando o exemplo dado pela norma API MPMS 13.2, apêndice B, os métodos computacionais alternados nem sempre conseguem o mesmo propósito pretendido pelos testes de *outliers*, rejeitando diferentes valores de MF.

No entanto, abordagens diferentes numericamente podem chegar a diferentes fatores de medidores, porém, estatisticamente sem qualquer diferença significativa entre eles.

Este trabalho também destaca que, mesmo após a detecção e tratamento de *outliers* usando abordagens diferentes (paramétricas e não paramétricas), a incerteza expandida da média de um conjunto de fatores de medidores pode estar acima do valor esperado para o número de corridas do provador.

Enfim, a dissertação tem as seguintes pretensões de trabalhos futuros:

- Propor subsídios à revisão, atualização e ampliação dos métodos estatísticos (paramétricos e não paramétricos) para tratamento de *outliers* em fatores de medidores do tipo turbina para derivados líquidos de petróleo descritos pelo API MPMS 13.2;
- A interpretação das demonstrações financeiras é uma ferramenta essencial da Contabilidade. A empresa utiliza desses subsídios com a finalidade de auxiliar na tomada de decisão dos gestores e a divulgação das mesmas contribui para despertar interesse de novos investidores. Sendo assim, levando-se em consideração que esta pesquisa utilizou dados reais de duas refinarias e dois terminais brasileiros e tendo o conhecimento da importância da indústria brasileira de petróleo no cenário mundial, esta dissertação propõe uma análise do possível impacto econômico-financeiro resultante das diferentes metodologias estatísticas (paramétricas e não paramétricas) aplicadas para detecção e tratamento de valores aberrantes. Este aspecto ressalta a interdisciplinaridade da Metrologia;
- Aplicar os testes para detecção e tratamento de *outliers* para novos conjuntos de dados e para outros tipos de medidores, como por exemplo, os do tipo ultrassônico;
- Identificar na literatura outros testes paramétricos e não paramétricos para detecção e tratamento de valores aberrantes.

6. Referências bibliográficas

ABOUL-GHEIT, A. K. et al. High Yields of LPG Via n-Hexane Hydrocracking Using Unloaded Acidic Zeolite Catalysts. **Petroleum Science and Technology**, v. 33, n. 12, p. 1314-1321, 2015.

AMERICAN PETROLEUM INSTITUTE. **Manual of Petroleum Measurement Standards**. Cap. 13 – Statistical Aspects of Measuring and Sampling – Seção 1 – Statistical Concepts and Procedures in Measurements. Washington D. C., 1985. (reafirmado 2016).

_____. **Manual of Petroleum Measurement Standards**. Cap. 4.8 – Operation of Proving Systems - 1.ed. Washington D. C., 1995. (reafirmado 2007).

_____. **Manual of Petroleum Measurement Standards**. Cap. 4.8 – Operation of Proving Systems - 2.ed. Washington D. C., 2013.

_____. **Manual of Petroleum Measurement Standards**. Cap. 5 – Metering – Seção 1 – General Considerations for Measurement by Meters – 4.ed. Washington D. C., 2005. (reafirmado 2016).

_____. **Manual of Petroleum Measurement Standards**. Cap. 13 – Statistical Aspects of Measuring and Sampling – Seção 2 – Methods of Evaluating Meter Proving Data. Washington D. C., 1994. (reafirmado 2016).

_____. **Manual of Petroleum Measurement Standards**. Cap. 5 – Metering – Seção 3 – Measurement of Liquid Hydrocarbons by Turbine Meters. Washington D. C., 2005. (alterações em 2009).

BRASIL. Empresa de Pesquisa Energética. Plano Nacional de Energia 2030. Rio de Janeiro: EPE, 2007.

BUENO, A. F. et al. Validation of the performance of process stream analyzer systems with nonparametric behavior. **Accreditation and Quality Assurance**, v. 19, n. 3, p. 185-193, Jun 2014.

CHAUVENET, W. A manual of spherical and practical astronomy, vol. II (5th ed., 1960). New York: Dover; 1863.

DE OLIVEIRA, E. C. simplified calibration methodology of chromatographs used in custody transfer measurement of natural gas. **Metrology and Measurement Systems**, v. 19, n. 2, p. 405-416, 2012. ISSN 0860-8229.

DE OLIVEIRA, E. C.; LOURENCO, T. C. Comparison of different approaches to calculate a final meter factor for rotary-type natural gas displacement meters. **Flow Measurement and Instrumentation**, v. 30, p. 160-165, Apr 2013.

DO AMARAL, C. E. F. et al. Image processing techniques for high-speed videometry in horizontal two-phase slug flows. **Flow Measurement and Instrumentation**, v. 33, p. 257-264, Oct 2013.

FICCO, G. Metrological performance of diaphragm gas meters in distribution networks. **Flow Measurement and Instrumentation**, v. 37, p. 65-72, Jun 2014.

GRUBBS, F. E. PROCEDURES FOR DETECTING OUTLYING OBSERVATIONS IN SAMPLES. **Technometrics**, v. 11, n. 1, p. 1-&, 1969.

INSTITUTO NACIONAL DE METROLOGIA, QUALIDADE E TECNOLOGIA, **Guia para a expressão de incerteza de medição – GUM 2008**. Duque de Caxias, RJ: INMETRO/CICMA/SEPIN, 2012, 141 p.

_____. Vocabulário Internacional de Metrologia – Conceitos fundamentais e gerais e termos associados (VIM). Portaria N° 232, de 08 maio de 2012.

_____. Vocabulário Internacional de Termos de Metrologia Legal (VIML). Portaria INMETRO N° 150, de 29 de março de 2016.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 13528**: Statistical methods for use in proficiency testing by interlaboratory comparison. Switzerland, 2015.

_____. **ISO 16269-4**: Statistical interpretation of data – Part 4: Detection and treatment of outliers. Switzerland, 2010.

_____. **ISO Guide 35**: Reference materials – General and statistical principles for certification. Switzerland, 2006.

IRAQUE. Iraq Ministry of oil. **Iraqi National Code for measurement of hydrocarbon fluids, Fiscal & custody transfer measurements**. Iraq, 2007. 69 p.

KAMBLE, B.; DOKE, K. Outlier Detection Approaches in Data Mining. **International Research Journal of Engineering and Technology (IRJET)**, India, mar. 2017. Disponível em: < <https://www.irjet.net/archives/V4/i3/IRJET-V4I3171.pdf> > Acesso em: 15 mai. 2017.

LIQUIGÁS, PETROBRAS. Disponível em: <<https://www.liquigas.com.br/>>. Acesso em: 20 ago. 2016.

MADANSKY, A. **Prescriptions for Working Statisticians**. Springer-Verlag, New York; 1988.

MASSART, D. L. et al. **Handbook of chemometrics and qualimetrics: Part A**. Elsevier, Amsterdam; 1997.

METROSUL – Congresso Latino – Americano de Metrologia, 4., 2004, Foz do Iguaçu, Paraná. Parâmetros que influenciam na medição de petróleo por medidores do tipo turbina. Paraná, 2004. Disponível em: <https://www.ipen.br/biblioteca/cd/metrosul/2004/anais/MetrosulIV_4196.pdf>. Acesso em: 08 set. 2016.

MILLER, J. N.; MILLER, J. C. **Statistics and chemometrics for analytical chemistry**. New York; 2005.

MONTGOMERY, D. C.; RUNGER, D. C. **Applied statistics and probability for engineers**, 3rd ed. Wiley, New York; 2003.

NIST/SEMATECH e-Handbook of Statistical Methods. Disponível em: <<http://www.itl.nist.gov/div898/handbook/>>. Acesso em: 21 dez. 2015.

PIPA, D. R. et al. Typical bubble shape estimation in two-phase flow using inverse problem techniques. **Flow Measurement and Instrumentation**, v. 40, p. 64-73, Dec 2014.

POPE, J. G. et al. Extended Lee model for the turbine meter & calibrations with surrogate fluids. **Flow Measurement and Instrumentation**, v. 24, p. 71-82, Apr 2012.

PORTAL ACTION. Disponível em: <http://www.portalaction.com.br/inferencia/64-teste-de-shapiro-wilk>. Acesso em: 28 jun. 2017.

READERHARRIS, M. J.; SATTARY, J. A.; SPEARMAN, E. P. the orifice plate discharge coefficient equation - further work. **Flow Measurement and Instrumentation**, v. 6, n. 2, p. 101-114, Apr 1995.

READER-HARRIS, M. J.; SATTARY, J. A.; SPEARMAN, E. P. The orifice plate discharge coefficient equation. **Flow Measurement and Instrumentation** 1990; 1:67–76.

REIS, E. A.; REIS, I. A. Análise descritiva de dados – Síntese numérica. Belo Horizonte: Universidade Federal de Minas Gerais – Instituto de Ciências Exatas – Departamento de Estatística, 2002, 36p. Relatório Técnico.

RIBEIRO, M. A. **Medição de Petróleo e Gás Natural**. 2º. ed. Salvador, 2003.

SABOOHI, Z.; SORKHKHAH, S.; SHAKEN, H. Developing a model for prediction of helical turbine flowmeter performance using CFD. **Flow Measurement and Instrumentation**, v. 42, p. 47-57, Apr 2015.

SCRIBD: banco de dados. Disponível em: <<https://pt.scribd.com/document/22992713/Tabela-Student-Chauvenet>>. Acesso em: 27 jun. 2017.

SMARTINS & PIRES COMÉRCIO E REPRESENTAÇÃO LTDA. Medidor de vazão tipo turbina. Disponível em: <http://www.smartinstec.com.br/medidor-de-vazao-turbina>. Acesso em: 06 ago. 2016.

SURAWSKI, N. C. et al. Performance and gaseous and particle emissions from a liquefied petroleum gas (LPG) fumigated compression ignition engine. **Fuel**, v. 133, p. 17-25, Oct 2014.

7. Anexo A – Artigo publicado na Revista Flow Measurement and Instrumentation - 2016

Flow Measurement and Instrumentation 48 (2016) 29–35



Contents lists available at ScienceDirect

Flow Measurement and Instrumentation

journal homepage: www.elsevier.com/locate/flowmeasinst



Comparison of different approaches for detection and treatment of outliers in meter proving factors determination



Elcio Cruz de Oliveira^{a,b,*}, Alcir de Faro Orlando^b, Anderson Luiz dos Santos Ferreira^b, Carlos Eduardo de Oliveira Chaves^c

^a Petrobras Transporte S.A., Project Management, 20091 060 Rio de Janeiro, RJ, Brazil

^b Postgraduate Programme in Metrology: Metrology for Quality and Innovation, Pontifical Catholic University of Rio de Janeiro, 22453 900 RJ, Brazil

^c Petróleo Brasileiro S.A., Supply, Measurement Control of Custody Transfer and Audit Losses, 20231 030 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:

Received 22 October 2015

Received in revised form

21 December 2015

Accepted 4 February 2016

Available online 6 February 2016

Keywords:

Meter proving factor determination

Detection and treatment of outliers

Nonparametric data

Turbine meter

ABSTRACT

A meter proving factor can be considered as a calibration parameter, by expressing the ratio the reference volume and the gross volume of liquid passed through a meter. The international guideline recommends Dixon's test for outliers to a meter proving factor set. However, the literature says that this statistic test is restricted only to data with Gaussian behavior, besides of not to be able to detect and treat two outliers at the same time. Here, Gaussian behavior of the meter proving factor set is evaluated, then different parametric and nonparametric approaches for detection and treating outliers applied to turbine meter proving factors for custody transfer of liquefied petroleum gas are compared. Afterwards, this effect is evaluated in relation to the number of outliers and how this handling affects the variable range criteria for expanded uncertainty in average meter proving factor. The results show that different average meter factors can be reached for each nonparametric and parametric test; anyway, no statistically significant effect between them is noticed.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Although not a renewable fuel, liquefied petroleum gas (LPG) is an alternative fuel, safe, with few atmospheric emissions, low-cost and considerable social benefits [1]. Recent studies in many global markets indicate that LPG can be an excellent alternative fuel for the road transport sector [2].

Despite of the growing use of ultrasonic meters, turbine flow meters still are widely used to measure this valuable fuel [3].

Pipeline transportation companies consider the meter calibration as one of the most important parameter in order to guarantee reliability in commercial transactions. Meter calibrations can be more relevant when the device meter is mechanical as a turbine meter that is more sensitive to friction and wear [4]. In this scenario, the term "proof" represents tests in meters. A volume meter is considered as proved, when a materialized measure of volume, called as prover, is compared to the totalized volume indication of the meter. The totalized volume by the meter and the prover are submitted to several calculations, using correction factors to convert volumes to reference conditions, establishing a meter factor

(MF). The meter factor is a dimensionless number obtained by ratio of the volume of liquid passing through the meter and the volume of the prover, both at reference conditions for a particular flow rate [5].

Meter proving factors or simply meter factors (MF) are usually monitored to detect and track down trends or sudden shifts as indications when carrying out maintenance and calibration of the meter or of the auxiliary measuring equipment.

The turbine flow meter response depends mainly on changes in flow rate, mechanical condition of the meter, physicochemical properties of the fluid, contaminants and flow impurities. These parameters can change the pulse numbers for each volumetric unit of liquid passing through the turbine flow meter, i.e., the meter factor [6].

Well defined acceptance criteria are usually used to evaluate the meter factor. Custody transfer players in petroleum industry reach a consensus in relation to a minimum number of proving runs that agree within a maximum range between high and low meter factors, to a meter proving interval and deviation limit between consecutive meter factors [7]. This latter parameter is very useful to shed lights in the reliability of the complete metering system, meter and proving systems. Fixed limits are based on the operator experience, however statistics methods can be used to decide if the variability of the meter proving is suitable or not.

In order to improve this essential control, a statistical analysis

* Corresponding author at: Petrobras Transporte S.A., Project Management, 20091 060 Rio de Janeiro, RJ, Brazil.

E-mail address: elciooliveira@petrobras.com.br (E.C. de Oliveira).

<http://dx.doi.org/10.1016/j.flowmeasinst.2016.02.002>
0955-5986/© 2016 Elsevier Ltd. All rights reserved.

of a single set of meter proving factor is required. A statistically based uncertainty criterion is recommended to determine the acceptability of a set of meter proving factor. The most representative guidelines to describe these tests, API MPMS 13.1 [8] and API MPMS 13.2 [7] calculate an average or arithmetic mean and indicating, exclusively, Dixon's test for outliers. However, it is worth mentioning that this approach is specific when data are parametric, i.e. they have Gaussian distribution [9].

The guideline API MPMS 13.2, Appendix B, recommends Dixon's test for outlier and makes the following consideration: "Alternate outlier tests are not expected to duplicate the exact results provided by these procedures; however, alternate computational methods should achieve the same purpose intended by the outlier tests in this appendix [7]".

Contradicting this guideline, this paper suggests, first, that the Gaussian data behavior be evaluated based on Shapiro-Wilk test [10]. Then, beyond this proposed test, other parametric and non-parametric approaches are discussed and compared.

The aim of this paper is to evaluate the whether or not the meter proving factor set follows a Gaussian or normal distribution, and then compares different parametric and nonparametric approaches for detection and treating outliers applied to turbine meter proving factors for custody transfer of liquefied petroleum gas.

2. Methodology

The methodology is divided into four parties. The two first parts are revisions of the Shapiro-Wilk procedure and parametric tests for detection and treating of outliers; the third part is an introduction to the nonparametric tests for detection and treating of outliers and the last one is the variable range criteria for expanded uncertainty in average meter factor.

Revision of the nonparametric statistical analysis

2.1. Shapiro-Wilk test [10]

To verify if one data set can be treated as Gaussian or normal distribution, this paper uses Shapiro-Wilk test.

In this test, the data number (n) is a limitation, $3 < n \leq 50$.

The test procedure is:

- Primarily, set of n data X_i ($i=1, 2, \dots, n$) is arranged in ascending order;
- The subtraction are calculated: $(X_{(n+1)-i} - X_i)$;
- The index i varies from 1 to $n/2$ or from 1 to $(n+1)/2$, according to n being even or odd, respectively;
- The multiplications are calculated: $a_i(X_{(n+1)-i} - X_i)$;
- The coefficients a_i are tabulated;
- The sum is calculated: $SW = \sum a_i(X_{(n+1)-i} - X_i)$;
- The sum of squared is calculated: $SQT = \sum (X_i - \bar{X})^2$, or $(n-1)S^2$, considering \bar{X} as the arithmetic mean of the set data;
- The ratio is calculated: $W_{calculated} = \frac{SW^2}{SQT}$;
- Compare the calculated value $W_{calculated}$ to the value $W_{critical}$. If $W_{calculated} > W_{critical}$, the data set can be treated as Gaussian or normal distribution.

2.2. Parametric tests for outliers

Scientific studies derived from petroleum industry flow measurements have inserted outlier tests in their approaches, always considering the data behavior as Gaussian, without testing them [11–15].

2.2.1. Dixon's test [16]

Dixon's Q test, or simply the Q test, is one way to assess if suspected data belong to a population. Dixon Q value is defined as the ratio of the difference between the suspect value and the closest to this value and the difference between the largest and the smallest value of the set.

Q calculated value by Eq. (1), Eq. (2) or Eq. (3), depends on the sample size, is compared to the Q critical value for a desired level of confidence. If it is not greater than the critical value, the suspect value is kept, otherwise it is rejected.

Considering a set of n data x_i ($i=1, 2, \dots, n$) arranged in ascending order. The statistical test, for $3 \leq n \leq 7$, Eq. (1) (depending if x_1 or x_n is the suspect value):

$$Q_{3 \rightarrow 7} = \frac{x_2 - x_1}{x_n - x_1} \quad \text{or} \quad \frac{x_n - x_{n-1}}{x_n - x_1} \quad (1)$$

For $8 \leq n \leq 12$, Eq. (2):

$$Q_{8 \rightarrow 12} = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad \text{or} \quad \frac{x_n - x_{n-1}}{x_n - x_2} \quad (2)$$

For $13 < n \leq 40$, Eq. (3):

$$Q_{13 \rightarrow 40} = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad \text{or} \quad \frac{x_n - x_{n-2}}{x_n - x_3} \quad (3)$$

One limitation of this approach is when it has two suspect results in high or low part of the sample data and when there are two suspect results one at each end of the data set [17].

Another drawback of Dixon's test is the fact that increasing the number n of measures, it also increases the probability of occurrence of large gaps in the set of measures. For example, to two thousand measures, the probability of detecting a deviation greater than 3.29 is large and there is no sense to discard the measure once the probability of a value from standard normal distribution greater than 3.29 is about 0.0005, i.e., $P(X > 3.29) \approx 0.0005$, where $X \sim N(0, 1)$. Chauvenet's criterion eliminates this problem.

2.2.2. Chauvenet's criterion [18]

Based on this criterion, one measure must be rejected if $|d_j| = |(y_j - \bar{y})| > d_{ch}$, where d_{ch} is the Chauvenet's limit for rejection, defined by: $p_0 = \int_{-\infty}^{-d_{ch}} G(\eta) d\eta + \int_{+d_{ch}}^{+\infty} G(\eta) d\eta = \int_{-d_{ch}}^{+d_{ch}} G(\eta) d\eta = \frac{1}{2n}$, where $G(\eta)$ is the Gaussian function. In other words, a measure may be excluded if the probability of obtaining the specific deviation from the mean is less than $1/(2n)$.

This criterion establishes that a measure x_i must be discarded if the r calculated value by Eq. (4) is larger than the critical value for those degrees of freedom, considering \bar{X} as the arithmetic mean and $s(X)$ as the standard deviation:

$$r = \frac{|x_i - \bar{X}|}{s(X)} \quad (4)$$

2.2.3. Grubbs' test [19]

Grubbs' test is firstly performed to verify the existence of a dispersed value in each extremity of the data set. If this first analysis, one of the two values is considered to be scattered, it is refused, withdrawn from data set and new test, checking for the existence of a dispersed value in each extremity of the data set is carried out and so on. Otherwise, if this first analysis, both values are accepted as not dispersed, the test is finished and the remaining data set is used for analysis. If in the second analysis, the two results of extremity are considered as dispersed, they must be discarded, removed from the data set and new test is carried out, verifying the existence of two outliers in each extremity of the data set, and so on, until both values are accepted as not dispersed.

For the detection of one outlier, the maximum normalized deviation test is based on Eq. (5):

$$G = \frac{|g_i - \bar{g}|}{s} \quad (5)$$

where g_i is the suspected value, \bar{g} is the arithmetic mean and s is the standard deviation. G calculated value ($G_{\text{calculated}}$) is compared to the G critical value (G_{critical}) for a desired level of confidence. If $G_{\text{calculated}} > G_{\text{critical}}$, g_i is considered as an outlier.

Alternatively, for the detection of two outliers, Eqs. (6) and (7) are recommended, two largest or two smallest values, respectively:

$$G_{\text{largest}} = \frac{S_{p-1,p}^2}{S_0^2} \quad (6)$$

$$G_{\text{smallest}} = \frac{S_{1,2}^2}{S_0^2} \quad (7)$$

where:

- $S_0^2 = \sum_{i=1}^p (g_i - \bar{g})^2$ is the sum of squared deviations from the mean for the original sample;
- $S_{p-1,p}^2 = \sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2$ and $S_{1,2}^2 = \sum_{i=3}^p (g_i - \bar{g}_{1,2})^2$ are the sum of squared deviations obtained after removal of the two highest or the lowest values, respectively;
- $\bar{g}_{p-1,p}$ and $\bar{g}_{1,2}$ are the means for the original sample, without the two highest or the lowest values, respectively.

Here outliers are detected if the test statistic of Eqs. (6) and (7) are smaller than the critical value.

2.3. Nonparametric tests for outliers

Some data set do not comply with the central limit theorem and cannot be treated as Gaussian or normal distributions. In these cases, tests which make no assumptions about the shape of the distribution from which the data are taken are called nonparametric or distribution-free tests. To these data sets with behavior that departs from normality, parametric approaches (Dixon, Chauvenet and Grubbs) are not recommended.

2.3.1. Median absolute deviation (MAD) [17]

MAD is an important robust univariate statistic, being more resilient to detect outliers than the standard deviation that is generally used in parametric statistic. When standard deviations are calculated, the distances from the mean are squared, thus large standard deviations are weighted more heavily and thus outliers can clumsily influence them. Instead of it, MAD uses the median value that separates an equal number of higher values from lower values of a set of data from the lower half. This approach seems to be more logical, when related to robust tests, once that it is based on the concept of a distance function.

The robust estimate of the variance of x_i values can be derived from MAD, Eq. (8):

$$MAD = \text{median}[x_i - \text{median}(x_i)] \quad (8)$$

For an extreme result (lowest or higher value of a data set), x_0 , if $|x_0 - \text{median}(x_i)|/MAD$ provides results greater than 5, the null hypothesis is rejected and it is considered as an outlier.

2.3.2. Interquartile range (IQR) [20]

Here, the standard deviation, generally used in parametric tests, is replaced by the interquartile range.

This approach takes into account the rules of the quartiles

considering as outliers those smaller or larger than 1.5 times the interquartile range (the difference between the third and first quartile) from the first and third quartiles, Q_1 and Q_3 , respectively, i.e., the values below of the Eq. (9) and above of the Eq. (10):

$$Q_1 - 1.5 \times (Q_3 - Q_1) \quad (9)$$

$$Q_3 + 1.5 \times (Q_3 - Q_1) \quad (10)$$

Q_1 is the median of the data equal or smaller than the median of the entire data set and Q_3 is the median of the data equal to or greater than the median of the entire data set.

2.4. Variable range criteria for expanded uncertainty in average meter factor

After treatment of the outliers, the expanded uncertainty (type A [21], only) of the average value of a set of meter factors generated in proving runs, $a(\overline{MF})$, can be estimated according to Chapter 13.1 API MPMS [8], Eq. (11):

$$a(\overline{MF}) = \frac{t \times w}{\sqrt{n} \times d_2} \quad (11)$$

where:

- t is Student- t distribution, $\alpha=0.05$ and $n-1$ degrees of freedom;
- w is range of values in the meter factor set generated in proving runs;
- n is the number of meter proving runs;
- d_2 is the conversion factor for estimating the standard deviation derived from the range for n data points.

API MPMS Chapter 13.1 assumes that random errors follow a Gaussian distribution. However, not always it occurs. In occasion when the real distribution of the random errors is such that the scatter in the data is less than it would be under a Gaussian distribution, the intervals used to estimate the values of the process variables can be a little longer than expected. Otherwise when the intervals are too short or are shifted away from the true mean value of the process variable being estimated, this situation results in intervals that contain the true process parameters less often than expected [22].

For the Brazilian current practice of five proving runs, if it is considered that the maximum range of values agrees within of 0.0005, the expanded uncertainty of the average of this set can be estimated by Eq. (11):

$$a(\overline{MF}) = \frac{t \times w}{\sqrt{n} \times d_2} = \frac{2.770 \times 0.0005}{\sqrt{5} \times 2.326} = 0.00027$$

Thus, from a relationship between the maximum range of values in meter proving set and the number of meter proving runs, this paper uses this parameter as acceptance criterion for evaluating the different approaches for testing outliers in a meter proving data set.

3. Experimental data

Data processed in this paper are based on:

- Review of the numerical example B-1, Appendix B, API MPMS 13.2;
- Set of meter proving factor for LPG by turbine meters in two Brazilian fuel storage terminals and in two Brazilian refineries.

This paper uses real experimental data of LPG, because as noted in practice, LPG is the product that generates more dispersion in the set of meter proving factors in the turbine meters.

4. Results and discussion

Set of meter proving factor are treated by different approaches. The number of outliers, the average value of a set of meter proving runs and its expanded uncertainties are compared.

4.1. Numerical example B-1, Appendix B, API MPMS 13.2

15 m factors in sequential and increasing order are shown in Table 1.

Parametric tests for outliers are applied, because based on Shapiro-Wilk test, data follow a normal distribution: $W_{calculated} = 0.934 > W_{critical} = 0.881$.

4.1.1. Dixon's test

Extreme values, in bold in Table 1, are tested using Eq. (3):

$$Q_1 = \frac{x_3 - x_1}{x_{15-2} - x_1} = \frac{1.0004 - 1.0000}{1.0009 - 1.0000} = 0.444 \quad \text{and} \quad Q_{15} = \frac{x_{15} - x_{15-2}}{x_{15} - x_3} = \frac{1.0015 - 1.0009}{1.0015 - 1.0004} = 0.546$$

The critical value of Q ($P=0.05$ for a two-sided test) for a sample size 15 is 0.565. The suspect values (1.0000 and 1.0015) are retained.

4.1.2. Chauvenet's criterion

The average value and the standard deviation from data of Table 1 are calculated, 1.00063 and 0.00034, respectively. Extreme values, in bold in Table 1, are tested using Eq. (4):

$$r_1 = \frac{|1.0000 - 1.00063|}{0.00034} = 1.88 \quad \text{and} \quad r_{15} = \frac{|1.0015 - 1.00063|}{0.00034} = 2.57$$

The critical value of Chauvenet's criterion ($P=0.05$ for a two-sided test) for a sample size 15 is 2.13 < 2.57. The suspect value 1.0015 is thus rejected.

The new average value and the standard deviation from data of Table 2 are calculated, 1.00057 and 0.00025, respectively. Extreme values, in bold in Table 2, are tested using Eq. (4):

$$r_1 = \frac{|1.0000 - 1.00057|}{0.00025} = 2.32 \quad \text{and} \quad r_{14} = \frac{|1.0009 - 1.00057|}{0.00025} = 1.33$$

Table 1
Meter factors.

MF sequence	MF	Ascending order	Initial run sequence	MF
1	1.0004	1	5	1.0000
2	1.0006	2	9	1.0003
3	1.0005	3	1	1.0004
4	1.0007	4	6	1.0004
5	1.0000	5	8	1.0005
6	1.0004	6	3	1.0005
7	1.0009	7	2	1.0006
8	1.0005	8	11	1.0006
9	1.0003	9	4	1.0007
10	1.0008	10	12	1.0007
11	1.0006	11	13	1.0007
12	1.0007	12	10	1.0008
13	1.0007	13	7	1.0009
14	1.0015	14	15	1.0009
15	1.0009	15	14	1.0015

Table 2
Meter factors after rejection of the first outlier.

1.0000	1.0005	1.0007
1.0003	1.0006	1.0008
1.0004	1.0006	1.0009
1.0004	1.0007	1.0009
1.0005	1.0007	

Table 3
Meter factors after rejection of the second outlier.

	1.0005	1.0007
1.0003	1.0006	1.0008
1.0004	1.0006	1.0009
1.0004	1.0007	1.0009
1.0005	1.0007	

The critical value of Chauvenet's criterion ($P=0.05$ for a two-sided test) for a sample size 14 is 2.11 < 2.32. The suspect value 1.0000 is thus rejected.

Once more, the average value and the standard deviation from data of Table 3 are calculated, 1.00062 and 0.00019, respectively. Extreme values, in bold in Table 3, are tested using Eq. (4):

$$r_1 = \frac{|1.0003 - 1.00062|}{0.00019} = 1.65 \quad \text{and} \quad r_{13} = \frac{|1.0009 - 1.00062|}{0.00019} = 1.49$$

The critical value of Chauvenet's criterion ($P=0.05$ for a two-sided test) for a sample size 13 is 2.08. The suspect values (1.0003 and 1.0009) are retained.

4.1.3. Grubbs' test

Similarly to Chauvenet's criterion, the average value and the standard deviation from data of Table 1 are calculated, respectively, 1.00063 and 0.00034, for the detection of one outlier, Eq. (5).

$$G_1 = \frac{|1.0000 - 1.00063|}{0.00034} = 1.877 \quad \text{and} \quad G_{15} = \frac{|1.0015 - 1.00063|}{0.00034} = 2.569$$

The critical value of G ($P=0.05$ for a two-sided test) for a sample size 15 is 2.549 < 2.569. The suspect value 1.0015 is thus rejected.

The new average value and the standard deviation from data of Table 2 are calculated, 1.00057 and 0.00025, respectively. Extreme values, in bold in Table 2, are tested using Eq. (5):

$$G_1 = \frac{|1.0000 - 1.00057|}{0.00025} = 2.320 \quad \text{and} \quad G_{14} = \frac{|1.0009 - 1.00057|}{0.00025} = 1.334$$

For sample size 14, the critical value of G is 2.507, ($P=0.05$ for a two-sided test). Since the calculated values of G do not exceed 2.507, the suspect measurements should be retained.

Outliers are not detected for two suspect values, based on Eqs. (6) and (7).

Table 4
Comparison of different approaches of the numerical example, API MPMS 13.2.

Approach	Numbers of outliers	MF	$u(MF)$
API MPMS 13.2 (Dixon)	2	1.00062	0.00011
Dixon	0	1.00063	0.00024
Chauvenet	2	1.00062	0.00011
Grubbs	1	1.00057	0.00015

4.1.4. Comparison

The original API MPMS 13.2 results and the ones of the three approaches discussed from 4.1.1 to 4.1.3 are summarized in Table 4, including the expanded uncertainty of the average value, $a(\overline{MF})$, based on Eq. (11). MF and $a(\overline{MF})$ parameters are expressed with one more digit in order to improve the sensitivity to evaluate the results.

Using Dixon's test, Example B-1, Appendix B, API MPMS 13.2 misinterprets its calculations and considerations, excluding two values that are not outliers. It occurs because, in Appendix B of API MPMS 13.2, Table B-1, the values presented on the meter factors at the initial sequence numbers 11, 12 and 15 are different from those presented in ascending order, Table B-1. In the data copy from Table B-1 to Table B-1, changes occurred. Although, Dixon's test has been applied correctly, the MF values treated for outliers do not represent the original ones.

Different approaches may exclude different numbers of outliers, resulting different meter factors and different average uncertainties.

In this example with simulated data from API MPMS 13.2, all expanded uncertainty of the average value of a set of meter proving runs ($n=5$) are smaller than the expected value, 0.00027, even the limited Dixon's test, 0.00024.

4.2. Set of meter proving data for LPG by turbine meters in two Brazilian terminals

Meter factors of two Brazilian terminals are shown in Table 5.

Parametric tests for outliers are applied, because based on Shapiro–Wilk test, data of terminal 1 and 2 follow a normal distribution, respectively: $W_{calculated} = 0.924 > W_{critical} = 0.881$ and $W_{calculated} = 0.931 > W_{critical} = 0.881$.

No outliers are found by Dixon, Chauvenet and Grubbs.

The expanded uncertainties of the average value are calculated by Eq. (11):

$$\text{Terminal 1: } a(\overline{MF}) = \frac{t(12) \times w}{\sqrt{n} \times d_2} = \frac{2.179 \times 0.0023}{\sqrt{13} \times 3.336} = 0.00042$$

$$\text{Terminal 2: } a(\overline{MF}) = \frac{t(14) \times w}{\sqrt{n} \times d_2} = \frac{2.145 \times 0.0025}{\sqrt{15} \times 3.472} = 0.00040$$

These results are interesting because although outliers have not been detected, the expanded uncertainties of the average value are above the expected value for $n=5$, 0.00027.

4.3. Set of meter proving data for LPG by turbine meters in two Brazilian refineries

Meter factors of two Brazilian refineries are shown in Table 6.

Parametric tests for outliers are applied, because based on Shapiro–Wilk test, data of Refinery 2 follow a normal distribution: $W_{calculated} = 0.916 > W_{critical} = 0.842$. No outliers are found by Dixon, Chauvenet and Grubbs.

Table 5
Meter factors of two Brazilian fuel storage terminals.

Terminal 1		Terminal 2	
1.0011	1.0005	1.0001	0.9995
1.0010	0.9998	0.9989	1.0009
0.9999	0.9988	0.9995	0.9995
1.0000	0.9998	0.9996	1.0000
0.9995	1.0000	0.9997	1.0008
0.9997	1.0009	0.9994	1.0014
0.9999		1.0002	0.9990
		1.0010	

Table 6
Meter factors of two Brazilian refineries.

Refinery 1		Refinery 2	
1.0007	0.9994	1.0007	1.0007
0.9977	0.9990	1.0002	1.0002
0.9992	0.9981	1.0000	0.9998
0.9996	0.9982	1.0000	0.9998
0.9988	0.9986	0.9995	1.0000
0.9988	0.9980		
0.9985	0.9982		
0.9980	0.9981		
0.9983	0.9983		
0.9984			

The expanded uncertainty of the average value is calculated by Eq. (11):

$$\text{Refinery 2: } a(\overline{MF}) = \frac{t(9) \times w}{\sqrt{n} \times d_2} = \frac{2.262 \times 0.0012}{\sqrt{10} \times 3.078} = 0.00028$$

Once more, although outliers have not been detected, the expanded uncertainty of the average value is above the expected value for $n=5$, 0.00027.

However, the data set of the Refinery 1 does not follow a Gaussian distribution based on Shapiro–Wilk. As the recommended guideline usually used in industry, API MPMS 13.2, does not make value judgment upon the data behavior, this paper aims to treat these meter factors (Refinery 1) both by parametric tests and by the nonparametric tests for outliers detection and treatment, in order to assess the compatibility between some statistical tests.

4.3.1. Dixon's test

Extreme values, in bold in Table 6, are tested using Eq. (3):

$$Q_1 = \frac{x_3 - x_1}{x_{19-2} - x_1} = \frac{0.9980 - 0.9977}{0.9994 - 0.9977} = 0.176 \quad \text{and} \quad Q_{19} = \frac{x_{19} - x_{19-2}}{x_{19} - x_3} = \frac{1.0007 - 0.9994}{1.0007 - 0.9980} = 0.482$$

The critical value of Q ($P=0.05$ for a two-sided test) for a sample size 19 is 0.501, greater than Q_1 and Q_{19} . Thus, the suspect values (0.9977 and 1.0007) are retained.

4.3.2. Chauvenet's criterion

The average value and the standard deviation from data of Table 6 (Refinery 1) are calculated, 0.9986 and 0.00072, respectively. Extreme values, in bold in Table 1, are tested using Eq. (4):

$$r_1 = \frac{|0.9977 - 0.9986|}{0.00072} = 1.29 \quad \text{and} \quad r_{19} = \frac{|1.0007 - 0.9986|}{0.00072} = 2.90$$

The critical value of Chauvenet's criterion ($P=0.05$ for a two-sided test) for a sample size 19 is $2.22 < 2.90$. The suspect value 1.0007 is thus rejected.

The new average value and the standard deviation from data of Table 6 (Refinery 1) are calculated, 0.9985 and 0.000525, respectively. Extreme values, in bold in Table 7, are tested using Eq. (4):

Table 7
Meter factors after rejection of the first outlier, Refinery 1.

0.9977	0.9982	0.9985	0.9992
0.9980	0.9982	0.9986	0.9994
0.9980	0.9983	0.9988	0.9996
0.9981	0.9983	0.9988	
0.9981	0.9984	0.9990	

$$r_1 = \frac{0.9977 - 0.9985}{0.000525} = 1.55 \text{ and } r_{18} = \frac{0.9996 - 0.9985}{0.000525} = 2.08$$

The critical value of Chauvenet's criterion ($P=0.05$ for a two-sided test) for a sample size 18 is 2.22. Now, the suspect values (0.9977 and 0.9996) are retained.

4.3.3. Grubbs' test

Similarly to Chauvenet's criterion, the average value and the standard deviation from data of Table 6 (Refinery 1) are calculated, respectively, 0.9986 and 0.00072, for the detection of one outlier, Eq. (5).

$$G_1 = \frac{|0.9977 - 0.9986|}{0.00072} = 1.294 \text{ and } G_{19} = \frac{|1.0007 - 0.9986|}{0.00072} = 2.898$$

The critical value of G ($P=0.05$ for a two-sided test) for a sample size 19 is $2.681 < 2.898$. The suspect value 1.0007 is thus rejected.

The new average value and the standard deviation from data of Table 7 are calculated, 0.9985 and 0.000525, respectively. Extreme values, in bold in Table 7, are tested using Eq. (5):

$$G_1 = \frac{|0.9977 - 0.9985|}{0.000525} = 1.546 \text{ and } G_{18} = \frac{|0.9996 - 0.9985|}{0.000525} = 2.076$$

For sample size 18, the critical value of G is 2.651, ($P=0.05$ for a two-sided test). Since the calculated values of G do not exceed 2.651, the suspect measurements should be retained.

For the detection of two outliers, variances of the data sets are compared based on Eqs. (6) and (7), respectively, $G_{largest}$ ($G_{18}=0.9996$ and $G_{19}=1.0007$) and $G_{smallest}$ ($G_1=0.9977$ and $G_2=0.9980$)

$$G_{largest} = \frac{S_{p-1,p}^2}{S_0^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0.3713 \text{ and}$$

$$G_{smallest} = \frac{S_{1,2}^2}{S_0^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0.8490$$

The critical value of G ($P=0.05$ for a two-sided test) for a sample size 19 is 0.4241. The suspect values ($G_{18}=0.9996$ and $G_{19}=1.0007$) are thus rejected, because $G_{largest}$ is smaller than the critical value.

Once more, for the detection of two outliers, variances of the data sets are compared based on Eqs. (6) and (7), respectively, $G_{largest}$ ($G_{16}=0.9992$ and $G_{17}=0.9994$) and $G_{smallest}$ ($G_1=0.9977$ and $G_2=0.9980$)

$$G_{largest} = \frac{S_{p-1,p}^2}{S_0^2} = \frac{\sum_{i=1}^{p-2} (g_i - \bar{g}_{p-1,p})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0.512 \text{ and}$$

$$G_{smallest} = \frac{S_{1,2}^2}{S_0^2} = \frac{\sum_{i=3}^p (g_i - \bar{g}_{1,2})^2}{\sum_{i=1}^p (g_i - \bar{g})^2} = 0.715$$

The critical value of G ($P=0.05$ for a two-sided test) for a sample size 17 is 0.3822. No outliers are detected because the critical value is smaller than the test statistic.

4.3.4. Median absolute deviation (MAD)

Extreme values, in bold in Table 6, are also tested using Eq. (8): $MAD = \text{median}|x_i - \text{median}(x_i)| = 0.0004$.

For 0.9977: $|x_0 - \text{median}(x_i)|/MAD = |0.9977 - 0.9984|/0.0004 = 1.75$ and for 1.0007: $|x_0 - \text{median}(x_i)|/MAD = |1.0007 - 0.9984|/0.0004 = 5.750$.

The test statistic for the suspect value 1.0007 generates a result superior than 5, rejecting the null hypothesis, so it is considered as an outlier.

New MAD value is calculated from data of Table 7 and the test is carried out again for extreme values, in bold:

$$MAD = \text{median}|x_i - \text{median}(x_i)| = 0.0003.$$

For 0.9977: $|x_0 - \text{median}(x_i)|/MAD = |0.9977 - 0.9984|/0.0003 = 2.167$ and for 0.9996: $|x_0 - \text{median}(x_i)|/MAD = |0.9996 - 0.9984|/0.0003 = 4.167$.

The test statistic for the suspect values generates no results superior than 5, accepting the null hypothesis, thus the suspect values are retained.

4.3.5. Interquartile range (IQR)

Extreme values, in bold in Table 6 (0.9977 and 1.0007), are also tested using Eqs. (9) and (10), based on quartiles: Median = $Q_2 = 0.9984$; $Q_1 = 0.9981$; $Q_3 = 0.9990$ and $Q_3 - Q_1 = 0.0009$. Then, $Q_1 - 1.5 \times (Q_3 - Q_1) = 0.9968$ and $Q_3 + 1.5 \times (Q_3 - Q_1) = 1.0004$.

As the suspect value 1.0007 is out of the interquartile range (0.9968–1.0004), it must be rejected and the test statistic must be repeated from data in Table 7.

The new parameters are: Median = $Q_2 = 0.9984$; $Q_1 = 0.9981$; $Q_3 = 0.9988$ and $Q_3 - Q_1 = 0.00075$. Then, $Q_1 - 1.5 \times (Q_3 - Q_1) = 0.9970$ and $Q_3 + 1.5 \times (Q_3 - Q_1) = 0.9999$.

As the suspect values (0.9977 and 0.9996) are within the interquartile range (0.9970–0.9999), no more outliers are detected.

4.3.6. Comparison

The results of the Refinery 1 m factors of the five approaches discussed (parametric and nonparametric) from 4.3.1 to 4.3.5 are summarized in Table 8, including the expanded uncertainty of the average value, $a(\overline{MF})$, based on Eq. (11). MF and $a(\overline{MF})$ parameters are expressed with one more digit in order to improve the sensitivity to evaluate the results.

Once more, it is noticed that different approaches may exclude different numbers of outliers, resulting different meter factors (average for parametric approaches and median for nonparametric ones) and different average uncertainties.

In this real data set from Refinery 1, except Dixon's test (0.00034), all expanded uncertainty of the average value of a set of meter proving runs ($n=5$) are smaller than the expected value 0.00027.

To this set of data, Grubbs' test shows to be more restrictive to delete two values; however, it is worth remembering that due to the data behavior that does not follow a normal distribution, the most suitable tests are MAD and IQR.

However, it is possible to compare statistically the approach recommended by API MPMS 13.2 (independent of the data behavior), Dixon's test and the nonparametric tests (MAD or IQR). There is no significant difference between them if the absolute value of the difference between the measurements is less or equal to the square root of the sum of the squares of the expanded uncertainties of the central values (average or median), Eq. (12) [23].

$$|MF_1 - MF_2| \leq \sqrt{a(\overline{MF})_1^2 + a(\overline{MF})_2^2} \quad (12)$$

where $MF_1 \pm a(\overline{MF})_1$ is the result of Dixon's test and $MF_2 \pm a(\overline{MF})_2$ is the result of MAD or IQR tests.

Based on Eq. (12): $|0.99863 - 0.99835| \leq \sqrt{0.00034^2 + 0.00026^2}$, i.e., $|0.00028| \leq 0.00043$, there are no significant difference between the calculated meter factors by these different approaches.

Table 8
Comparison of different approaches of the Refinery 1 m factors.

Approach	Numbers of outliers	MF	$a(\overline{MF})$
Dixon	0	0.99863	0.00034
Chauvenet	1	0.99851	0.00026
Grubbs	2	0.99845	0.00024
MAD	1	0.99835	0.00026
IQR	1	0.99835	0.00026

5. Conclusions

This paper statistically evaluated the detection and treating of outliers applied to turbine meter proving factors for custody transfer of liquefied petroleum gas.

This paper recommends that the international guideline reassesses its used example to illustrate the application of Dixon's test outlier and before applying any outlier test, to check the meter proving factor set follows a Gaussian behavior, in other words, a normal distribution. It is noticed that not always data sets of meter factors follow a normal distribution. Therefore, this paper suggests the application of the Shapiro–Wilk test to evaluating the data behavior, prior to detecting and treatment outliers. In situations that data set with behavior that departs from normality, MAD and IQR tests are recommended.

Example B-1. Appendix B, API MPMS 13.2 presents Dixon's test as the only one to detecting and treating outliers. This appendix applies correctly Dixon's test although there have been changes in three meter factors when they were arranged in ascending order of magnitude, generating doubts to the reader in relation to its use. Moreover, contradicting this appendix, alternate computational methods does not always achieve the same purpose intended by the outlier tests, rejecting different MF values.

Dixon's test shows to be limited in some cases by not treating two outliers at the same time as Grubbs' test does. Furthermore, the Chauvenet's criterion can be an excellent alternative to be more restrictive than Grubbs' test for the detection of one outlier. Nevertheless different approaches can numerically reach to different meters factors, however statistically without any significant difference between them.

Finally, this paper also highlights that even after detecting and treating outliers using different approaches (parametric and non-parametric ones), the expanded uncertainty of the average value of a set of meter factors can be above the expected value to the number of meter proving runs, what it is recommend to be explored in future works.

References

- [1] N.C. Surawski, B. Miljevic, T.A. Bodisco, R. Situ, R.J. Brown, Z.D. Ristovski, Performance and gaseous and particle emissions from a liquefied petroleum gas (LPG) fumigated compression ignition engine, *Fuel* 133 (2014) 17–25.
- [2] A.K. Aboul-Gheit, A.E. Awadallah, N.A.K. Aboul-Gheit, D.S. El-Desouki, S. M. Abdel-Hamid, High yields of LPG via n-hexane hydrocracking using unloaded acidic zeolite catalysts, *Pet. Sci. Technol.* 33 (2015) 1314–1321.
- [3] J.G. Pope, J.D. Wright, A.N. Johnson, M.R. Moldover, Extended Lee model for the turbine meter & calibrations with surrogate fluids, *Flow Meas. Instrum.* 24 (2012) 71–82.
- [4] E.C. Oliveira, T.C. Lourenço, Comparison of different approaches to calculate a final meter factor for rotary-type natural gas displacement meters, *Flow Meas. Instrum.* 30 (2013) 160–165.
- [5] Iraq Ministry of Oil, Iraqi National Code for Measurement of Hydrocarbon Fluids, Fiscal & Custody Transfer Measurements, 1st ed., Iraq, 2007.
- [6] Z. Saboohi, S. Sorkhkhah, H. Shakeri, Developing a model for prediction of helical turbine flowmeter performance using CFD, *Flow Meas. Instrum.* 42 (2015) 47–57.
- [7] A.P.I. MPMS Chapter 13.2, Manual of Petroleum Measurement Standards Chapter 13 – Statistical Aspects of Measuring and Sampling, Section 2 – Methods of Evaluating Meter Proving Data, American Petroleum Institute, API Publishing Services, 1220L Street, N.W., Washington, DC, 20005, November 1994, Reaffirmed February 2011.
- [8] A.P.I. MPMS Chapter 13.1, Manual of Petroleum Measurement Standards Chapter 13 – Statistical Aspects of Measuring and Sampling, Section 1 – Statistical Concepts and Procedures in Measurement, American Petroleum Institute, API Publishing Services, 1220L Street, N.W., Washington, DC, 20005, November 1985, Reaffirmed February 2011.
- [9] E.C. Oliveira, et al., Validation of the performance of process stream analyzer systems with nonparametric behavior, *Accredit. Qual. Assur.* 19 (2014) 185–193.
- [10] A. Madansky, *Prescriptions for Working Statisticians*, Springer-Verlag, New York, 1988.
- [11] M.J. Reader-Harris, J.A. Sattary, E.P. Spearman, The orifice plate discharge coefficient equation – further work, *Flow Meas. Instrum.* 6 (1995) 101–114.
- [12] M.J. Reader-Harris, J.A. Sattary, E.P. Spearman, The orifice plate discharge coefficient equation, *Flow Meas. Instrum.* 1 (1990) 67–76.
- [13] D.R. Pipa, M.J. Silva, R.E.M. Moraes, M.V.W. Zibetti, Typical bubble shape estimation in two-phase flow using inverse problem techniques, *Flow Meas. Instrum.* 40 (2014) 64–73.
- [14] G. Ficco, Metrological performance of diaphragm gas meters in distribution networks, *Flow Meas. Instrum.* 37 (2014) 65–72.
- [15] C.E.F. Amaral, et al., Image processing techniques for high-speed videometry in horizontal two-phase slug flows, *Flow Meas. Instrum.* 33 (2013) 257–264.
- [16] D.L. Massart, et al., *Handbook of Chemometrics and Qualimetrics Part A*, Elsevier, Amsterdam, 1997.
- [17] J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 5th ed., Pearson Education Canada, New York, 2005.
- [18] W. Chauvenet, *A Manual of Spherical and Practical Astronomy*, 5th ed., Dover, New York (1960), p. 1863.
- [19] F.E. Grubbs, Procedures for detecting outlying observations in samples, *Technometric* 11 (1969) 1–21.
- [20] D.C. Montgomery, D.C. Runger, *Applied Statistics and Probability for Engineers*, 3rd ed., Wiley, New York, 2003.
- [21] BIPM, IEC, IFCC, ILAC, I.S.O., IUPAC, IUPAP, OIML, *Evaluation of Measurement Data: Guide to the Expression of Uncertainty in Measurement – GUM 2008*, ISBN: 978–85–86920–13–4, 2008.
- [22] NIST/SEMATECH e-Handbook of Statistical Methods, (<http://www.itl.nist.gov/div898/handbook/>) (accessed 21.12.15).
- [23] E.C. Oliveira, Simplified calibration methodology used in custody transfer measurement of natural gas, *Metrolog. Meas. Syst.* 2 (2012) 405–416.

8. Anexo B – Tabelas para testes estatísticos

Tabela 13 – Teste de Shapiro-Wilk: Coeficiente $a_n(W)$ para cálculo de W

$i \backslash n$	2	3	4	5	6	7	8	9	10	11	12	13
1	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6062	0,5888	0,5739	0,5601	0,5475	0,5359
2			0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291	0,3315	0,3325	0,3325
3					0,0875	0,1401	0,1743	0,1976	0,2141	0,2260	0,2347	0,2412
4							0,0561	0,0947	0,1224	0,1429	0,1586	0,1707
5									0,0399	0,0695	0,0922	0,1099
6											0,0303	0,0539
$i \backslash n$	14	15	16	17	18	19	20	21	22	23	24	25
1	0,5251	0,5150	0,5056	0,4968	0,4886	0,4808	0,4734	0,4643	0,4590	0,4542	0,4493	0,4450
2	0,3318	0,3306	0,3290	0,3273	0,3253	0,3232	0,3211	0,3185	0,3156	0,3126	0,3098	0,3069
3	0,2460	0,2495	0,2521	0,2540	0,2553	0,2561	0,2565	0,2578	0,2571	0,2563	0,2554	0,2543
4	0,1802	0,1878	0,1939	0,1988	0,2027	0,2059	0,2085	0,2119	0,2131	0,2139	0,2145	0,2148
5	0,1240	0,1353	0,1447	0,1524	0,1587	0,1641	0,1686	0,1736	0,1764	0,1787	0,1807	0,1822
6	0,0727	0,0880	0,1005	0,1109	0,1197	0,1271	0,1334	0,1399	0,1443	0,1480	0,1512	0,1539
7	0,0240	0,0433	0,0593	0,0725	0,0837	0,0932	0,1013	0,1092	0,115	0,1201	0,1245	0,1283
8			0,0196	0,0359	0,0496	0,0612	0,0711	0,0804	0,0878	0,0941	0,0997	0,1046
9					0,0163	0,0303	0,0422	0,0530	0,0618	0,0696	0,0764	0,0823
10							0,0140	0,0263	0,0368	0,0459	0,0539	0,061
11									0,0122	0,0228	0,0321	0,0403
12											0,0107	0,0200
13												0,0000
$i \backslash n$	26	27	28	29	30	31	32	33	34	35	36	37
1	0,4407	0,4366	0,4328	0,4291	0,4254	0,4220	0,4188	0,4156	0,4127	0,4096	0,4068	0,4040
2	0,3043	0,3018	0,2992	0,2968	0,2944	0,2921	0,2898	0,2876	0,2854	0,2834	0,2813	0,2794
3	0,2533	0,2522	0,2510	0,2499	0,2487	0,2475	0,2463	0,2451	0,2439	0,2427	0,2415	0,2403
4	0,2151	0,2152	0,2151	0,2150	0,2148	0,2145	0,2141	0,2137	0,2132	0,1227	0,2121	0,2116

5	0,1836	0,1848	0,1857	0,1864	0,1870	0,1874	0,1878	0,1880	0,1882	0,1883	0,1883	0,1883	
6	0,1563	0,1584	0,1601	0,1616	0,1630	0,1641	0,1651	0,1660	0,1667	0,1673	0,1678	0,1683	
7	0,1316	0,1346	0,1372	0,1395	0,1415	0,1433	0,1449	0,1463	0,1475	0,1487	0,1496	0,1505	
8	0,1089	0,1128	0,1162	0,1192	0,1219	0,1243	0,1265	0,1284	0,1301	0,1317	0,1331	0,1344	
9	0,0876	0,0923	0,0965	0,1002	0,1036	0,1066	0,1093	0,1118	0,1140	0,1160	0,1179	0,1196	
10	0,0672	0,0728	0,0778	0,0822	0,0862	0,0899	0,0931	0,0961	0,0988	0,1013	0,1036	0,1056	
11	0,0476	0,0540	0,0598	0,065	0,0697	0,0739	0,0777	0,0812	0,0844	0,0873	0,0900	0,0924	
12	0,0284	0,0358	0,0424	0,0483	0,0537	0,0585	0,0629	0,0669	0,0706	0,0739	0,0770	0,0798	
13	0,0094	0,0178	0,0253	0,032	0,0381	0,0435	0,0485	0,0530	0,0572	0,0610	0,0645	0,0677	
14		0,0000	0,0084	0,0159	0,0227	0,0289	0,0344	0,0395	0,0441	0,0484	0,0523	0,0559	
15				0	0,0076	0,0144	0,0206	0,0262	0,0314	0,0361	0,0404	0,0444	
16						0,0000	0,0068	0,0131	0,0187	0,0239	0,0287	0,0331	
17								0,0000	0,0062	0,0119	0,0172	0,0220	
18										0,0000	0,0057	0,0110	
19												0,0000	
i\ñ	38	39	40	41	42	43	44	45	46	47	48	49	50
1	0,4015	0,3989	0,3964	0,3940	0,3917	0,3894	0,3872	0,3850	0,3830	0,3808	0,3789	0,3770	0,3751
2	0,2774	0,2755	0,2737	0,2719	0,2701	0,2684	0,2667	0,2651	0,2635	0,2620	0,2604	0,2589	0,2574
3	0,2391	0,2380	0,2368	0,2357	0,2345	0,2334	0,2323	0,2313	0,2302	0,2291	0,2281	0,2271	0,2260
4	0,2110	0,2104	0,2098	0,2091	0,2085	0,2078	0,2072	0,2065	0,2058	0,2052	0,2045	0,2038	0,2032
5	0,1881	0,1880	0,1878	0,1876	0,1874	0,1871	0,1868	0,1865	0,1862	0,1859	0,1855	0,1851	0,1847
6	0,1686	0,1689	0,1691	0,1693	0,1694	0,1695	0,1695	0,1695	0,1695	0,1695	0,1693	0,1692	0,1691
7	0,1513	0,1520	0,1526	0,1531	0,1535	0,1539	0,1542	0,1545	0,1548	0,1550	0,1551	0,1553	0,1554
8	0,1356	0,1366	0,1376	0,1384	0,1392	0,1398	0,1405	0,1410	0,1415	0,1420	0,1423	0,1427	0,1430
9	0,1211	0,1225	0,1237	0,1249	0,1259	0,1269	0,1278	0,1286	0,1293	0,1300	0,1306	0,1312	0,1317
10	0,1075	0,1092	0,1108	0,1123	0,1136	0,1149	0,1160	0,1170	0,1180	0,1189	0,1197	0,1205	0,1212
11	0,0947	0,0967	0,0986	0,1004	0,1020	0,1035	0,1049	0,1062	0,1073	0,1085	0,1095	0,1105	0,1113
12	0,0824	0,0848	0,0870	0,0891	0,0909	0,0927	0,0943	0,0959	0,0972	0,0986	0,0998	0,1010	0,1020
13	0,0706	0,0733	0,0759	0,0782	0,0804	0,0824	0,0842	0,0860	0,0876	0,0892	0,0906	0,0919	0,0932
14	0,0592	0,0622	0,0651	0,0677	0,0701	0,0724	0,0745	0,0765	0,0783	0,0801	0,0817	0,0832	0,0846
15	0,0481	0,0515	0,0546	0,0575	0,0602	0,0628	0,0651	0,0673	0,0694	0,0713	0,0731	0,0748	0,0764

16	0,0372	0,0409	0,0444	0,0476	0,0506	0,0534	0,0560	0,0584	0,0607	0,0628	0,0648	0,0667	0,0685
17	0,0264	0,0305	0,0343	0,0379	0,0411	0,0442	0,0471	0,0497	0,0522	0,0546	0,0568	0,0588	0,0608
18	0,0158	0,0203	0,0244	0,0283	0,0318	0,0352	0,0383	0,0412	0,0439	0,0465	0,0489	0,0511	0,0532
19	0,0053	0,0101	0,0146	0,0188	0,0227	0,0263	0,0296	0,0328	0,0357	0,0385	0,0411	0,0436	0,0459
20		0,0000	0,0049	0,0094	0,0136	0,0175	0,0211	0,0245	0,0277	0,0307	0,0335	0,0361	0,0386
21				0,0000	0,0045	0,0087	0,0126	0,0163	0,0197	0,0229	0,0259	0,0288	0,0314
22						0,0000	0,0042	0,0081	0,0118	0,0153	0,0185	0,0215	0,0244
23								0,0000	0,0039	0,0076	0,0111	0,0143	0,0174
24										0,0000	0,0037	0,0071	0,0104
25												0,0000	0,0350

Tabela 14 – Valores críticos da estatística W de Shapiro-Wilk

N	Nível de significância								
	0,01	0,02	0,05	0,1	0,5	0,9	0,95	0,98	0,99
3	0,753	0,756	0,767	0,789	0,959	0,998	0,999	1,000	1,000
4	0,687	0,707	0,748	0,792	0,935	0,987	0,992	0,996	0,997
5	0,686	0,715	0,762	0,806	0,927	0,979	0,986	0,991	0,993
6	0,713	0,743	0,788	0,826	0,927	0,974	0,981	0,986	0,989
7	0,730	0,760	0,803	0,838	0,928	0,972	0,979	0,985	0,988
8	0,749	0,778	0,818	0,851	0,932	0,972	0,978	0,984	0,987
9	0,764	0,791	0,829	0,859	0,935	0,972	0,978	0,984	0,986
10	0,781	0,806	0,842	0,869	0,938	0,972	0,978	0,983	0,986
11	0,792	0,817	0,850	0,876	0,940	0,973	0,979	0,984	0,986
12	0,805	0,828	0,859	0,883	0,943	0,973	0,979	0,984	0,986
13	0,814	0,837	0,866	0,889	0,945	0,974	0,979	0,984	0,986
14	0,825	0,846	0,874	0,895	0,947	0,975	0,980	0,984	0,986
15	0,835	0,855	0,881	0,901	0,950	0,975	0,980	0,984	0,987
16	0,844	0,863	0,887	0,906	0,952	0,976	0,981	0,985	0,987
17	0,851	0,869	0,892	0,910	0,954	0,977	0,981	0,985	0,987
18	0,858	0,874	0,897	0,914	0,956	0,978	0,982	0,986	0,988
19	0,863	0,879	0,901	0,917	0,957	0,978	0,982	0,986	0,988
20	0,868	0,884	0,905	0,920	0,959	0,979	0,983	0,986	0,988
21	0,873	0,888	0,908	0,923	0,960	0,980	0,983	0,987	0,989
22	0,878	0,892	0,911	0,926	0,961	0,980	0,984	0,987	0,989
23	0,881	0,895	0,914	0,928	0,962	0,981	0,984	0,987	0,989
24	0,884	0,898	0,916	0,930	0,963	0,981	0,984	0,987	0,989
25	0,888	0,901	0,918	0,931	0,964	0,981	0,985	0,988	0,989
26	0,891	0,904	0,920	0,933	0,965	0,982	0,985	0,988	0,989
27	0,894	0,906	0,923	0,935	0,965	0,982	0,985	0,988	0,990
28	0,896	0,908	0,924	0,936	0,966	0,982	0,985	0,988	0,990
29	0,898	0,910	0,926	0,937	0,966	0,982	0,985	0,988	0,990
30	0,900	0,912	0,927	0,939	0,967	0,983	0,985	0,988	0,990
31	0,902	0,914	0,929	0,940	0,967	0,983	0,986	0,988	0,990
32	0,904	0,915	0,930	0,941	0,968	0,983	0,986	0,988	0,990

33	0,906	0,917	0,931	0,942	0,968	0,983	0,986	0,989	0,990
34	0,908	0,919	0,933	0,943	0,969	0,983	0,986	0,989	0,990
35	0,910	0,920	0,934	0,944	0,969	0,984	0,986	0,989	0,990
36	0,912	0,922	0,935	0,945	0,970	0,984	0,986	0,989	0,990
37	0,914	0,924	0,936	0,946	0,970	0,984	0,987	0,989	0,990
38	0,916	0,925	0,938	0,947	0,971	0,984	0,987	0,989	0,990
39	0,917	0,927	0,939	0,948	0,971	0,984	0,987	0,989	0,991
40	0,919	0,928	0,940	0,949	0,972	0,985	0,987	0,989	0,991
41	0,920	0,929	0,941	0,950	0,972	0,985	0,987	0,989	0,991
42	0,922	0,930	0,942	0,951	0,972	0,985	0,987	0,989	0,991
43	0,923	0,932	0,943	0,951	0,973	0,985	0,987	0,990	0,991
44	0,924	0,933	0,944	0,952	0,973	0,985	0,987	0,990	0,991
45	0,926	0,934	0,945	0,953	0,973	0,985	0,988	0,990	0,991
46	0,927	0,935	0,945	0,953	0,974	0,985	0,988	0,990	0,991
47	0,928	0,936	0,946	0,954	0,974	0,985	0,988	0,990	0,991
48	0,929	0,937	0,947	0,954	0,974	0,985	0,988	0,990	0,991
49	0,929	0,938	0,947	0,955	0,974	0,985	0,988	0,990	0,991
50	0,930	0,939	0,947	0,955	0,974	0,985	0,988	0,990	0,991

Fonte: PORTAL ACTION. Disponível em: <http://www.portalaction.com.br/inferencia/64-teste-de-shapiro-wilk>. Acesso em: 28 jun. 2017.

Tabela 15 – Valores críticos para o teste de Dixon

n		Bi caudal, α	
		0,05	0,01
3	Q ₁₀	0,970	0,994
4		0,829	0,926
5		0,710	0,821
6		0,628	0,740
7		0,569	0,680
8	Q ₁₁	0,608	0,717
9		0,564	0,672
10		0,530	0,635
11		0,502	0,605
12		0,479	0,579
13	Q ₂₂	0,611	0,697
14		0,586	0,670
15		0,565	0,647
16		0,546	0,627
17		0,529	0,610
18		0,514	0,594
19		0,501	0,580
20		0,489	0,567

Fonte: Tabela adaptada (Massart et al., 1997).

Tabela 16 – Critério de Chauvenet para rejeição de valor medido

n	d_{max} / σ	n	d_{max} / σ
1	-	16	2,16
2	-	17	2,18
3	1,38	18	2,20
4	1,54	19	2,22
5	1,65	20	2,24
6	1,73	21	2,26
7	1,80	22	2,28
8	1,85	23	2,30
9	1,91	24	2,31
10	1,96	25	2,33
11	1,99	26	2,35
12	2,03	27	2,36
13	2,06	28	2,37
14	2,10	29	2,38
15	2,13	30	2,39

Fonte: Tabela adaptada. SCRIBD: banco de dados. Disponível em:
 <<https://pt.scribd.com/document/22992713/Tabela-Student-Chauvenet>>.

Tabela 17 – Valores críticos para o teste de Grubbs

p	Um valor		Dois valores		p	Um valor		Dois valores	
	1 %	5 %	1 %	5 %		1 %	5 %	1 %	5 %
-	-	-	-	-	22	3,060	2,758	0,3927	0,4711
3	1,155	1,155	-	-	23	3,087	2,781	0,4085	0,4857
4	1,496	1,481	0,0000	0,0002	24	3,112	2,802	0,4234	0,4994
5	1,764	1,715	0,0018	0,0090	25	3,135	2,822	0,4376	0,5123
6	1,973	1,887	0,0116	0,0349	26	3,157	2,841	0,4510	0,5245
7	2,139	2,020	0,0308	0,0708	27	3,178	2,859	0,4638	0,5360
8	2,274	2,126	0,0563	0,1101	28	3,199	2,876	0,4759	0,5470
9	2,387	2,215	0,0851	0,1492	29	3,218	2,893	0,4875	0,5574
10	2,482	2,290	0,1150	0,1864	30	3,236	2,908	0,4985	0,5672
11	2,564	2,355	0,1448	0,2213	31	3,253	2,924	0,5091	0,5766
12	2,636	2,412	0,1738	0,2537	32	3,270	2,938	0,5192	0,5856
13	2,699	2,462	0,2016	0,2836	33	3,286	2,952	0,5288	0,5941
14	2,755	2,507	0,2280	0,3112	34	3,301	2,965	0,5381	0,6023
15	2,806	2,549	0,2530	0,3367	35	3,316	2,979	0,5469	0,6101
16	2,852	2,585	0,2767	0,3603	36	3,330	2,991	0,5554	0,6175
17	2,894	2,620	0,2990	0,3822	37	3,343	3,003	0,5636	0,6247
18	2,932	2,651	0,3200	0,4025	38	3,356	3,014	0,5714	0,6316
19	2,968	2,681	0,3398	0,4214	39	3,369	3,025	0,5789	0,6382
20	3,001	2,709	0,3585	0,4391	40	3,381	3,036	0,5862	0,6445
21	3,031	2,733	0,3761	0,4556	-	-	-	-	-

Fonte: Tabela adaptada (Massart et al.,1997).

Tabela 18 – Tabela t-Student

Graus de liberdade v	Frações p em porcentagem					
	68,27 ^(a)	90	95	95,45 ^(a)	99	99,73 ^(a)
1	1,84	6,31	12,71	13,97	63,66	235,80
2	1,32	2,92	4,30	4,53	9,92	19,21
3	1,20	2,35	3,18	3,31	5,84	9,22
4	1,14	2,13	2,78	2,87	4,60	6,62
5	1,11	2,02	2,57	2,65	4,03	5,51
6	1,09	1,94	2,45	2,52	3,71	4,90
7	1,08	1,89	2,36	2,43	3,50	4,53
8	1,07	1,86	2,31	2,37	3,36	4,28
9	1,06	1,83	2,26	2,32	3,25	4,09
10	1,05	1,81	2,23	2,28	3,17	3,96
11	1,05	1,80	2,20	2,25	3,11	3,85
12	1,04	1,78	2,18	2,23	3,05	3,76
13	1,04	1,77	2,16	2,21	3,01	3,69
14	1,04	1,76	2,14	2,20	2,98	3,64
15	1,03	1,75	2,13	2,18	2,95	3,59
16	1,03	1,75	2,12	2,17	2,92	3,54
17	1,03	1,74	2,11	2,16	2,90	3,51
18	1,03	1,73	2,10	2,15	2,88	3,48
19	1,03	1,73	2,09	2,14	2,96	3,45
20	1,03	1,72	2,09	2,13	2,85	3,42
25	1,02	1,71	2,06	2,11	2,79	3,33
30	1,02	1,70	2,04	2,09	2,75	3,27
35	1,01	1,70	2,03	2,07	2,72	3,23
40	1,01	1,68	2,02	2,06	2,70	3,20
45	1,01	1,68	2,01	2,06	2,69	3,18
50	1,01	1,68	2,01	2,05	2,68	3,16
100	1,005	1,660	1,984	2,025	2,626	3,077
∞	1,000	1,645	1,960	2,000	2,576	3,000

^(a) Para a grandeza z descrita por uma distribuição normal, com esperança μ_2 e desvio padrão σ , o intervalo $\mu_2 \pm k\sigma$ abrange $p = 68,27, 95,45$ e $99,73$ por cento da distribuição para $k = 1, 2$ e 3 , respectivamente.

Fonte: GUM, 2008.

Tabela 19 – Fator de conversão para estimar o desvio padrão para n pontos a partir da amplitude

n	d_2
2	1,128
3	1,693
4	2,059
5	2,326
6	2,534
7	2,704
8	2,847
9	2,970
10	3,078
11	3,173
12	3,258
13	3,336
14	3,407
15	3,472
16	3,532
17	3,588
18	3,640
19	3,689
20	3,765

Fonte: Montgomery; Runger, 2003