

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

**DESCRIÇÃO E COMPARAÇÃO DE METODOLOGIAS  
ESTATÍSTICAS UTILIZADAS EM ENSAIOS DE PROFICIÊNCIA**

**GENÍLSON DE SOUZA PACHECO**

Porto Alegre  
2019

**GENÍLSON DE SOUZA PACHECO**

**DESCRIÇÃO E COMPARAÇÃO DE METODOLOGIAS  
ESTATÍSTICAS UTILIZADAS EM ENSAIOS DE PROFICIÊNCIA**

Trabalho de Conclusão do Curso para  
obtenção do grau de Bacharel em Estatística

Orientador  
Doutor Danilo Marcondes Filho

Porto Alegre  
2019

## CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Pacheco, Genílson de Souza  
Descrição e Comparação de Metodologias Estatísticas utilizadas  
em Ensaio de Proficiência / Genílson de Souza Pacheco, - Porto  
Alegre: Departamento de Estatística / Instituto de Matemática e  
Estatística, 2019.

Nº de páginas: 45il.

Orientador: Doutor Danilo Marcondes Filho.

Trabalho de Conclusão (Graduação) - UFRGS, Instituto de  
Matemática e Estatística, Curso de Bacharel em Estatística, Porto  
Alegre, 2019.

1. Ensaio de Proficiência. 2. ISO 13528. 3. ASTM E691. 4.  
Média Robusta. 5. Desvio Padrão Robusto. I. Marcondes Filho,  
Danilo, orient. II. Doutor.

Instituto de Matemática e Estatística  
Departamento de Estatística

**DESCRIÇÃO E COMPARAÇÃO DE METODOLOGIAS  
ESTATÍSTICAS UTILIZADAS EM ENSAIOS DE PROFICIÊNCIA**

GENÍLSON DE SOUZA PACHECO

Banca examinadora:

Doutor Danilo Marcondes Filho (orientador)  
Universidade Federal do Rio Grande do Sul

Doutora Liane Werner  
Universidade Federal do Rio Grande do Sul

## **AGRADECIMENTOS**

Agradeço a todas as pessoas que de uma forma ou outra acompanharam minha trajetória e meu aprendizado na graduação em Estatística e que torceram por mim para concluir essa etapa da minha vida.

Em especial a meus filhos Eduarda e Miguel, foram meu maior incentivo para chegar até aqui, agradeço por entender os momentos de minha ausência, por sempre me apoiarem nas minhas escolhas e pelo incentivo incondicional.

Ao meu irmão Jeverson, que sempre fez o possível e o impossível para possibilitar meus estudos desde minha adolescência.

A Diretora do ISI Engenharia de Polímeros Viviane Lovison, que sempre me motivou e foi muito paciente em momentos de minha ausência.

Ao Professor Danilo, por todo aprendizado, motivação, paciência e auxílio na orientação deste trabalho.

Obrigado a Deus e a todos por tudo, sem vocês nada disso seria possível.

## **DEDICATÓRIA**

Dedico este trabalho a minha querida mãe Geniria, que me educou de forma exemplar, não poupando esforços junto com meu irmão para garantir o sustendo e a felicidade da família. Dedico também ao meu pai Genino que junto com minha mãe tinham muito orgulho de seu filho estar estudando em uma universidade federal, mas infelizmente não estarão presentes fisicamente para presenciar este momento tão importante.

## RESUMO

Os Ensaio de Proficiências (EPs) são mundialmente conhecidos na área da metrologia como a ferramenta mais completa para a verificação da qualidade e confiabilidade das medidas de uma mesma característica de qualidade por diferentes participantes. As normas ISO 13528 e ASTM E691 descrevem as metodologias estatísticas mais utilizadas nos EPs. Entretanto, não há obrigatoriedade de escolha de qualquer uma delas, dando a ambas o mesmo valor na avaliação de um participante envolvido no estudo de EP. Trabalhos apresentando estudos comparativos dos procedimentos estatísticos presentes nas normas ISO e ASTM são escassos na literatura. Neste contexto, este trabalho descreve detalhadamente metodologias estatísticas utilizadas em ambas as normas, bem como apresenta um estudo comparativo da ISO e ASTM a partir da análise de 27 estudos de casos envolvendo EPs.

**Palavras-chave:** Ensaio de Proficiência. ISO 13528. ASTM E691. Média Robusta. Desvio Padrão Robusto.

## ABSTRACT

Proficiency Testing (PTs) are known worldwide in the field of metrology as the most complete tool for verifying the quality and reliability of measurements of the same quality characteristic by different participants (industries). ISO 13528 and ASTM E691 describe the statistical methodologies most commonly used in PTs. However, there is no obligation to choose any of them, that is, both have the same value in the certification of a participant involved in the PT study. Studies presenting comparative studies of the statistical procedures present in the ISO and ASTM standards are scarce in the literature. In this context, this work describes in detail the statistical methodologies used in both standards, as well as presents a comparative study of ISO and ASTM through the analysis of 27 case studies involving PTs.

**Keywords:** Proficiency Testing. ISO 13528. ASTM E691. Robust Mean. Robust Standard Deviation.

## LISTA DE FIGURAS

Figura 1 - Fluxograma básico para execução de um Ensaio de Proficiência.....	16
Figura 2 - Fluxograma da metodologia de cálculo da norma ISO13528.....	20
Figura 3 - Fluxograma da metodologia de cálculo da norma ASTM E691.....	25
Figura 4 - Distribuição dos valores dos coeficientes de Kappa nos EPs.....	38
Figura 5 - Região da mediana do desvio absoluto (MAD).....	42

## LISTA DE TABELAS

Tabela 1 - Modelo de Tabela para registro dos resultados dos participantes.....	17
Tabela 2 - Modelo de Tabela utilizada no Teste de Wilcoxon.....	30
Tabela 3 - Modelo de Tabela de contingência utilizada para o coeficiente de Kappa....	32
Tabela 4 - Faixa de classificação de Kappa.....	33
Tabela 5 - EPs agrupados por variáveis.....	35
Tabela 6 - Resultados dos EPs e das análises estatísticas.....	37
Tabela 7 - Comparação das avaliações de desempenho dos participantes.....	38
Tabela 8 - Comparação das avaliações de desempenho dos participantes acumulada.....	39
Tabela 9 - Valores de $\beta$ e $\tau$ .....	43

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>12</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b> .....	<b>14</b>
<b>2.1</b>	<b>Estruturação de um Ensaio de Proficiência (EP)</b> .....	<b>14</b>
<b>2.2</b>	<b>Norma ISO 13528</b> .....	<b>18</b>
<b>2.3</b>	<b>Norma ASTM E691</b> .....	<b>23</b>
<b>2.4</b>	<b>Técnicas estatísticas utilizadas na Metodologia</b> .....	<b>30</b>
2.4.1	Teste de Wilcoxon para amostras pareadas.....	30
2.4.2	Coeficiente de concordância de Kappa.....	32
<b>3</b>	<b>METODOLOGIA</b> .....	<b>34</b>
<b>3.1.</b>	<b>Banco de dados</b> .....	<b>34</b>
<b>3.2.</b>	<b>Análise Comparativa</b> .....	<b>35</b>
<b>4</b>	<b>ESTUDO DE CASO</b> .....	<b>36</b>
<b>4.1</b>	<b>Comparação entre estimativas das médias e dos desvios padrão</b> .....	<b>36</b>
<b>4.2</b>	<b>Comparação da avaliação de desempenho dos participantes</b> .....	<b>38</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>40</b>
	<b>REFERÊNCIAS</b> .....	<b>41</b>
	<b>ANEXO A</b> .....	<b>42</b>

## 1. INTRODUÇÃO

Desde o início da humanidade existiu a necessidade da realização de medições, como por exemplo nas pirâmides egípcias, construções gregas e romanas, até chegar na atualidade. Em todos esses casos, sempre estavam presentes sistemas de medidas compatíveis com sua época. Logo o homem percebeu que não era necessário somente medir, mas que suas medidas deveriam coincidir com as dos outros homens. Para verificar a diferença entre medições realizadas por diferentes pessoas, empresas, laboratórios, equipamentos e/ou outras fontes, utiliza-se as comparações interlaboratoriais. Entende-se por Comparação Interlaboratorial a organização, realização e avaliação de medições ou ensaios nos mesmos ou em itens similares por dois ou mais laboratórios, de acordo com as condições predeterminadas (ABNT NBR ISO/IEC 17043, 2011). As comparações interlaboratoriais com a inclusão da etapa de avaliação de desempenho dos participantes são chamadas de Ensaio de Proficiência (EP). Os EPs avaliam o desempenho do participante contra critérios preestabelecidos por meio de comparações interlaboratoriais (ABNT NBR ISO/IEC 17043, 2011).

No início da década de 80 com o acontecimento da globalização mundial (processo de interação entre as nações, proporcionado pelos progressos dos meios de transportes e telecomunicações), uma medida realizada na China por exemplo, deveria ser compatível com uma medida semelhante realizada nos USA, por este motivo os principais Institutos Nacionais de Metrologias (USA, Inglaterra, Alemanha, Japão, Austrália, etc...) iniciaram um forte processo de reconhecimento mútuo através da utilização de EPs. No Brasil, a globalização iniciou fortemente nos anos 90, quando o INMETRO<sup>1</sup> começa um importante trabalho de quebra de barreira técnicas para possibilitar a entrada dos produtos nacionais em outros países, através do reconhecimento por parte destes, de que as medidas realizadas no Brasil eram compatíveis com as realizadas pelos mesmos. O reconhecimento das medidas nacionais requereu por parte do INMETRO a exigência que sua rede de laboratórios acreditados participasse obrigatoriamente e periodicamente de EPs. Historicamente o primeiro setor de forma compulsória a realizar comparações interlaboratoriais foram os laboratórios de

---

<sup>1</sup> INMETRO: Instituto Nacional de Metrologia, Qualidade e Tecnologia

medicina (THOLEN, 2004).

Os requisitos para a realização de um EP estão definidos na norma ABNT NBR ISO/IEC 17043 “Avaliação de Conformidade - Requisitos gerais para ensaios de proficiência”. A mesma requisita que devem ser utilizadas técnicas estatísticas para análise dos resultados dos participantes, porém não define um método/norma específico. Dessa forma, os provedores, entidades que organizam os EPs, utilizam técnicas estatísticas variadas que apresentam resultados diferentes. Para tal, as normas mais utilizadas são:

- ISO<sup>2</sup> 13528 *Statistical methods for use in proficiency testing by interlaboratory comparison*
- ASTM<sup>3</sup> E691 *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method.*

Este trabalho tem como objetivos descrever de maneira detalhada as metodologias estatísticas utilizadas nas normas ISO 13528 e ASTM E691, bem como apresenta um estudo comparativo de ambas as normas a partir de estudos de casos contendo dados de participantes de EP.

O trabalho de conclusão está organizado em suas seções da seguinte forma: A seção 2 apresenta uma revisão das Normas ISO 13528 e ASTM E691 é apresentada com foco no detalhamento da metodologia estatística empregada para realização do estudo de EP. A seção 3 apresenta a metodologia aplicada no presente trabalho. A seção 4 apresenta análise dos estudos de casos apresentados através da metodologia descrita na anterior. A seção 6 apresenta as conclusões.

---

<sup>2</sup> ISO: International Organization for Standardization

<sup>3</sup> ASTM: American Society for Testing and Materials

## **2. REVISÃO BIBLIOGRÁFICA**

Esta seção inicialmente descreve a estrutura para execução de um ensaio de proficiência. Na sequência, as duas principais normas para análise estatística de ensaios de proficiência (as normas ISO 13528 e ASTM E691) são apresentadas com foco no detalhamento da metodologia estatística empregada.

### **2.1. Estrutura de um Ensaio de Proficiência (EP)**

O objetivo principal de um EP é a avaliação de desempenho dos participantes, para isso todas as etapas que influenciam nesta avaliação devem ser planejadas e cuidadosamente executadas. O EP é organizado por um órgão chamado “Provedor do Ensaio de Proficiência”, que é o responsável por todas as etapas de execução do EP. O provedor do EP deve atender aos requisitos da norma ABNT NBR ISO/IEC 17043 “Avaliação da conformidade – Requisitos gerais para ensaios de proficiência”, a mesma apresenta todas as etapas que devem ser controladas e executadas de modo a garantir a qualidade e confiabilidade de um EP.

As principais etapas executadas pelo provedor e seus participantes para a realização de um EP são as seguintes:

- Planejamento: o provedor planeja as etapas que influenciam na qualidade do EP, onde são definidos o cronograma, as medições e as amostras incluídas no EP.
- Divulgação: o provedor realiza a divulgação do EP, onde são informadas as medições incluídas, o cronograma, os critérios e custo para participação do EP.
- Inscrição dos participantes: o provedor realiza a listagem dos participantes em cada medição incluída no EP.
- Preparação das amostras: o provedor prepara as amostras de acordo com as medições incluídas no EP e o número de participantes inscritos.
- Homogeneidade das amostras: o provedor realiza as medições e análise estatísticas nas amostras a serem utilizadas no EP, para verificar se as mesmas são homogêneas de modo que não influenciem significativamente nos resultados dos participantes.

- Envio das Amostras: o provedor embala adequadamente as amostras e as envia todos os participantes do EP.
- Recebimento das amostras: o participante verifica se as amostras estão em condições adequadas para a realização das medições do EP.
- Realização das medições: o participante realiza as medições nas amostras enviadas, seguindo um método de medição pré-definido.
- Envio dos resultados: após a realização dos ensaios o participante deve enviar ao provedor os resultados de suas medições nas amostras no EP.
- Recebimento dos resultados: a provedor recebe os resultados das medições dos participantes.
- Tabulação dos resultados: o provedor organiza os resultados das medições dos participantes do EP conforme Tabela 1.
- Análise estatística dos resultados: o provedor realiza a análise estatística dos resultados dos ensaios enviados pelos participantes, conforme alguma norma de referência (ISO 13528, ASTM E691, etc...).
- Emissão do Relatório Final: o provedor emite o Relatório Final com os resultados das medições e a avaliação de desempenho de cada participante no EP.
- Envio do Relatório Final: o provedor envia o Relatório Final a todos participantes do EP.
- Recebimento do Relatório Final: o participante recebe o Relatório Final do EP.
- Análise do Relatório Final: o participante realiza a análise crítica dos resultados de suas medições e de sua avaliação de desempenho no EP.

O fluxograma básico para execução de um EP atendendo aos requisitos da ABNT NBR ISO/IEC 17043 é apresentado na figura 1.

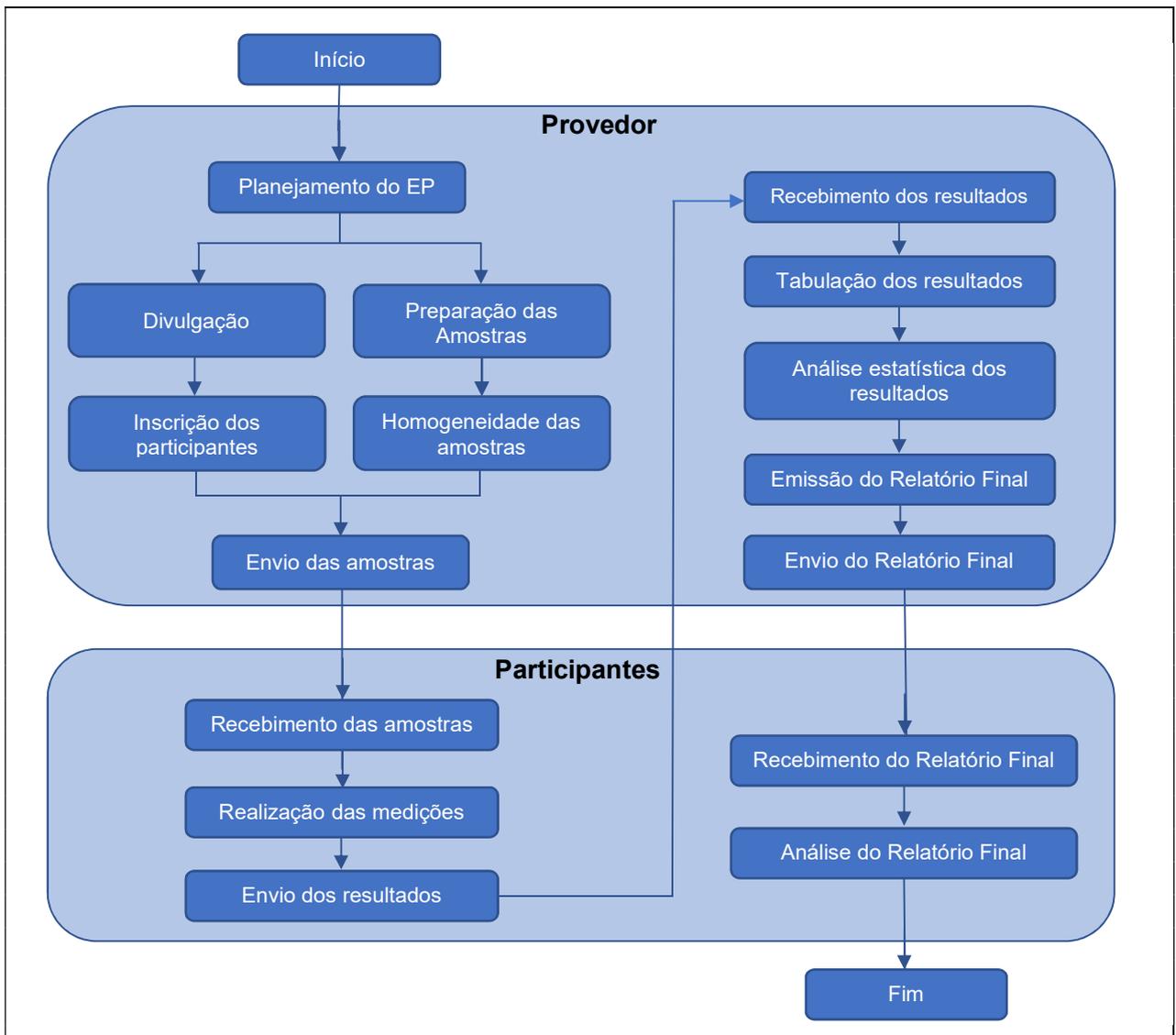


Figura 1 – Fluxograma básico para execução de um Ensaio de Proficiência

O modelo matemático básico de um EP implica que valor verdadeiro da medida ( $\mu$ ) pode ser estimado a partir dos resultados dos participantes remanescentes após a análise prévia de medidas discrepantes. Consideramos as observações  $x_{ri}$  possuindo distribuição Normal com média  $\mu$  e variância  $\sigma^2$ , isto é,  $x_{ri} \sim N(\mu, \sigma^2)$ . Decorre então que  $\varepsilon_i \sim N(0, \sigma^2)$ . Assim a média  $\bar{x}_i$  dos resultados do participante seria uma estimativa aproximada do valor verdadeiro da medição. Quando  $\bar{x}_i$  possui valores discrepantes (“outlier”), cada uma das metodologias para estudos de EP possuem critérios próprios

para obter um valor de consenso mais confiável para a avaliação de desempenho subsequente das amostras (WONG, 2018).

Considerando um grupo de  $p$  participantes de uma EP, onde cada um gera  $n$  medições da característica de qualidade de interesse. O modelo matemático básico de um EP é apresentado abaixo:

$$x_{ri} = \mu + B_i + \varepsilon_i, \text{ para } i = 1, \dots, p \quad (1)$$

onde:  $x_{ri}$  é a  $r$ -ésima medição do  $i$ -ésimo participante,

$\mu$  é o valor “verdadeiro” da medida,

$B_i$  é o componente de tendência do  $i$ -ésimo participante,

$\varepsilon_i$  é o erro de medição do  $i$ -ésimo participante,

$n$  é o número de repetições realizadas por cada participante.

A Tabela 1 apresenta a estrutura genérica das medições enviadas pelos participantes.

Valores discrepantes (“outlier”) é a observação em um conjunto de dados que parece incompatível com o restante conjunto de dados (ABNT NBR ISO/IEC 17043, 2011).

Tabela 1 - Modelo de Tabela para registro dos resultados dos participantes

Repetições	Participantes				
	1	...	$i$	...	$p$
1	$x_{11}$	...	$x_{1i}$	...	$x_{1p}$
.	.	.	.	.	.
.	.	.	.	.	.
$r$	$x_{r1}$	...	$x_{ri}$	...	$x_{rp}$
.	.	.	.	.	.
.	.	.	.	.	.
$n$	$x_{n1}$	...	$x_{ni}$	...	$x_{np}$
Média ( $\bar{x}$ )	$\bar{x}_1$	...	$\bar{x}_i$	...	$\bar{x}_p$

A partir das médias  $\bar{x}_i$  de cada participante, obtém-se uma estimativa do valor “verdadeiro”  $\mu$  da medida através da grande média dos participantes, a qual denotamos por  $\bar{\bar{x}}$ . Também são computados para o estudo uma medida de desvio padrão amostral ( $s$ ) das medições, a qual é considerada para estimar  $\sigma$ .

A metodologias ISO e ASTM que serão descritas utilizam as estatísticas apresentadas acima, porém de diferenciam em vários aspectos, utilizando diferentes procedimentos para o cálculo das estimativas de  $\mu$  e de  $\sigma$ , para detecção de outliers e para construção de regras para comparação dos resultados dos participantes.

### 3.2. Norma ISO 13528

A norma internacional ISO 13528 “*Statistical methods for use in proficiency testing by interlaboratory comparisons*” é um documento de referência importante para provedores de EPs, que fornece orientação sobre o uso de métodos estatísticos para EPs.

A primeira edição da ISO 13528 foi publicada em 2005 como um complemento ao então ISO/IEC Guide 43- Parte 1:1997 “*Proficiency testing by interlaboratory comparisons – Part 1: Development and operation of proficiency testing schemes*”, norma que define os requisitos para organização de EPs. Em 2010, a ISO/IEC 17043 foi publicada para substituir as partes I e II do ISO/IEC Guide 43:1997, que atualizou os princípios para a operação de EPs. A segunda e última edição da ISO 13528 de setembro de 2015 após uma revisão técnica da versão anterior, bem como para harmonizar o documento com a ISO/IEC 17043:2010. Em suma, a nova edição se concentra no projeto e na análise estatística de Ensaios de Proficiência. Novas seções também foram adicionadas para abranger procedimentos para ensaios qualitativos de testes de proficiência.

A metodologia estatística definida para analisar os resultados dos participantes consiste na utilização da média individual do participante ( $\bar{x}_i$ ), a “grande média” ( $\bar{\bar{x}}$ ) e a variação da grande média dos participantes (o desvio padrão  $s$ ), de modo permitir determinar o quanto o resultado de um participante se aproxima de um “valor de consenso”.

Numa primeira fase do estudo de EP, a norma ISO utiliza um procedimento visando minimizar a influência de medições atípicas que irão distorcer a avaliação do participante do EP. Os resultados de média e de desvio padrão são sensíveis à presença de valores discrepantes. Para diminuir estes efeitos a norma ISO 13528 define a utilização do Algoritmo A, que é um procedimento baseado em estatísticas descritivas para substituição dos valores discrepantes, desta forma minimizando a influência que estes podem ter sobre estimativas de média e desvio padrão. Este procedimento se repete iterativamente para as medições que permaneceram no estudo até que nenhuma nova medição seja classificada como outlier. A cada nova rodada os valores das estatísticas descritivas (média do participante, grande média e desvio padrão) são atualizados.

A Figura 2 apresenta o fluxograma resumido da metodologia da ISO 13528, que por sua vez será descrita nos parágrafos subsequentes.

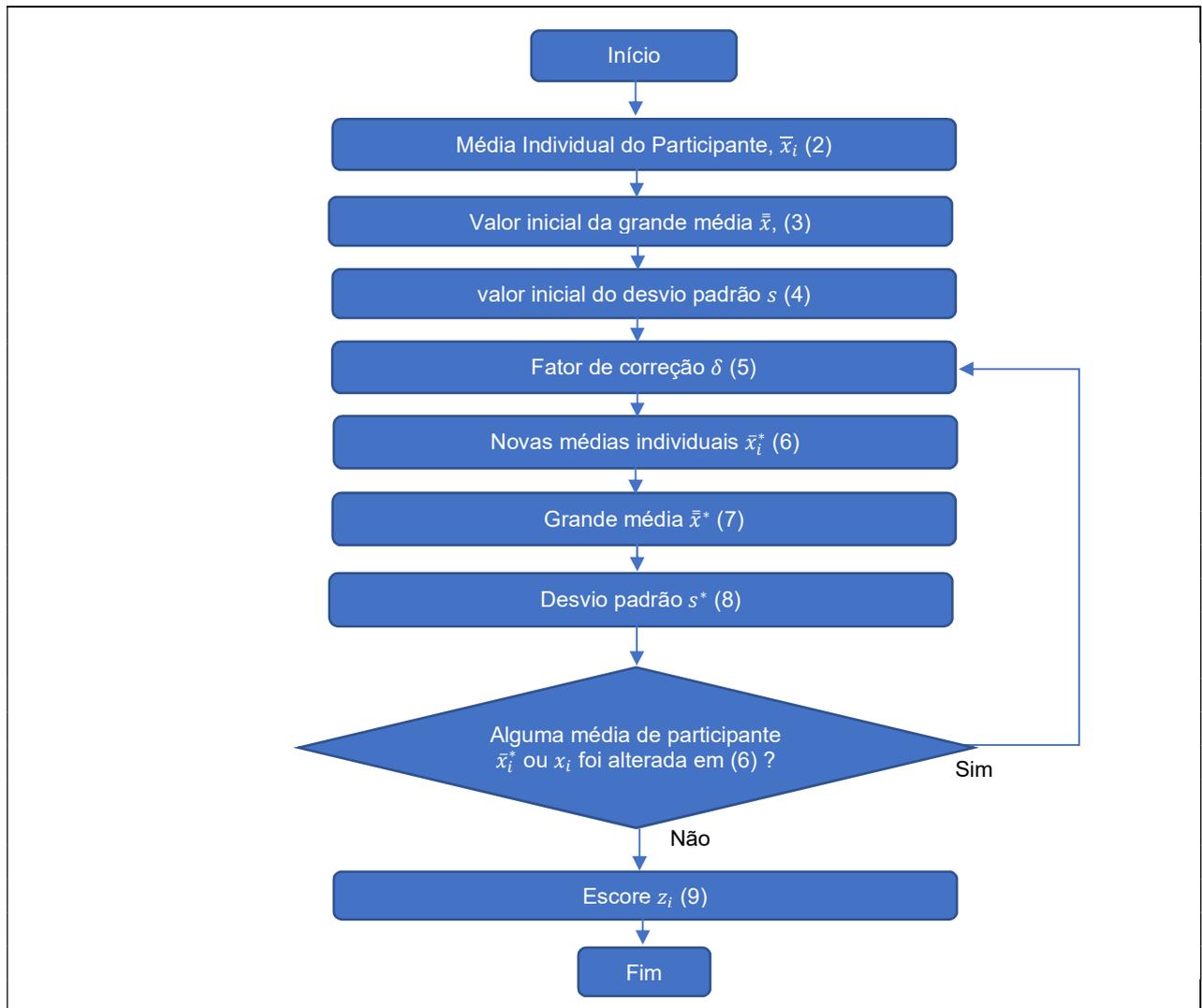


Figura 2 – Fluxograma da metodologia de cálculo da norma ISO 13528

Numa segunda fase do estudo de EP, após a retirada das observações discrepantes, os valores finais das estatísticas descritivas (média do participante, grande média e desvio padrão) são obtidos e utilizadas para construir uma regra de classificação das medições dos participantes.

Considere  $p$  participantes, cada um enviando  $n$  mensurações, obtemos os dados estruturados como na Tabela 1. O algoritmo A definido na norma ISO 13528 é detalhado a seguir:

a) Obter as médias  $\bar{x}_i$  de cada participante da seguinte forma:

$$\bar{x}_i = \frac{\sum_{r=1}^n x_{ri}}{n} \quad (2)$$

Onde  $x_{ri}$  é a  $r$ -ésima medição fornecida pelo  $i$ -ésimo participante  
 $n$  é o número de repetições realizadas por cada participante.

b) Obter a estimativa inicial da grande média  $\bar{\bar{x}}$ , da seguinte forma:

$$\bar{\bar{x}} = \text{mediana}(\bar{x}_i) \quad (3)$$

c) Obter a estimativa inicial do desvio padrão  $s$  da seguinte forma:

$$s = \alpha * MAD \quad , \text{ onde } MAD = \text{mediana} |\bar{x}_i - \bar{\bar{x}}| \quad (4)$$

O valor usualmente utilizado para  $\alpha$  é 1,482 , mais detalhes no Anexo A.

d) Obter o fator de correção  $\delta$  dos valores das medições da seguinte forma:

$$\delta = \tau * s \quad (5)$$

O valor usualmente utilizado para  $\tau$  é 1,5 , mais detalhes no Anexo A.

e) Atualizar as médias  $\bar{x}_i$  das medições dos participantes. As novas médias  $\bar{x}_i^*$  são obtidas da seguinte forma:

$$\bar{x}_i^* = \begin{cases} \bar{\bar{x}} - \delta, & \text{se } \bar{x}_i < \bar{\bar{x}} - \delta \\ \bar{\bar{x}} + \delta, & \text{se } \bar{x}_i > \bar{\bar{x}} + \delta \\ \bar{x}_i, & \text{caso contrário} \end{cases} \quad (6)$$

Os valores discrepantes são identificados e substituídos por valores estimados através da Eq. (6). Na etapa seguinte, a partir das novas médias  $\bar{x}_i^*$  obtém-se a grande média e o desvio padrão atualizados:

- f) Obter a estimativa grande média  $\bar{\bar{x}}^*$  atualizada das novas médias  $\bar{x}_i^*$  dos participantes:

$$\bar{\bar{x}}^* = \frac{\sum_{i=1}^p \bar{x}_i^*}{p} \quad (7)$$

- g) Obter a estimativa do desvio padrão  $s^*$  atualizado das novas médias dos participantes:

$$s^* = \sqrt{\frac{1}{\beta}} \sqrt{\frac{\sum_{i=1}^p (\bar{x}_i^* - \bar{\bar{x}}^*)^2}{(p-1)}} \quad (8)$$

O valor usualmente utilizado para  $\beta$  é 0,778 , mais detalhes no Anexo A.

O algoritmo A é recursivo, assim, utilizando na Eq. (5) o valor do desvio padrão  $s^*$  obtido na Eq. (8), as etapas (d) até (g) são repetidas até que nenhuma média de participante  $\bar{x}_i^*$  ou  $x_i$  seja alterado.

Nesta etapa, foram obtidos a grande média  $\bar{\bar{x}}^*$  e o desvio padrão  $s^*$  “robustos”, isto é, estimativas com efeito minimizado dos valores extremos para  $\mu$  e  $\sigma$ , respectivamente. De fato, no modelo de EP proposto em (1) considera-se as medições como sendo oriundas de uma distribuição Normal de média  $\mu$  e variância  $\sigma^2$  (desvio padrão  $\sigma$ ). Utilizando  $\bar{\bar{x}}^*$  e  $s^*$  como estimadores de  $\mu$  e  $\sigma$ , respectivamente, pode-se calcular o escore padronizado  $z_i$  onde  $z \sim N(0, 1)$ .

Finalmente, considere novamente as médias iniciais  $\bar{x}_i$  de cada participante conforma Tabela 1. O passo final consiste em calcular o escore  $z_i$  de cada participante da seguinte forma:

$$\text{h) } z_i = \frac{\bar{x}_i - \bar{x}^*}{s^*} \quad (9)$$

Os participantes tem sua avaliação de desempenho realizadas de acordo com os seus valores  $z_i$ , através do seguinte critério:

- $|z_i| \leq 2,0$  desempenho aceitável;
- $2,0 < |z_i| < 3,0$  desempenho questionável, deve rever as condições de ensaio e monitorar os resultados futuros;
- $|z_i| \geq 3,0$  desempenho insatisfatório, deve parar imediatamente realização da medição, bem como rever as condições de ensaio para identificar as causas deste desempenho.

### 2.3. Norma ASTM E691

A norma internacional ASTM E691 “*Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*” é um outro documento de referência importante para provedores de EPs, que fornece orientação sobre o uso de métodos estatísticos nos mesmos. A primeira edição da ASTM E691 foi publicada em 1999, a última atualização é a 12ª versão publicada em 2018.

Assim como na norma ISO 13528 descrita na seção anterior, a metodologia estatística definida para analisar os resultados dos participantes consiste na utilização da média individual do participante ( $\bar{x}_i$ ), a grande média ( $\bar{x}$ ) e a variação da grande média (o desvio padrão  $s$ ), de modo permitir determinar o quanto o resultado de um participante se aproxima da média geral dos participantes.

Numa primeira fase do estudo de EP, a norma ASTM também utiliza um procedimento visando descartar medições atípicas que irão distorcer a análise. Os resultados de média e de desvio padrão são sensíveis à presença de valores discrepantes. Para diminuir estes efeitos, diferentemente da ISO 13528, a norma ASTM E691 define a utilização das estatísticas  $h$  e  $k$  de Mandel (MANDEL, 1974), que é um procedimento baseado em estatísticas descritivas para substituição dos valores discrepantes, desta forma minimizando a influência que estes podem ter sobre estimativas de  $\bar{x}$  e  $s$ .

Este procedimento é realizado somente uma vez nos dados, ou seja, não é iterativo como o procedimento descrito na norma ISO. Desta forma os resultados dos participantes classificados como valor discrepante são substituídos, a identificação de outliers não é realizada somente na média dos participantes em relação a grande média, como na norma ISO, mas também nos valores individuais do participante, de modo a identificar um desvio padrão muito elevado (discrepante) do participante em relação aos outros participantes.

Numa segunda fase do estudo de EP, após a retirada das observações discrepantes, os valores finais das estatísticas descritivas (média do participante, grande média e desvio padrão) são obtidos e utilizadas para construir uma regra de classificação das medições dos participantes.

O Fluxograma de aplicação da metodologia de cálculo da norma ASTM E691 é apresentado na figura 4.



Figura 4 – Fluxograma da metodologia de cálculo apresentada na norma ASTM E691

Considere novamente  $p$  participantes, cada um enviando  $n$  medições, obtém-se os dados estruturados como na Tabela 1. O procedimento estatístico definido na norma ASTM E691 é detalhado a seguir:

a) Obter as médias  $\bar{x}_i$  de cada participante da seguinte forma:

$$\bar{x}_i = \frac{\sum_{r=1}^n x_{ri}}{n} \quad (10)$$

Onde  $x_{ri}$  é a  $r$ -ésima medição fornecida pelo  $i$ -ésimo participante  
 $n$  é o número de repetições realizadas por cada participante.

b) Obter o desvio padrão Individual ( $s_i$ ) dos resultados enviado pelo participante, da seguinte forma:

$$s_i = \sqrt{\frac{\sum_{r=1}^n (x_{ri} - \bar{x}_i)^2}{n - 1}} \quad (11)$$

onde

$\bar{x}_i$  é a média das  $n$  medições do  $i$ -ésimo participante,

$i$  é o número do  $i$ -ésimo participante,

$r$  é o número da  $r$ -ésima repetição da medição do mesmo participante.

c) Obter a estimativa inicial da grande média  $\bar{\bar{x}}$  da seguinte forma:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^p \bar{x}_i}{p} \quad (12)$$

d) Obter a estimativa do desvio ( $d_i$ ) dos resultados enviado pelo participante da seguinte forma:

$$d_i = \bar{x}_i - \bar{\bar{x}} \quad (13)$$

- e) Obter a estimativa do desvio padrão da grande média ( $s_{\bar{x}}$ ) (reprodutibilidade) da seguinte forma:

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^p d_i^2}{p-1}} \quad (14)$$

- f) Realizar a identificação e exclusão de valores discrepantes (variação entre as médias dos participantes) com o auxílio da estatística  $h_i$ , comparando-se a média dos resultados enviado pelo participante ( $\bar{x}_i$ ), a grande média ( $\bar{\bar{x}}$ ) e o seu desvio padrão ( $s_{\bar{x}}$ ), da seguinte forma:

$$h_i = \frac{\bar{x}_i - \bar{\bar{x}}}{s_{\bar{x}}} = \frac{d_i}{s_{\bar{x}}} \quad (15)$$

$$h_{crit} = \frac{(p-1) \cdot t_{(\alpha=0,05; gl=p-2)}}{\sqrt{[p(t^2 + p - 2)]}} \quad (16)$$

Onde

$h_i$  é a estatística  $h$  do  $i$ -ésimo participante,

$h_{crit}$  é a estatística  $h$  crítica para um  $\alpha = 0,05$ ,

$t$  é da distribuição t-student's para um  $\alpha = 0,05$  e  $gl = p - 2$ .

Quando  $|h_i| > h_{crit}$  indica que a média dos resultados enviado pelo participante ( $\bar{x}_i$ ) é diferente significativamente da grande média dos participantes ( $\bar{\bar{x}}$ ), para um  $\alpha = 0,05$  de nível de significância, ou seja, a média dos resultados enviado pelo participante ( $\bar{x}_i$ ) é um valor discrepante. Devemos entender que a reprodutibilidade deste participante é pior quando comparada aos demais participantes.

g) Obter a estimativa do desvio padrão da repetibilidade ( $S_r$ ) da seguinte forma:

$$S_r = \sqrt{\frac{\sum_{i=1}^p S_i^2}{p}} \quad (17)$$

h) Mesmo após eventual exclusão dos participantes discrepantes, consideramos novamente os  $p$  participantes para a análise de repetibilidade. Realizamos agora a identificação e exclusão de valores discrepantes entre as medidas de um mesmo participante (variação entre os resultados individuais dos participantes, repetibilidade) com o auxílio da estatística  $k_i$ . Comparamos o desvio padrão do participante ( $S_i$ ) e o desvio padrão da repetibilidade ( $S_r$ ), da seguinte forma:

$$k_i = \frac{S_i}{S_r} \quad (18)$$

$$k_{crit} = \sqrt{\frac{\frac{p}{p-1}}{1 + \frac{F_{(\alpha=0,05; gl=(n-1); gl=(p-1)(n-1))}}{p-1}}} \quad (19)$$

onde:

$p$  representa o número de participantes,

$k_i$  é a estatística  $k$  do  $i$ -ésimo participante,

$k_{crit}$  é a estatística  $k$  crítica para um  $\alpha = 0,05$ ,

$F$  é da distribuição F de Snedecor, com  $\alpha = 0,05$ ,  $gl = n - 1$  no numerador e  $gl = (p - 1)(n - 1)$  no denominador.

Quando  $|k_i| > k_{crit}$  indica que o desvio padrão do participante ( $S_i$ ) é diferente significativamente do desvio padrão da repetibilidade ( $S_r$ ), para um  $\alpha = 0,05$  de nível de significância, ou seja, medições da mesma unidade feitas pelo mesmo participante apresentam grande discrepância.

Nesta etapa deve-se recalculer a grande média ( $\bar{\bar{x}}$ ) dos participantes (Eq. 12) e o desvio padrão da grande média (Eq. 14), excluindo-se o(s) participante(s) identificado(s) como valor discrepante em algumas das análises (reprodutibilidade e repetibilidade,  $h$  e  $k$  de Mandel respectivamente).

Dessa forma, considerando as médias iniciais  $\bar{x}_i$  de cada um do  $p$  participantes conforme Tabela 1, o passo final consiste em calcular o escore  $z_i$  de cada participante da seguinte forma:

$$i) \quad z_i = \frac{\bar{x}_i - \bar{\bar{x}}}{s_{\bar{x}}} \quad (20)$$

Os participantes tem sua avaliação de desempenho realizadas de acordo com os seus valores  $z_i$ , através do seguinte critério:

- $|z_i| \leq 2,0$  desempenho aceitável;
- $2,0 < |z_i| < 3,0$  desempenho questionável, deve rever as condições de ensaio e monitorar os resultados futuros;
- $|z_i| \geq 3,0$  desempenho insatisfatório, deve parar imediatamente realização da medição, bem como rever as condições de ensaio para identificar as causas deste desempenho.

## 2.4. Técnicas estatísticas para análise comparativa das normas ISO e ASTM

### 2.4.1 Teste de Wilcoxon para amostras pareadas

O Teste de Wilcoxon é um teste de hipótese não paramétrico para comparação de duas amostras pareadas que se baseia no posto ( $r$ ) das diferenças entre pares. (SIEGEL, S. CASTELAN, NJ, 1988).

Tabela 2 - Modelo de Tabela utilizada no Teste de Wilcoxon

	Amostras	
	$X$	$Y$
1	$x_1$	$y_1$
2	$x_2$	$y_2$
...	...	...
$i$	$x_i$	$y_i$
...	...	...
$n$	$x_n$	$y_n$

Os dados das duas amostras pareadas são dispostos conforme Tabela 2, onde a observação  $(x_i, y_i)$  representa o par valores do  $i$ -ésimo indivíduo. Considerando tamanhos de amostra superiores a 25 ( $n > 25$ ), o conjunto de passos para o teste de hipótese é descrito a seguir:

a) Hipóteses do teste:

$$\begin{cases} H_0: \text{Não existe diferença dentro dos pares} \\ H_1: \text{Existe diferença dentro dos pares} \end{cases}$$

b) Obter o desvio ( $D_i$ ) entre os pares da amostra, da seguinte forma:

$$D_i = x_i - y_i \tag{24}$$

c) Obter a soma dos postos positivos ( $T^+$ ) e a soma dos postos negativos ( $T^-$ ), da seguinte forma:

$$T^+ = \sum_{i=1}^N I_i * r(|D_i|) \quad (25)$$

$$T^- = \sum_{i=1}^N (1 - I_i) * r(|D_i|) \quad (26)$$

onde:

$I_i$  é o indicador do sinal do desvio ( $D_i$ ), ou seja  $I_i = \begin{cases} 1 & \text{se } D_i > 0 \\ 0 & \text{se } D_i \leq 0 \end{cases}$ ;

$r(|D_i|)$  é o posto do módulo dos desvios ( $D_i$ ), começando com 1 para o menor  $D_i$ . Valores repetidos recebem um posto igual a média dos postos que eles abrangem.

d) A estatística do Teste de Wilcoxon ( $T$ ) é aproximada pela normal padronizada, da seguinte forma:

$$T = \begin{cases} T^+ & \text{se } T^+ \leq T^- \\ T^- & \text{se } T^- < T^+ \end{cases} \quad (27)$$

$$Z = \frac{T - \mu_T}{\sigma_T} \approx N(0,1) \quad (28)$$

onde:

$\mu_T$  é a esperança de  $T$ , dado por  $\mu_{T^+} = E(T) = \frac{N(N+1)}{4}$

$\sigma_T$  é desvio padrão de  $T$ , dado por  $\sigma_{T^+} = \sqrt{Var(T)} = \sqrt{\frac{N(N+1)(2N+1)}{24}}$

- e) A partir de um nível de significância  $\alpha$ , definimos a região de aceitação da hipótese  $H_0$ . Valor de Z superior ao limite crítico obtido a partir de  $\alpha$  sugere que existe diferença significativa dentro dos pares. O nível de significância a ser adotado neste trabalho é de  $\alpha=0,05$ . O teste de Wilcoxon será realizado no software IBM SPSS Statistics. Versão 20.

#### 2.4.2 Coeficiente de concordância de Kappa

O coeficiente de concordância de Kappa é utilizado para descrever a concordância entre dois ou mais métodos de avaliação quando realizam uma avaliação nominal ou ordinal de uma mesma amostra. Os dados são dispostos conforme Tabela 3.

Tabela 3 - Modelo de Tabela de contingência utilizada para o coeficiente de Kappa.

		Método de avaliação Y			Total
		1	...	r	
método de avaliação X	1	$n_{11}$	...	$n_{1i}$	$n_{1.}$
	2	$n_{21}$	...	$n_{2i}$	$n_{2.}$
	...	...	...	...	...
	r	$n_{r1}$	...	$x_{rr}$	$n_{r.}$
Total		$n_{.1}$	...	$n_{.r}$	$n$

O coeficiente de concordância de Kappa é obtido da seguinte forma (LANDIS JR, KOCH GG., 1977):

$$p_0 = \sum_{i=1}^r \frac{n_{ii}}{n}, \quad (29)$$

$$p_e = \sum_{i=1}^r \frac{n_{i.} * n_{.i}}{n^2}, \quad (30)$$

$$K = \frac{p_0 - p_e}{1 - p_e}, \quad (31)$$

onde  $p_0$  indica a proporção de avaliações concordantes entre os métodos X e Y e  $p_e$  indica a proporção de concordâncias “randômicas” entre X e Y. O valor  $K$  do coeficiente de concordância de Kappa pode variar de  $(p_e/1 - p_e)$  até 1. Quanto mais próximo de 1 for seu valor, maior é o indicativo de que existe uma concordância entre os métodos de avaliação, isto é, a proporção de concordantes é substancialmente superior à proporção randômica. De maneira oposta, quanto mais próximo de zero, maior é o indicativo de que a concordância é substancialmente aleatória. Uma descrição detalhada do coeficiente de Kappa pode ser encontrado em Landis JR e Koch GG. A Tabela 4 apresentam uma interpretação dos os valores  $K$  (Landis JR. E Koch GG, 1977).

Tabela 4 - Faixa de classificação de Kappa

<b>Valor de Kappa</b>	<b>Interpretação</b>
$K < 0$	insignificante
$0 \leq K \leq 0,2$	fraca
$0,2 < K \leq 0,4$	razoável
$0,4 < K \leq 0,6$	moderada
$0,6 < K \leq 0,8$	forte
$0,8 < K < 1,0$	quase perfeita
$K = 1,0$	perfeita

Fonte: Landis JR. E Koch GG, 1977

O cálculo do coeficiente de Kappa será realizada no software Microsoft Excel. Versão 2016.

### 3. METODOLOGIA

Esta seção descreve a metodologia a ser utilizada para a realização do estudo comparativo das normas ISO 13528 e a ASTM E691, que será realizado a partir de dados de uma série de Estudos de Proficiência disponibilizados pelo Instituto SENAI de Inovação em Engenharia de Polímeros em 2018. As subseções apresentam informações detalhadas sobre estes estudos, incluindo as variáveis de análise e o número de participantes. Adicionalmente, será detalhado o conjunto de passos da análise que é proposta neste trabalho, bem como as estatísticas que serão utilizados para comparar os resultados das duas normas.

#### 3.1. Banco de dados

Os dados utilizados foram obtidos dos Ensaios de Proficiência (EP) realizados no ano de 2018 pelo Instituto SENAI de Inovação em Engenharia de Polímeros, na 33ª Rodada do EP em ensaios em Elastômeros. Nestes EPs tiveram participantes do Brasil e também de outros países, incluindo USA, México, China, Espanha, Portugal, Colômbia, Peru, Argentina e Uruguai. No total foram utilizados resultados de 27 EPs, cujo o número de participantes foi de pelo menos 15 ( $p \geq 15$ ).

Os resultados dos participantes dos EPs oriundos de ensaios físico-mecânicos e químicos realizados em corpos de prova de borracha. As variáveis (ou características) utilizadas nestes estudos incluem medições de dureza, tração, rasgamento e envelhecimento. A Tabela 5 descreve resumidamente os EPs incluídos neste trabalho agrupados por tipo de variável utilizada.

Tabela 5 - EPs agrupados por variáveis

Variável	Número de EPs	Número de participantes por EP
Dureza Shore A	3	53
Tração – Tensão na Ruptura (TR)	3	40
Tração – Alongamento na Ruptura (AL)	3	40
Tração – Módulo a 100%	3	38
Tração – Módulo a 300%	3	36
Rasgamento tipo C	3	23
Envelhecimento Acelerado – Variação da Dureza Shore A	3	16
Envelhecimento Acelerado – Variação da TR	3	16
Envelhecimento Acelerado – Variação do AL	3	16

Importante ressaltar que para cada variável existe um protocolo para preparação dos corpos de prova, bem como realização da medida. Entretanto, neste trabalho não apresentamos a descrição desta etapa. Ressaltamos que este protocolo é seguido na íntegra nos EPs e não varia entre as normas ISO 13528 e ASTM E691.

### 3.2. Análise Comparativa

Nesta seção apresentamos os procedimentos estatísticos utilizados no estudo comparativo dos resultados das EPs entre as normas ISO 13528 e a ASTM E691. Em cada EP descrito na Tabela 3 iremos realizar dois estudos comparativos, descritos a seguir:

- 1) Comparar os resultados da estimação da média e do desvio padrão robustos de ambos os métodos através do teste não paramétrico de Wilcoxon (descrito na seção 2.4.1).
- 2) Comparar o grau de concordância na avaliação de desempenho dos participantes por ambos os métodos, através do coeficiente de concordância de Kappa. (descrito na seção 2.4.2).

## 4. ESTUDO DE CASOS

Nesta seção apresentamos os resultados da comparação dos procedimentos estatísticos utilizados nas normas ISO 13528 e a ASTM E691 através de 27 EPs, descritos na seção anterior. A análise de EP foi realizada conforme a norma ISO 13528 [Equações (2) até (9)]; e via a norma ASTM E691 [Equações (2) e (14) até (23)]. A subseção 5.1 apresenta o resultado da comparação das estimativas de média e desvio padrão robustos de ambos os métodos. O teste de Wilcoxon foi realizado para realizar tal comparação [Equações (24) até (28)]. A subseção 5.2 apresenta a análise de concordância das duas normas na avaliação dos participantes em cada uma das 27 EPs analisadas. Tal comparação foi realizada através do coeficiente de Kappa [Equações (29) até (31)].

### 4.1 Comparação entre estimativas das médias e dos desvios padrão

A Tabela 6 (colunas 4 até 7) apresenta em cada estudo de EP as estimativas da média e do desvio padrão robustos de ambos os métodos. O Teste Wilcoxon não indicou diferença significativa ao nível de 5% (valor  $p = 0,962 > 0,05$ ) entre as estimativas das médias das duas normas (colunas 4 e 6 da Tabela 6, respectivamente). De maneira análoga, o Teste Wilcoxon indicou não indicou diferença significativa ao nível de 5% (valor  $p = 0,249 > 0,05$ ) entre as estimativas do desvio padrão das duas normas (colunas 5 e 7 da Tabela 6, respectivamente). Estes resultados evidenciam a consistência nas estatísticas utilizadas em cada método.

Tabela 6 - Resultados dos EPs e das análises estatísticas

Ensaio	EP	Nº de participantes ( <i>p</i> )	ISO 13528		ASTM E691		Coeficiente de concordância de Kappa			
			Média ( $\bar{x}$ )	Desvio Padrão ( $s^*$ )	Média ( $\bar{x}$ )	Desvio Padrão ( $s_{\bar{x}}$ )	<i>K</i>	classificação	Discordâncias	
									ISO mais rigorosa	ASTM mais rigorosa
Dureza Shore A	1	53	41,02	1,11	41,13	1,32	0,56	moderada	7	0
	2	53	62,99	1,59	63,02	1,50	0,89	quase perfeita	0	1
	3	53	84,04	1,45	84,06	1,36	0,89	quase perfeita	0	1
Tração – Tensão na Ruptura (TR)	4	40	7,09	0,66	7,07	0,77	0,79	forte	1	0
	5	40	16,21	1,13	16,17	1,15	1,00	perfeita	0	0
	6	40	22,00	1,20	21,90	1,33	0,85	quase perfeita	1	0
Tração – Alongamento na Ruptura (AL)	7	40	348,96	32,40	349,69	33,88	0,88	quase perfeita	1	0
	8	40	640,66	36,65	638,26	43,73	0,75	forte	3	0
	9	40	435,26	26,45	436,70	32,07	0,81	quase perfeita	2	0
Tração – Módulo a 100%	10	38	1,33	0,14	1,32	0,18	0,76	forte	2	0
	11	38	1,49	0,11	1,50	0,19	0,65	forte	4	0
	12	38	2,86	0,26	2,86	0,39	0,80	forte	2	0
Tração – Módulo a 300%	13	36	5,54	0,57	5,56	0,56	1,00	perfeita	0	0
	14	36	5,92	0,52	5,91	0,53	0,88	quase perfeita	1	0
	15	36	14,22	1,08	14,19	1,15	0,73	forte	2	0
Rasgamento tipo C	16	22	19,22	4,42	19,33	4,02	0,65	forte	0	1
	17	23	42,18	3,55	42,16	3,67	1,00	perfeita	0	0
	18	23	51,33	6,34	52,03	7,04	1,00	perfeita	0	0
Envelhecimento Acelerado – Variação da Dureza Shore A	19	16	1,78	1,09	1,86	1,14	0,64	forte	1	0
	20	16	1,97	1,05	1,95	0,96	0,64	forte	0	1
	21	16	2,00	0,95	1,99	0,90	1,00	perfeita	0	0
Envelhecimento Acelerado – Variação da TR	22	16	-6,44	8,14	-6,60	8,27	0,64	forte	1	0
	23	16	-2,53	4,76	-1,64	3,76	0,64	forte	0	1
	24	16	-3,83	2,53	-4,39	3,07	1,00	perfeita	0	0
Envelhecimento Acelerado – Variação do AL	25	16	-9,43	5,66	-11,02	5,38	1,00	perfeita	0	0
	26	16	-8,32	4,48	-8,77	4,24	1,00	perfeita	0	0
	27	16	-8,27	7,18	-8,14	6,08	1,00	perfeita	0	0

## 4.2 Comparação da avaliação de desempenho dos participantes

Para a comparação do grau de concordância das normas ISO 13528 e a ASTM E691 na avaliação de desempenho dos participantes, utilizamos o coeficiente de concordância de Kappa, conforme Tabela 7.

Tabela 7 - Comparação da avaliação de desempenho dos participantes

		Avaliação de desempenho pela ISO 13528			Total
		aceitável $ z_i  \leq 2,0$	questionável $2,0 <  z_i  < 3,0$	insatisfatório $ z_i  \geq 3,0$	
Avaliação de desempenho pela ASTM E691	aceitável $ z_i  \leq 2,0$			ISO mais rigorosa	
	questionável $2,0 <  z_i  < 3,0$				
	insatisfatório $ z_i  \geq 3,0$	ASTM mais rigorosa			
Total					

Concordância entre as normas

A Tabela 6 (coluna 8) apresenta os valores do coeficiente de Kappa de cada EP. Segundo a escala de classificação dos valores de Kappa (Tabela 5), através da Figura 4 podemos observar que 96% (26 das 27 dos EPs) apresentam nível de concordância superior a “forte”. Apenas em uma EP investigada o grau de concordância foi “moderado”.

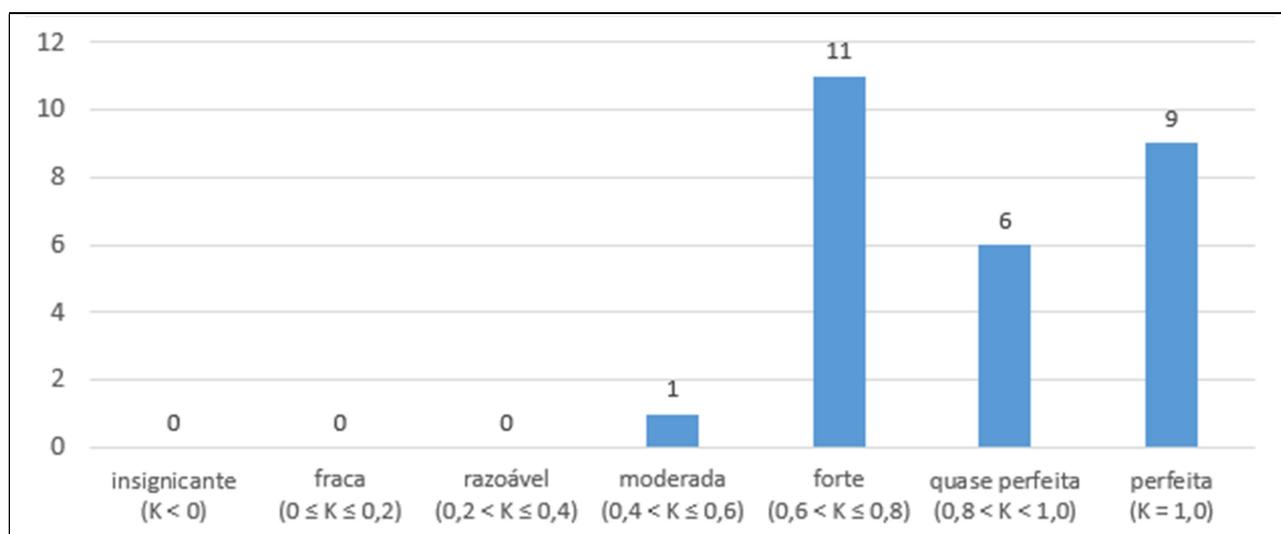


Figura 4 - Distribuição dos valores dos coeficientes de Kappa nos EPs

Observando as colunas 10 e 11 da Tabela 6, podemos fazer uma análise mais apurada dos casos discordantes. Estes resultados discordantes estão totalizados (considerando todas as EPs conjuntamente) e detalhados na Tabela 8. A diagonal principal apresenta o conjunto total de participantes analisados nos estudos que receberam a mesma avaliação de ambas as normas (em verde na Tabela 8). O valor do coeficiente de Kappa resultante das EPs acumuladas é de  $K=0,79$ , o que indica concordância “forte”. Em relação aos resultados discordantes, podemos notar que a norma ISO 13528 foi em 85% dos casos mais rigorosa do que a norma ASTM E691, pois dentre as 33 discordâncias na avaliação de desempenho dos participantes (em amarelo na Tabela 8), em 28 a norma ISO 13528 foi mais rigorosa. Observando mais detalhadamente as discordâncias com maior rigor da norma ISO 13528, podemos notar que, dentre as 28 discordâncias, 19 participantes receberam a classificação “questionável” por esta norma e “aceitável” pela norma ASTM E691. Este resultado nos revela que, embora na análise geral das EPs o percentual de discordâncias tenha sido baixo, evidenciamos uma tendência de que a norma ISO questione participantes que são classificados apenas como “aceitáveis” na norma ASTM. Em termos práticos, estes 19 participantes são classificados como aceitáveis nas pela ASTM e “em alerta” pela ISO.

Em relação às discordâncias com maior rigor da norma ASTM E691, observamos na Tabela 8 um número de apenas 5 discordâncias. Dentre elas, tivemos frequências semelhantes de classificações discrepantes do tipo (ISO “aceitável” e ASTM “questionável”) com 2 casos; e do tipo (ISO “questionável” e ASTM “insatisfatória”) com 3 casos.

Tabela 8 - Comparação da avaliação de desempenho dos participantes acumulada

		Avaliação de desempenho pela ISO 13528			Total
		aceitável $ z_i  \leq 2,0$	questionável $2,0 <  z_i  < 3,0$	insatisfatório $ z_i  \geq 3,0$	
Avaliação de desempenho pela ASTM E691	aceitável $ z_i  \leq 2,0$	724	19	0	743
	questionável $2,0 <  z_i  < 3,0$	2	31	9	42
	insatisfatório $ z_i  \geq 3,0$	0	3	42	45
Total		726	53	51	830

## 6. CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma descrição detalhada das metodologias estatísticas contidas nas normas para Ensaios de Proficiência (EPs) ISO 13528 e ASTM E691. Adicionalmente, a partir de um estudo de casos contendo 27 EPs foi realizado um estudo comparativo dessas normas, incluindo a análise comparativa das estimativas da média e desvio padrão “robustos” de ambas, bem como uma análise do grau de concordância das classificações dos participantes entre as normas.

Em linhas gerais, conforme o esperado, os resultados evidenciam o bom grau de classificações concordantes entre as normas. Em relação aos poucos resultados discrepantes, o estudo evidenciou que a norma ISO é mais rígida do que a ASTM na grande maioria dos casos discordantes, sendo que a discordância mais frequente nos participantes classificados como “aceitáveis” pela norma ASTM é “questionáveis” pela norma ISO.

Em trabalhos futuros seria interessante o estudo mais detalhado da norma ISO 13528, avaliando o impacto de diferentes valores de  $\alpha$ ,  $\tau$  e  $\beta$  do Algoritmo A nas estimativas da média, do desvio padrão e na identificação de valores dispersos dos dados das EPs. Adicionalmente, através de dados simulados com diversas estruturas (assimétricos a direita, à esquerda, com curtose elevada, com outliers propositais) poderíamos realizar um estudo comparativo mais completo das normas ISO e ASTM.

## REFERÊNCIAS

- [1] AMERICAN SOCIETY FOR TESTING AND MATERIALS. **ASTM E691**: Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. West Conshohocken, ASTM, 2018.
- [2] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 13528**: Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva, ISO, 2015.
- [3] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC Guide 43-1**: Proficiency testing by interlaboratory comparisons - Part 1: Development and operation of proficiency testing schemes. Geneva, ISO, 1997.
- [4] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC Guide 43-2**: Proficiency testing by interlaboratory comparisons - Part 2: Election and use of proficiency testing schemes by laboratory accreditation bodies. Geneva, ISO, 1997.
- [5] ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 17043**: Avaliação de Conformidade - Requisitos gerais para ensaios de proficiência. Brasil, ABNT, 2011.
- [6] THOLEN, DW. Impact of international standards and initiatives on proficiency testing for medical laboratories. **Accreditation and Quality Assurance**. v.9, n.11-12, p. 653-656, 2004.
- [7] WONG, Siu Kay. Review of the new edition of ISO 13528. **Accreditation and Quality Assurance**. Berlin, v.21, p. 249-254, 2016.
- [8] Robust Statistics - How Not to Reject Outliers. **Analytical Methods Committee**. London, v.114, p. 1693-1697, 1989.
- [9] MANDEL, J and ASHOF, T.W. Interpretation and Generalization of Youden's Two-Sample Diagram, **Journal of Quality Technology**. v6, p. 22-36, 1974.
- [10] LANDIS JR, KOCH GG. The Measurement of Observer Agreement for Categorical Data. **International Biometric Society**. v33, n1, p. 159-74, 1977.
- [11] SIEGEL, S. CASTELAN, NJ. **Nonparametric Statistics for the Behavioral Sciences**. 2nd ed. New York: McGraw-Hill, 1988. 399p.
- [12] ISI, Engenharia de Polímeros. Programa de Ensaios de Proficiência (PEP) por Comparação Interlaboratorial - **Relatório Final da 33ª Rodada** - Ensaios em Elastômeros. São Leopoldo. 2018. p59.
- [13] Software IBM SPSS Statistics. Versão 20.  
(tutorial disponível em <https://www.ibm.com/support/knowledgecenter/search/spss> acesso em 18/06/2019).
- [14] Software Microsoft Excel. Versão 2016.  
(tutorial disponível em <https://support.office.com/pt-br/excel> acesso em 18/06/2019).

## ANEXO A

Este anexo apresenta uma descrição breve da base teórica das constantes  $\alpha$ ,  $\tau$  e  $\beta$ . A descrição detalhada da teoria que sustenta a obtenção desses valores é encontrada em *Analytical Methods Committee (1989)*.

Para definição do  $\alpha$  utilizado na Eq. (4), no cálculo do valor inicial do desvio padrão  $s$ , utiliza-se a mediana do desvio absoluto [MAD, presente na Eq. (4)]. O MAD é uma estatística robusta, sendo mais resiliente a valores dispersos (outliers) em um conjunto de dados do que o desvio padrão. No desvio padrão, as distâncias da média são ao quadrado, de modo que os grandes desvios são mais pesados e, portanto, os desvios podem influenciá-lo fortemente.

Para dados com distribuição normal de média  $\mu$  e variância  $\sigma^2$  (desvio padrão  $\sigma$ ),  $\alpha$  é o inverso da função de distribuição cumulativa ( $\Phi^{-1}$ ), sendo obtido da seguinte forma:

$$\alpha = \frac{1}{(\Phi_{(1/2;3/4)}^{-1})} \approx \frac{1}{0,67449 \dots} = 1,4824 \dots$$

Para uma distribuição normal padronizada  $z = (\bar{x} - \mu)/\sigma$ , o argumento  $3/4$  é tal que  $\pm$  MAD cobre 50% (entre  $1/4$  e  $3/4$ ) da função de distribuição cumulativa da distribuição normal padrão (Fig. 5), temos:

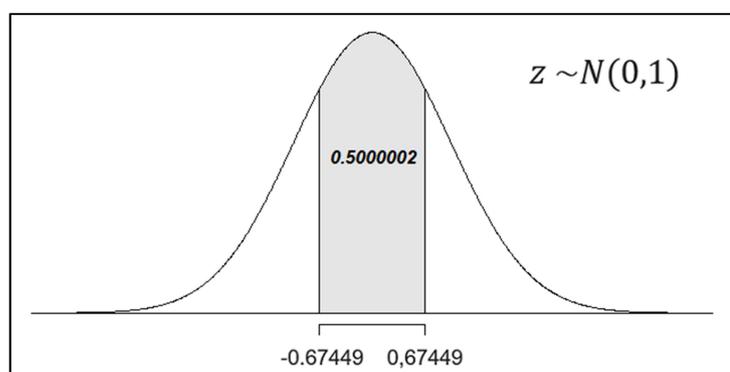


Figura 5 - Região da mediana do desvio absoluto (MAD)

$$\frac{1}{2} = P(|\bar{x} - \mu| \leq MAD) = P\left(\left|\frac{\bar{x} - \mu}{\sigma}\right| \leq \frac{MAD}{\sigma}\right) = P\left(|z| \leq \frac{MAD}{\sigma}\right)$$

O valor de  $\tau$  utilizado na Eq. (5), representa um limiar para um intervalo simétrico cuja a probabilidade é obtida na variável normal padronizada ( $Z$ ). Esta probabilidade, chamada de  $\theta$ , é utilizada para obtenção de  $\beta$ .

$$\theta = P(|Z| < \tau)$$

$$\beta = \frac{\theta + \tau^2(1 - \theta) - 2\tau \exp\left(-\tau^2/2\right)}{\sqrt{2\pi}}$$

Os valores de  $\tau$  são apresentados na Tabela 9, onde  $\tau$  deve ser escolhido dependendo da frequência da presença de outliers,  $\tau = 1,4$  para 5% de outliers e  $\tau = 2,0$  para 1% de outliers, um valor menor de  $\tau$  deve ser escolhido em caso de dúvida. O valor de  $\tau = 1,5$  é amplamente utilizado, fornecendo bons resultados em termos de estimativas de  $\mu$  e  $\sigma$ .

Tabela 9 - Valores de  $\beta$  e  $\tau$

	$\tau$										
	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
$\beta$	0,516	0,578	0,635	0,688	0,736	0,778	0,816	0,849	0,877	0,900	0,921

Fonte: *Analytical Methods Committee, 1989*

A definição do  $\beta$  utilizado na Eq. (8), no cálculo do desvio padrão  $s^*$ , depende do valor escolhido do valor de corte  $\tau$ . Para atualizar a forma de obtenção do desvio padrão  $s^*$ , com o valor de corte  $\tau = 1,5$  deve ser utilizado o  $\beta = 0,778$ , conforme Tabela 9.