

UNIVERSIDADE FEDERAL DO RIO GRANDE
INSTITUTO DE MATEMÁTICA, FÍSICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL
CURSO DE MESTRADO EM MODELAGEM COMPUTACIONAL

Dissertação de Mestrado

**ANÁLISE DE ALGUMAS PROPRIEDADES DA
TENDÊNCIA CENTRAL DE UM MÉTODO DE
GERAÇÃO DE NÚMEROS *FUZZY*: O CASO DO
MÉTODO CHENG**

Tamara Ost Fracari

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande, como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional

Orientadora: Prof^a. Dr^a. Viviane Leite Dias de Mattos

Rio Grande, 2017

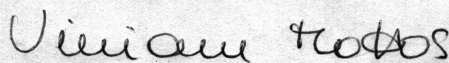
Tamara Ost Fracari

“Análise de algumas propriedades da tendência central de um método de geração de números fuzzy: o caso do método Cheng”

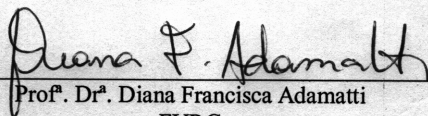
Dissertação apresentada ao Programa de Pós Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Mestre. Área concentração: Modelagem Computacional.

Aprovada em

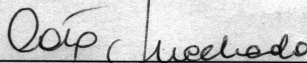
BANCA EXAMINADORA



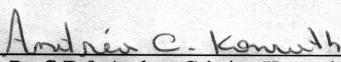
Prof.ª. Dr.ª. Viviane Leite Dias de Mattos
Orientadora – FURG



Prof.ª. Dr.ª. Diana Francisca Adamatti
FURG



Prof.ª. Dr.ª. Catia Maria dos Santos Machado
FURG



Prof.ª. Dr.ª. Andrea Cristina Konrath
UFSC

Rio Grande - RS
2017

Dados de catalogação na fonte:

Bibliotecário Me. João Paulo Borges da Silveira – CRB 10/2130

Biblioteca Central – FURG

F797a Fracari, Tamara Ost

ANÁLISE DE ALGUMAS PROPRIEDADES DA TENDÊNCIA CENTRAL DE UM MÉTODO DE GERAÇÃO DE NÚMEROS *FUZZY*: O CASO DO MÉTODO CHENG / Tamara Ost Fracari. – Rio Grande, 2017. – 98 f: gráf. – Dissertação (Mestrado) – Programa de Pós-Graduação em Modelagem Computacional. Universidade Federal do Rio Grande. Instituto de Matemática, Física e Estatística. Rio Grande, 2017. – Orientadora Viviane Leite Dias de Mattos.

1. Método Cheng. 2. Lógica *fuzzy*. 3. Comportamento de distribuição amostral. I. Mattos, Viviane Leite Dias de. II. Título.

CDU: 004

Dedico este trabalho a minha mãe Reni Maria Ost Fracari, ao meu pai Dilso Reus Fracari e ao meu irmão Tiago Ost Fracari.

AGRADECIMENTOS

Primeiramente, à Deus pela vida e por me conduzir durante todo esse tempo de mestrado.

À minha mãe, Reni Maria Ost Fracari, ao meu pai, Dilso Reus Fracari, pelas incontáveis conversas de ânimo, sempre me fazendo seguir e frente e me ensinando a nunca desistir.

Ao meu irmão, Tiago Ost Fracari, por sempre estar disposto a me ouvir, por compartilhar minhas alegrias e tristezas e por sempre me visitar, quando ficava difícil viajar.

Ao meu namorado, Maicon Machado de Abreu, por me dar forças e ânimo para concluir esse trabalho.

À minha família e aos meus amigos por sempre me apoiarem e por entenderem as minhas ausências.

Ao Matheus Korb, pela implementação do método Cheng.

Aos colegas, em especial, Fernando Pereira de Toledo, por todas as ajudas de programação e \LaTeX .

À minha orientadora, Viviane Leite Dias de Mattos, por sempre estar disposta a ouvir minhas dúvidas durante a pesquisa e sempre me ajudar a achar soluções para os problemas que apareciam no decorrer da pesquisa.

Aos demais professores, por sempre estarem dispostos a ajudar e ensinar o que seria necessário para o bom andamento da pesquisa.

Ao Programa de Pós-Graduação em Modelagem Computacional (PPGMC) e a Universidade Federal do Rio Grande (FURG) pela oportunidade de estudo.

À Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES) pelo auxílio financeiro em forma de Bolsa de Mestrado - Demanda Social.

RESUMO

FRACARI, Tamara Ost. **ANÁLISE DE ALGUMAS PROPRIEDADES DA TENDÊNCIA CENTRAL DE UM MÉTODO DE GERAÇÃO DE NÚMEROS FUZZY: O CASO DO MÉTODO CHENG.** 2017. 98 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Modelagem Computacional. Universidade Federal do Rio Grande, Rio Grande.

Este estudo analisa o comportamento da distribuição amostral de uma medida, denominada de *mCheng*, usada para determinar o valor central de um número *fuzzy* em um método de transformação de valores *crisp* em valores *fuzzy*, o método Cheng. O valor central do número *fuzzy* triangular é associado ao grau de pertinência um, razão pela qual se espera que se situe onde os dados apresentam maior concentração. Depois da verificação de algumas propriedades matemáticas desta medida, a sua distribuição amostral foi comparada com as distribuições amostrais de duas medidas de tendência central bastante utilizadas em análises estatísticas de dados: a média aritmética e a mediana. Para o desenvolvimento do estudo foram geradas amostras aleatórias retiradas de populações que apresentavam distribuição normal e exponencial, com diferentes variabilidades, sendo considerados três tamanhos amostrais: 5, 30 e 100. Antes da comparação das três distribuições amostrais, o que foi feito separadamente para cada um dos doze experimentos, foi realizada uma análise exploratória dos dados, utilizando métodos gráficos e analíticos, além de testes de hipóteses (Kolmogorov-Smirnov Lilliefors e Shapiro-Wilk). Outros testes de hipóteses (Levene, Welch, Friedman, Wilcoxon) foram usados para comparar a centralidade e dispersão das distribuições amostrais. A análise desenvolvida não encontrou evidências de diferença na centralidade das distribuições amostrais das três medidas (média, mediana e *mCheng*) nos experimentos em que as amostras foram geradas a partir de uma população com distribuição normal, o mesmo não acontecendo em relação à variabilidade. A menor variabilidade foi encontrada na distribuição amostral da média, considerada o melhor estimador. Nos experimentos em que as amostras foram geradas a partir de populações que apresentavam distribuição exponencial, foram encontradas evidências de diferença entre a centralidade e entre a dispersão das três distribuições amostrais, sendo a medida *mCheng* considerada melhor estimador para pequenas amostras.

Palavras-chave: Método Cheng, Lógica *fuzzy*, Comportamento de distribuição amostral.

ABSTRACT

FRACARI, Tamara Ost. **Analysis of some properties of a generating fuzzy numbers method: the case of the Cheng method.** 2017. 98 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Modelagem Computacional. Universidade Federal do Rio Grande, Rio Grande.

This job analyzes the behavior of the measure sample distribution, denominated *mCheng*, used to determine the central value of a fuzzy number in a method of transforming crisp values into fuzzy values, the Cheng method. The central value of the triangular fuzzy number is associated with the degree of pertinence one, which is why it is expected to be located where the data presents a higher concentration. After verifying some mathematical properties of this measure, its sample distribution was compared with the sample distribution of two measures of the central tendency much used in statistical analysis of data: the arithmetic mean and the median. For the development of the job, random samples were taken generated from populations with normal and exponential distribution, with different variabilities, being considered three sample sizes: 5, 30 and 100. Before comparing the three sample distributions, which was done separately for each of the twelve experiments, an exploratory analysis of the data, it was performed using graphical and analytical methods, yet hypothesis tests (Kolmogorov-Smirnov Lilliefors and Shapiro-Wilk). Other hypothesis tests (Levene, Welch, Friedman, Wilcoxon) were used to compare a centrality and dispersion of sample distributions. A developed analysis not found evidences of difference in the centrality of the sample distributions of the three measures (mean, median and *mCheng*) in the experiments in which the samples were generated from a population with normal distribution, the same was not happening in relation to variability. The lowest variability was found in the sample distribution of the mean, considered the best estimator. In the experiments where the samples were generated from populations that presented exponential distribution, it was found evidence of difference between the centrality and the dispersion of the three sample distributions, being a *mCheng* measure considered best estimator for small samples.

Keywords: Cheng Method, Fuzzy Logic, Sample Distribution Behavior.

LISTA DE FIGURAS

Figura 1	Distribuição normal: representação gráfica	26
Figura 2	Áreas importantes sobre a curva normal	26
Figura 3	Distribuição normal padrão: representação gráfica	27
Figura 4	Distribuição exponencial: representação gráfica	28
Figura 5	Distribuição exponencial acumulada: representação gráfica	28
Figura 6	Número <i>fuzzy</i> trapezoidal: representação gráfica	38
Figura 7	Número <i>fuzzy</i> triangular: representação gráfica	39
Figura 8	Fluxograma representando a análise realizada	45
Figura 9	Média: experimentos nº 1 à nº 6	50
Figura 10	Desvio padrão: experimentos nº 1 à nº 6	51
Figura 11	Média: experimentos nº 7 à nº 12	51
Figura 12	Desvio padrão: experimentos nº 7 à nº 12	52
Figura 13	Histogramas: experimento nº 1	54
Figura 14	Gráficos de probabilidade normal: experimento nº 1	54
Figura 15	Boxplots: experimento nº 1	55
Figura 16	Histogramas: experimento nº 2	55
Figura 17	Gráficos de probabilidade normal: experimento nº 2	55
Figura 18	Boxplots: experimento nº 2	56
Figura 19	Histogramas: experimento nº 3	58
Figura 20	Gráficos de probabilidade normal: experimento nº 3	58
Figura 21	Boxplots: experimento nº 3	59
Figura 22	Histogramas: experimento nº 4	59
Figura 23	Gráficos de probabilidade normal: experimento nº 4	59
Figura 24	Boxplots: experimento nº 4	60
Figura 25	Histogramas: experimento nº 5	62
Figura 26	Gráficos de probabilidade normal: experimento nº 5	62
Figura 27	Boxplots: experimento nº 5	62
Figura 28	Histogramas: experimento nº 6	63
Figura 29	Gráficos de probabilidade normal: experimento nº 6	63
Figura 30	Boxplots: experimento nº 6	63
Figura 31	Histogramas: experimento nº 7	65
Figura 32	Gráficos de probabilidade normal: experimento nº 7	66
Figura 33	Boxplots: experimento nº 7	66
Figura 34	Histogramas: experimento nº 8	66
Figura 35	Gráficos de probabilidade normal: experimento nº 8	67
Figura 36	Boxplots: experimento nº 8	67

Figura 37	Histogramas: experimento nº 9	69
Figura 38	Gráficos de probabilidade normal: experimento nº 9	69
Figura 39	Boxplots: experimento nº 9	70
Figura 40	Histogramas: experimento nº 10	70
Figura 41	Gráficos de probabilidade normal: experimento nº 10	70
Figura 42	Boxplots: experimento nº 10	71
Figura 43	Histogramas: experimento nº 11	72
Figura 44	Gráficos de probabilidade normal: experimento nº 11	73
Figura 45	Boxplots: experimento nº 11	73
Figura 46	Histogramas: experimento nº 12	73
Figura 47	Gráficos de probabilidade normal: experimento nº 12	74
Figura 48	Boxplots: experimento nº 12	74

LISTA DE TABELAS

Tabela 1	Experimento 1: simulação	42
Tabela 2	Propriedades verificadas e não verificadas para a medida <i>mCheng</i> . .	50
Tabela 3	Análise estatística: experimento nº 1 e experimento nº 2	53
Tabela 4	Análise estatística: experimento nº 3 e experimento nº 4	57
Tabela 5	Análise estatística: experimento nº 5 e experimento nº 6	61
Tabela 6	Análise estatística: experimento nº 7 e experimento nº 8	65
Tabela 7	Comparações múltiplas pelo teste de postos de Wilcoxon: experimen- tos nº 7 e nº 8	67
Tabela 8	Análise estatística: experimento nº 9 e experimento nº 10	68
Tabela 9	Comparações múltiplas pelo teste de postos de Wilcoxon: experimen- tos nº 9 e nº 10	71
Tabela 10	Análise estatística: experimento nº 11 e experimento nº 12	72
Tabela 11	Comparações múltiplas pelo teste de postos de Wilcoxon: experimen- tos nº 11 e nº 12	75

LISTA DE ABREVIATURAS E SIGLAS

iid	Independentes e identicamente distribuídas
fp	Função de probabilidade
VAD	Variável aleatória discreta
fdp	Função densidade de probabilidade
VAC	Variável aleatória contínua

LISTA DE SÍMBOLOS

$\{x_i\}$	Conjunto de variáveis aleatórias
(a, m, b)	Número <i>fuzzy</i> triangular
m	<i>Core</i> do número <i>fuzzy</i> triangular
a	Extremo esquerdo do número <i>fuzzy</i> triangular
b	Extremo direito do número <i>fuzzy</i> triangular
\bar{X}	Média aritmética
c	Constante
n	Número de elementos
G	Média Geométrica
H	Média Harmônica
md	Mediana
Ω	Espaço amostral
\mathcal{P}	Função
\mathcal{F}	Família de subconjuntos
P	Probabilidade
A_1, A_2, A_3	Sucessão de eventos
$(\Omega, \mathcal{F}, \mathcal{P})$	Espaço de probabilidade
I	Imagem
\mathbb{R}	Conjunto dos números reais
$p(x_i)$	Função de probabilidade
$E(X)$	Valor esperado da variável X
$Var(X)$	Variância de variável X
μ	Média populacional
σ^2	Variância populacional
σ	Desvio padrão populacional
$A(x)$	Suporte de X

s^2	Variância amostral
s	Desvio padrão amostral
Z	Estatística de teste normal
λ	Parâmetro taxa da distribuição exponencial
β	Parâmetro escala da distribuição exponencial
θ	Parâmetro
$\hat{\theta}$	Estimador do parâmetro θ
$\mu_F(x)$	Função de pertinência número <i>fuzzy</i>
$supp(F)$	Suporte do subconjunto <i>fuzzy</i> F
$[F]^\alpha$	α -corte (α -nível) do subconjunto <i>fuzzy</i> F
D_n	Estatística do teste Kolmogorov-Smirnov
W	Estatística do teste Shapiro-Wilk
gl	Graus de liberdade
F^*	Estatística do teste Levene
F_c	Estatística do teste Welch
χ	Estatística do teste não paramétrico de Friedman
$mCheng$	Nome utilizado para o valor central do número <i>fuzzy</i> no método de <i>Cheng</i> .

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	17
1.1.1	Objetivo Geral	17
1.1.2	Objetivos Específicos	17
1.2	Estruturação do trabalho	17
2	REVISÃO DE LITERATURA	19
2.1	Medidas Descritivas	19
2.2	Distribuições de Probabilidade	22
2.2.1	Conceitos básicos	22
2.2.2	Distribuições teóricas	25
2.3	Convergência, Lei dos Grandes Números e Teorema do Limite Central	29
2.4	Distribuições Amostrais	32
2.5	Estimadores	34
2.6	Lógica Fuzzy	37
2.7	MCheng	39
3	METODOLOGIA	42
4	RESULTADOS	46
4.1	Verificação de algumas propriedades da medida mCheng	46
4.2	Análise dos resultados dos experimentos	50
4.2.1	Visão geral	50
4.2.2	Distribuição Normal	52
4.2.3	Distribuição Exponencial	64
5	CONSIDERAÇÕES FINAIS	76
	REFERÊNCIAS	79
	ANEXO A EXPRESSÕES MATEMÁTICAS	82
	APÊNDICE A PROGRAMAS CONSTRUÍDOS PARA AS SIMULAÇÕES	84
A.1	Simulações com distribuição normal	84
A.2	Simulações com distribuição exponencial	91

1 INTRODUÇÃO

Ao trabalhar com pesquisas quantitativas necessita-se utilizar várias técnicas estatísticas que podem colaborar na coleta, organização, análise e interpretação dos dados. A análise de dados quantitativos normalmente se inicia com uma análise exploratória para que o pesquisador se familiarize com os dados. De acordo com MATTOS; KONRATH; AZAMBUJA (2017), é através dela que são obtidas as primeiras informações sobre os dados sem se preocupar com suposições de algum modelo probabilístico. Entre estas informações iniciais relevantes estão a tendência central e a dispersão dos dados, que costumam ser utilizadas sob o enfoque da lógica formal.

De acordo com WEISBERG (1992), as medidas de tendência central têm por objetivo determinar o valor central do conjunto de dados. De acordo com MATTOS; KONRATH; AZAMBUJA (2017), de maneira geral, os dados tendem a ser mais numerosos em torno de um valor central, diminuindo gradativamente sua frequência a medida que se afastam deste valor, que pode ser chamado de centro da distribuição e pode ser calculado para um número finito de valores ou para uma distribuição teórica, como a distribuição normal. De acordo com DODGE (2003) e UPTON; COOK (2008), ocasionalmente, alguns autores utilizam a centralidade significando a tendência de dados quantitativos de se agruparem ao redor de um valor central. Nas análises estatísticas, as medidas de tendência central clássicas mais utilizadas são a média aritmética e a mediana. A moda, a média harmônica e a média geométrica podem também ser utilizadas.

De acordo com CASTRO (1970) e CRESPO (2009), as medidas de dispersão servem para caracterizar como o conjunto de valores se comporta em torno das medidas de tendência central. São importantes por poder destacar o grau de homogeneidade e heterogeneidade que existe nos valores que compõem o conjunto. As principais medidas de dispersão utilizadas são a variância, o desvio padrão e o desvio interquartilico. Ainda podem ser utilizadas a amplitude total, o desvio médio e o coeficiente de variação.

Muitos estudos têm sido desenvolvidos com a utilização da métrica *fuzzy*. Esta métrica é especialmente indicada para mensurar variáveis complexas, subjetivas ou mal definidas, estando fundamentada na teoria dos conjuntos *fuzzy*, a qual é decorrente de estudos do matemático Zadeh, desenvolvidos a partir de 1965. Para propor a lógica *fuzzy*, Zadeh

partiu do princípio da incompatibilidade que diz que "alta complexidade é incompatível com alta precisão". Isto significa que, se a mensuração for feita de maneira menos precisa, ela terá menos chance de errar. Por exemplo, se for necessário definir a idade de um indivíduo sem olhar seus documentos de identidade nem perguntar-lhe, é melhor dizer que ele tem em torno de 25 anos do que afirmar que ele tem 25 anos. Da mesma forma, em vez de associar o nível de satisfação de um indivíduo a um escore, por exemplo 95, numa escala de zero a cem, é melhor dizer que o indivíduo está muito satisfeito.

A lógica *fuzzy* pode trabalhar com variáveis linguísticas e propriedades que possibilitam desenvolver raciocínios mais próximos do humano em estruturas matemáticas computacionais. Alguns autores a consideram como uma extensão da lógica booleana, que trabalha apenas com o puramente verdadeiro e o puramente falso, enquanto outros autores consideram que a lógica booleana é um caso particular da lógica *fuzzy*, pois esta última admite vários graus de verdade entre o puramente verdadeiro e o puramente falso.

Para CAMPOS FILHO (2004), na lógica *fuzzy* o raciocínio exato corresponde a um limite do raciocínio aproximado, reportando mais informações e reduzindo a perda de importantes informações. Segundo BARROS; BASSANEZI (2006), ao propor a lógica *fuzzy*, a principal intenção de Zadeh era dar um tratamento matemático a termos linguísticos subjetivos e imprecisos, tais como: aproximadamente, em torno de, grande, etc. Dentro deste enfoque, ganham importância as variáveis linguísticas.

SHAW; SIMOES (2007) afirmam que nas áreas onde é necessário lidar com a imprecisão, a subjetividade e o desconhecimento, essa teoria tem demonstrado grande capacidade de aplicação, ajudando profissionais a produzir modelos condizentes com as necessidades e realidades, mesmo com dados que apresentem alto grau de incerteza. Consequentemente, pode ser dito que foi um dos primeiros passos para programar e armazenar conceitos vagos em computadores, possibilitando cálculos a partir de informações vagas.

Esses dados que apresentam um certo grau de incerteza e que podem estar associados a termos linguísticos podem ser definidos por conjuntos *fuzzy*, que se utilizam de funções de pertinência.

Entretanto, ao trabalhar com dados quantitativos, como as observações são associadas a números *fuzzy*? Um procedimento que pode ser utilizado com esta finalidade é o método apresentado em CHENG (2005), que encontra um número *fuzzy* a partir de um conjunto de dados obtidos sob o enfoque da lógica formal (números *crisp*). Neste método, um conjunto de números *crisp*, $\{x_1, x_2, x_3, \dots, x_n\}$, gera um número *fuzzy* triangular, (a, m, b) , existindo necessidade de estimar o centro do conjunto, identificado como m , e uma medida utilizada para estimar a variabilidade, usada na determinação dos valores extremos do número *fuzzy* triangular, identificados por a e b . A determinação desses valores na geração dos números *fuzzy* é importante, pois o valor central é o único valor do suporte do conjunto *fuzzy* com grau de pertinência máximo (um) e os limites são os valores do suporte com grau de pertinência mínimo (zero).

Os números *crisp* podem ser considerados como uma amostra de n observações. Neste caso, o valor central m , também denominado *core* em um número *fuzzy* triangular, está associado à sua tendência central e os limites estão associados à sua variabilidade, podendo ser tratados como variáveis aleatórias, e os valores encontrados para m , a e b podem ser considerados estimativas dos verdadeiros valores, razão pela qual é importante conhecer as suas distribuições amostrais, o que possibilita avaliar a qualidade da medida como estimador. O conhecimento da distribuição amostral de uma variável aleatória possibilita associar uma confiabilidade ao resultado encontrado. Porém, não foram encontrados estudos de como essas medidas se comportam em amostras, o que justifica o desenvolvimento de estudos com esta finalidade.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo geral deste estudo é conhecer a distribuição amostral da medida de tendência central utilizada na estimação do centro de um número *fuzzy* triangular pelo método proposto em CHENG (2005).

1.1.2 Objetivos Específicos

- Aprimorar os conhecimentos nas áreas de medidas descritivas e distribuições amostrais;
- analisar criticamente algumas propriedades de distribuições amostrais de algumas medidas de tendência central;
- analisar o procedimento de determinação da tendência central do número *fuzzy* pelo método proposto em Cheng (2005);
- realizar simulações para analisar o comportamento de algumas medidas de tendência central em amostras extraídas de populações com diferentes parâmetros;
- comparar a distribuição amostral da medida m_{Cheng} com as demais distribuições amostrais das estatísticas estudadas (média aritmética e mediana).

1.2 Estruturação do trabalho

Neste primeiro capítulo é apresentada a introdução com o tema e sua importância, o problema e a justificativa, além dos objetivos geral e específicos que nortearam o estudo. O capítulo seguinte traz a revisão de literatura, abordando os principais conceitos estudados para a realização dos experimentos, além de alguns conceitos básicos de lógica *fuzzy* e o método *Cheng*. No capítulo 3 é apresentada a metodologia do trabalho que

explica a forma de realização deste estudo. O capítulo 4 apresenta os resultados encontrados e respectivas discussões, abordando desde a análise das diversas etapas do método *Cheng* e suas propriedades, bem como a análise dos resultados dos experimentos. Por fim, são apresentadas as considerações finais. Complementam esta apresentação, o referencial teórico e dois anexos, o primeiro com algumas expressões matemáticas utilizadas e o segundo com os programas desenvolvidos e usados nas simulações.

2 REVISÃO DE LITERATURA

2.1 Medidas Descritivas

Na análise estatística de dados, um conceito muito abordado em relação a descrição dos dados é a tendência central. O objetivo é obter informações sobre o valor central, ou seja, um valor que represente o centro dos dados obtidos.

Segundo CRESPO (2009), a média aritmética (\bar{X}), que nesse estudo será chamada apenas de média, de um conjunto de n valores $\{x_1, x_2, \dots, x_n\}$ é obtida pelo quociente entre a soma desses e sua quantidade, conforme Eq. (1).

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Segundo CASTRO (1970), a média aritmética é um valor resumo que se propõe a representar com maior ou menor capacidade o conjunto de dados estudados. Para os autores CASTRO (1970) e CRESPO (2009), a média aritmética possui as seguintes propriedades:

Propriedade 2.1 *A soma algébrica dos desvios tomados em relação a média é nula, conforme Eq. (2):*

$$\sum_{i=1}^n (x_i - \bar{X}) = 0 \quad (2)$$

Propriedade 2.2 *Somando ou subtraindo uma constante (c) de todos os valores de uma variável, a média do conjunto fica aumentada ou diminuída dessa constante, conforme Eq. (3):*

$$y_i = x_i \pm c \Rightarrow \bar{Y} = \bar{X} \pm c \quad (3)$$

Propriedade 2.3 *Multiplicando ou dividindo todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada ou dividida por essa constante, conforme Eqs. (4) e (5):*

$$y_i = x_i \cdot c \Rightarrow \bar{Y} = \bar{X} \cdot c \quad (4)$$

$$y_i = \frac{x_i}{c} \Rightarrow \bar{Y} = \frac{\bar{X}}{c} \quad (5)$$

Propriedade 2.4 *A soma dos quadrados dos afastamentos contados a partir da média aritmética é um mínimo, da função, apresentada na Eq. (6):*

$$S = \sum_{i=1}^n (x_i - a)^2 \quad (6)$$

Se $a = \bar{X}$, S é um ponto de mínimo.

Além da média aritmética, pode-se utilizar outras médias para identificar a centralidade, como a geométrica e harmônica.

A média geométrica ou média proporcional para dois números x_1 e x_2 está definida em CASTRO (1970) como a razão entre esses dois números e um terceiro, de acordo com a Eq. (7).

$$\frac{x_1}{G} = \frac{G}{x_2} \quad (7)$$

Na Eq. (7), G representa o valor da média geométrica, calculada conforme Eq. (8).

$$G = \sqrt{x_1 \cdot x_2} \quad (8)$$

Para n valores, a média geométrica pode ser calculada conforme Eq. (9).

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n} \quad (9)$$

Essa forma de média é indicada para cálculos que envolvam variações percentuais em sequência. Também pode ser utilizada quando os dados apresentam assimetria positiva acentuada.

Ainda existe a média harmônica indicada para cálculos que envolvem valores que tem como unidade o quociente entre duas unidades de medidas, sendo calculada para dois números x_1 e x_2 , segundo CASTRO (1970), por meio da Eq. (10).

$$H = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} \quad (10)$$

A Eq. (10) pode ser transformada na Eq. (11):

$$H = \frac{2x_1x_2}{x_1 + x_2} \quad (11)$$

Para n valores, as Eqs. (10) e (11) podem ser generalizadas conforme Eq. (12).

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (12)$$

De acordo com MATTOS; KONRATH; AZAMBUJA (2017), a igualdade entre os resultados das três médias só ocorrerá quando o conjunto de dados for formado por observações idênticas. À medida que a variabilidade entre as observações aumentar, crescerá a diferença entre os resultados desses três tipos de média, que se relacionam de acordo com a expressão: $H < G < \bar{X}$. A média aritmética é mais influenciada pelos valores mais altos das observações, enquanto a média harmônica, pelos valores mais baixos.

Para casos em que a utilização da média aritmética é duvidosa, como em casos que os dados apresentam assimetria acentuada, pode-se utilizar a mediana como principal medida de tendência central. Segundo CRESPO (2009), a mediana é definida como o número que se encontra na posição central do conjunto de números, estando estes ordenados em ordem crescente. A mediana pode ser definida conforme a Eq. (13).

$$md(X) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ par} \end{cases} \quad (13)$$

Segundo ANGELINI; MILONE (1993a), a mediana possui as seguintes propriedades:

Propriedade 2.5 *A mediana não é influenciada pelos valores extremos.*

Propriedade 2.6 *Somando ou subtraindo uma constante (c) de todos os valores de uma variável, a mediana do conjunto fica aumentada ou diminuída dessa constante, conforme Eq. (14):*

$$y_i = x_i \pm c \Rightarrow md_{(Y)} = md_{(X)} \pm c \quad (14)$$

Propriedade 2.7 *Multiplicando ou dividindo todos os valores de uma variável por uma constante (c), a mediana do conjunto fica multiplicada ou dividida por essa constante, conforme Eqs. (15) e (16):*

$$y_i = x_i \cdot c \Rightarrow md_{(Y)} = md_{(X)} \cdot c \quad (15)$$

$$y_i = \frac{x_i}{c} \Rightarrow md_{(Y)} = \frac{md_{(X)}}{c} \quad (16)$$

De acordo com MATTOS; KONRATH; AZAMBUJA (2017), quando a distribuição dos dados for simétrica, as duas medidas de tendência central, média aritmética e mediana são iguais, ou seja, $\bar{X} = Md$. No caso de existir uma assimetria positiva, a média

aritmética tem o valor maior que a mediana, $\bar{X} > Md$. Se a assimetria for negativa, a média aritmética apresenta valor menor que a mediana, $\bar{X} < Md$.

Salienta-se que, quando estas medidas representam a tendência central de uma amostra, elas se caracterizam como variáveis aleatórias, sendo seu comportamento descrito por uma distribuição de probabilidade.

2.2 Distribuições de Probabilidade

Na vida real, muitas vezes é necessário tomar decisões que dependem de resultados de fenômenos aleatórios, sendo necessário avaliar todas as possibilidades de resultados e suas respectivas probabilidades de ocorrência, o que pode ser feito por meio das distribuições de probabilidade. A própria inferência estatística se fundamenta na teoria da probabilidade para tomar decisões sobre uma população a partir da análise de uma pequena parcela desta população.

2.2.1 Conceitos básicos

Chama-se espaço amostral (Ω) o conjunto de todos os possíveis resultados de um experimento. Ao lançar uma moeda e observar a face que cai voltada para cima, por exemplo, pode-se obter a face cara (C) ou a face coroa (K), logo o espaço amostral é: $\Omega = \{C, K\}$.

Um evento é qualquer subconjunto de possíveis resultados de um experimento aleatório, incluindo o próprio espaço amostral. No espaço amostral correspondente ao lançamento de uma moeda, por exemplo, podem ser definidos os eventos representados por: \emptyset , $\{C\}$, $\{K\}$ e $\{C, K\}$. Neste caso, \emptyset representa a não ocorrência da face cara nem da face coroa, $\{C\}$ representa a ocorrência da face cara, $\{K\}$ representa a ocorrência da face coroa e $\{C, K\}$ representa a ocorrência da face cara ou da face coroa.

Existem vários conceitos de probabilidade. Por volta de 1930, Kolmogorov apresentou um conjunto de axiomas para uma formulação matemática mais rigorosa de probabilidade, onde as demais definições se enquadram como caso particular.

Uma função \mathcal{P} , definida numa classe \mathcal{F} de subconjuntos de Ω , e assumindo valores no intervalo $[0; 1]$, é uma probabilidade se satisfaz os Axiomas de Kolmogorov:

Axioma 2.1 $P(\Omega) = 1$;

Axioma 2.2 Para todo subconjunto $A \in \mathcal{F}$, $P(A) \geq 0$;

Axioma 2.3 Para toda a sequência $A_1, A_2, A_3 \in \mathcal{F}$, mutuamente exclusivos, tem-se a Eq. (17).

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots \quad (17)$$

A Eq. (17), pode ser escrita como o apresentado na Eq. (18).

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (18)$$

A terna $(\Omega, \mathcal{F}, \mathcal{P})$ é denominada espaço de probabilidade e os conjuntos que estão em \mathcal{F} , que são os subconjuntos de Ω , são os possíveis eventos.

Seja $(\Omega, \mathcal{F}, \mathcal{P})$ um espaço de probabilidade. Denomina-se variável aleatória a uma função do espaço amostral Ω em \mathbb{R} , que associa cada ponto amostral ou conjunto pertencente à classe \mathcal{F} a um número real. Isto significa que para cada valor $x \in \Omega$, existe um $f(x) \in I \subset \mathbb{R}$.

Mais formalmente, uma variável aleatória pode ser assim definida. Seja: $f : \Omega \rightarrow \mathbb{R}$ uma variável aleatória. Então $f^{-1}(x) = \{x \in \Omega : f(x) \in I\} \in \mathcal{F}$, sendo $I \subset \mathbb{R}$. Isto significa que o domínio da função é o espaço amostral e a sua imagem está contida no conjunto dos números reais.

Observe que I pode ser um conjunto enumerável ou não enumerável. Se for um conjunto enumerável, a variável aleatória é dita discreta; em caso contrário, é dita contínua.

Para uma variável aleatória discreta (VAD), a função de probabilidade (fp) ou função de massa de probabilidade (fmp) de uma variável aleatória X , definida em $(\Omega, \mathcal{F}, \mathcal{P})$, é uma função que associa cada um de seus possíveis valores à sua probabilidade de ocorrência.

Considerando $X = \{x_1, x_2, x_3, \dots, x_i, \dots\}$, tem-se $p(x_i) = P(X = x_i)$.

Esta função de probabilidade expressa como a probabilidade total é distribuída entre os valores possíveis da variável aleatória X e tem de satisfazer as duas condições expressas nas Eqs. (19) e (20).

$$0 \leq p(x_i) \leq 1, \text{ para } i = 1, 2, 3, \dots \quad (19)$$

$$\sum_i p(x_i) = 1 \quad (20)$$

Para uma variável aleatória contínua (VAC), a função de densidade de probabilidade (fdp) de uma variável aleatória X , definida em $(\Omega, \mathcal{F}, \mathcal{P})$, é uma função $f(x)$ não negativa definida de tal forma que a área compreendida entre esta função e o eixo das abcissas representa probabilidade. Para que uma função represente uma fdp é necessário que satisfaça as condições expressas nas Eqs. (21) e (22).

$$f(x) \geq 0 \text{ para todo } x \in \mathbb{R} \quad (21)$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad (22)$$

Assim como é útil usar medidas para resumir um conjunto de dados, é possível sumarizar uma distribuição de probabilidade através de medidas, o que é feito representando a tendência central e a variabilidade.

Para representar a tendência central é utilizada uma média aritmética calculada entre os possíveis valores da variável aleatória X , ponderados por sua probabilidade de ocorrência. Esta medida normalmente é chamada de valor esperado, podendo também ser denominada esperança matemática, expectância ou, simplesmente, média.

Em uma distribuição de probabilidade discreta, esse valor é expresso pela Eq. (23).

$$E(X) = \mu = \sum_{i=1}^n [x_i \cdot p(x_i)] \quad (23)$$

Em uma distribuição de probabilidade contínua, o valor esperado é expresso pela Eq. (24).

$$E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)dx \quad (24)$$

Além do valor esperado de uma distribuição, é interessante conhecer a sua variabilidade que indica o quanto os possíveis resultados se afastam deste centro. Valores pequenos significam que é muito provável que valores de X estejam muito próximos de seu valor esperado. Nas distribuições de probabilidade, a variância é uma média quadrática dos afastamentos em relação ao valor esperado, ponderados pela respectiva probabilidade de ocorrência.

Em uma distribuição de probabilidade discreta, a variância é encontrada pela Eq. (25).

$$Var(X) = \sum_{i=1}^n [(x_i - \mu)^2 \cdot p(x_i)] \quad (25)$$

A Eq. (25) pode ser reescrita como a Eq. (26), onde o cálculo do valor de $E(X^2)$ é descrito na Eq. (27).

$$Var(X) = E(X^2) - [E(X)]^2 \quad (26)$$

$$E(X^2) = \sum_{i=1}^n [x_i^2 \cdot p(x_i)] \quad (27)$$

Em uma distribuição de probabilidade contínua, a variância é encontrada pela Eq. (28).

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx \quad (28)$$

Segundo BOLFARINE; SANDOVAL (2010), o modelo probabilístico (ou estatístico) é de suma importância para inferir resultados da amostra para a população. Algumas variáveis podem apresentar padrões de comportamento similares, diferenciando-se apenas por algumas especificidades. Dentro desse contexto, assumem importante papel os modelos teóricos.

2.2.2 Distribuições teóricas

No estudo de uma variável é importante que, o modelo teórico selecionado, na medida do possível, represente a complexidade do comportamento da variável na população em estudo. Dentre os principais modelos teóricos de probabilidade que descrevem o comportamento de variáveis contínuas, estão o modelo normal e o modelo exponencial, descritos a seguir.

2.2.2.1 Distribuição Normal

A distribuição normal é sem dúvida a mais importante distribuição contínua. Essa importância se deve a vários fatores, um deles é o Teorema do Limite Central. Essa distribuição é também conhecida como distribuição gaussiana e é utilizada para modelar diversos fenômenos naturais. Uma das primeiras aparições dessa distribuição foi nos estudos de Abraham de Moivre, em 1733, com um estudo aprofundado de fatorial ($n!$). Em 1756, Moivre publicou o livro *A Doutrina das Chances*, em que a distribuição normal aparece como limite de uma distribuição binomial. Em 1781, Laplace publicou a primeira tabela da distribuição normal, e em 1821, ao elaborar o primeiro Teorema do Limite Central (chamado por teorema de Laplace), essa distribuição é totalmente definida.

Definição 2.1 *X tem distribuição normal com média μ e variância σ^2 , escrito como $X \sim N(\mu, \sigma^2)$, se a função densidade de probabilidade de X é dada pela Eq. (29) (BOLFARINE; SANDOVAL (2010)).*

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad (29)$$

Na Eq. (29), $-\infty < \mu < \infty$ e $\sigma^2 > 0$. Nesse caso, μ e σ^2 são denominados parâmetros da distribuição e o suporte de X, isto é $A(x) = \{x, f(x) > 0\}$, é uma reta. A representação gráfica dessa distribuição é apresentada na Fig. 1.

O modelo normal, é bastante utilizado no estudo de características populacionais que envolvem peso, altura, pressão arterial, quociente de inteligência, entre outros. Essa distribuição é simétrica em relação à média μ e tem como valor esperado $E(X) = \mu$ e variância $Var(X) = \sigma^2$.

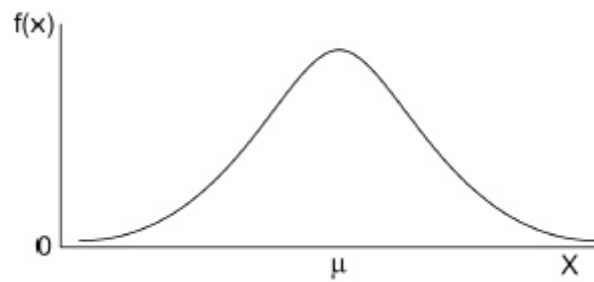


Figura 1: Distribuição normal: representação gráfica

Para achar a área sob a curva, é necessário conhecer a média (μ) e o desvio padrão (σ). Na Fig. 2 estão representadas algumas áreas importantes.

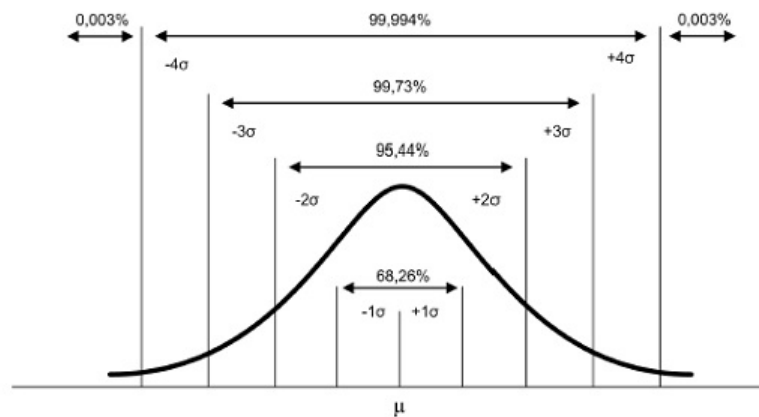


Figura 2: Áreas importantes sobre a curva normal

Para casos mais comuns em que os parâmetros populacionais (μ e σ) são desconhecidos, pode-se utilizar os respectivos parâmetros amostrais (\bar{X} e s , respectivamente). A média amostral (\bar{X}) pode ser calculada conforme a Eq. (1), apresentada anteriormente, e o desvio padrão amostral (s) pode ser calculado conforme Eq. (30).

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2} \quad (30)$$

Para cada valor diferente de μ e σ , no caso populacional, ou \bar{X} e s , no caso amostral, a curva normal tem uma representação diferente. Para o cálculo das áreas específicas se utiliza a distribuição normal padrão, que pode ser também chamada de distribuição normal padronizada ou reduzida, e se caracteriza por apresentar média $\mu = 0$ e desvio padrão $\sigma = 1$. A representação gráfica da distribuição normal padrão é apresentada na Fig. 3.

Para obter uma curva normal padrão a partir de uma curva normal com parâmetros diferentes de $\mu = 0$ e $\sigma = 1$ essa deve ser reduzida a uma variável Z , conforme Eq. (31).

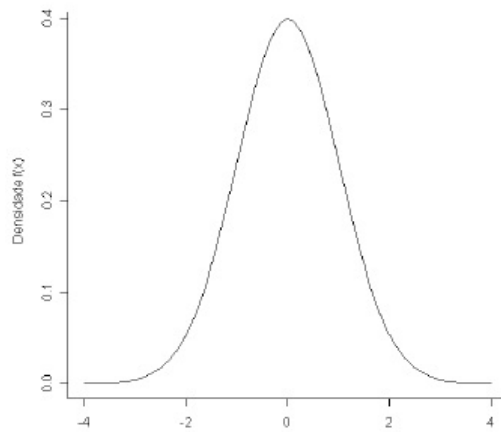


Figura 3: Distribuição normal padrão: representação gráfica

$$Z = \frac{x - \mu}{\sigma} \quad (31)$$

A partir da redução efetuada pela Eq. (31), a distribuição passa a ter média $\mu = 0$ e desvio padrão $\sigma = 1$. Pelo fato da distribuição normal ser simétrica em relação à média $\mu = 0$, as áreas a direita e a esquerda de μ são iguais. Por ser uma distribuição muito utilizada, existem tabelas que apresentam o resultado das integrais, que fazem o cálculo conforme Eq. (32).

$$\Phi(Z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (32)$$

Essa integral não pode ser resolvida pelo método tradicional (teorema fundamental do cálculo), somente por métodos numéricos. Por esse motivo, utilizam-se tabelas que apresentam os resultados das integrais.

2.2.2.2 Distribuição Exponencial

O outro modelo estudado, o exponencial, se caracteriza por apresentar um único parâmetro que indica a taxa de ocorrência por unidade de medida, normalmente tempo, distância ou volume. Ele é considerado um dos mais simples em termos matemáticos e tem sido usado frequentemente para descrever o tempo que se leva para completar uma tarefa ou tempo de duração de um equipamento.

Definição 2.2 Uma variável aleatória X , assumindo valores não negativos, segue o modelo Exponencial com parâmetro $\lambda > 0$, escrita como $X \sim \text{Exp}(\lambda)$, se sua densidade é conforme Eq. (33).

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0; \\ 0, & \text{caso contrário.} \end{cases} \quad (33)$$

Na Eq. (33), λ é o parâmetro taxa da distribuição, satisfazendo a condição $\lambda > 0$. O parâmetro deve ter a mesma unidade da variável x , e o suporte de X , isto é $A(x) = \{x, x > 0\}$. A representação gráfica dessa distribuição é apresentada na Fig. 4.

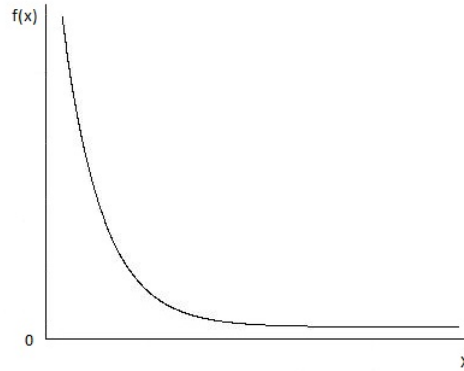


Figura 4: Distribuição exponencial: representação gráfica

Essa distribuição tem uma assimetria acentuada positiva em relação à média μ , e tem como valor esperado $E(X) = \frac{1}{\lambda}$ e variância $Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}$. A função de distribuição acumulada $F(x)$ é dada pela Eq. (34) e sua representação gráfica é apresentada na Fig. 5.

$$F(X) = P(X \leq x_0) = \int_0^{x_0} f(x)dx = \begin{cases} 1 - e^{-\lambda x}, & \text{se } x \geq 0; \\ 0, & \text{se } x < 0 \end{cases} \quad (34)$$

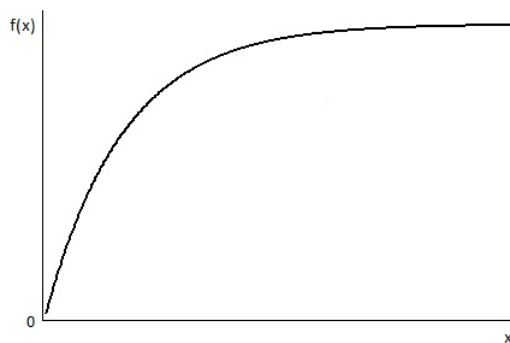


Figura 5: Distribuição exponencial acumulada: representação gráfica

A distribuição Exponencial pode ser parametrizada de uma forma alternativa segundo a função densidade de probabilidade dada pela Eq. (35).

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}}, & \text{se } x \geq 0; \\ 0, & \text{se } x < 0 \end{cases} \quad (35)$$

Neste caso, $\beta > 0$ é o parâmetro de escala da distribuição e é o inverso do parâmetro taxa na definição 2.2. A denotação nesse caso é $X \sim Exp(\beta)$ e essa definição alternativa torna-se ambígua. Portanto, sempre deve-se verificar qual das duas especificações está

sendo utilizada, se o parâmetro utilizado é o parâmetro taxa (λ) ou o parâmetro escala (β) da distribuição.

2.3 Convergência, Lei dos Grandes Números e Teorema do Limite Central

Segundo WASSERMAN (2003), um dos aspectos mais importantes da teoria da probabilidade diz respeito ao comportamento de sequências de variáveis aleatórias, podendo ser chamada de teoria da grande amostra, teoria limite ou ainda de teoria assintótica. Essa teoria baseia-se em três conceitos principais: convergência, Lei dos Grandes Números e Teorema do Limite Central, apresentados a seguir.

Definição 2.3 (Convergência em Probabilidade) *Uma sequência de variáveis aleatórias X_1, X_2, \dots, X_n , converge em probabilidade para uma variável aleatória X se, para cada $\varepsilon > 0$, verifica-se a Eq. (36).*

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0 \quad (36)$$

Isto significa que a probabilidade de X_n divergir de X tende a zero a medida que n tende ao infinito. Portanto, a partir da Eq. (36), deduz-se a Eq. (37).

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1 \quad (37)$$

Observe que, se a sequência de variáveis aleatórias X_n converge em probabilidade para X a medida que n tende ao infinito, a probabilidade da diferença $|X_n - X|$ ser infinitesimal tende a um. Salienta-se que as variáveis da sequência são independentes e identicamente distribuídas (iid).

Uma consequência do conceito de convergência em probabilidade para uma sequência de variáveis aleatórias é a Lei Fraca dos Grandes Números, enunciada a seguir.

Lei 2.1 (Lei Fraca dos Grandes Números) *Sejam X_1, X_2, \dots, X_n variáveis aleatórias iid, com $E(X_i) = \mu$ e $Var(X_i) = \sigma^2 < \infty$. A partir da média aritmética, Eq. (1), para cada $\varepsilon > 0$, define-se a Eq. (38).*

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1 \quad (38)$$

A Eq. (38) significa que a média amostral, \bar{X}_n , converge em probabilidade para μ . Portanto, de acordo com esta lei, sob condições gerais, a média amostral se aproxima da média da população a medida que o tamanho da amostra tende ao infinito. A propriedade definida por esta lei, de que a média de uma sequência de amostras de mesmo tamanho

se aproxima de uma constante a medida que o tamanho da amostra tende ao infinito, é conhecida como consistência.

Definição 2.4 (Convergência Quase Certa) *Uma sequência de variáveis aleatórias, X_1, X_2, \dots, X_n , converge quase certamente para uma variável aleatória X se, para cada $\varepsilon > 0$, tem-se a Eq. (39).*

$$P\left(\lim_{n \rightarrow \infty} |X_n - X| < \varepsilon\right) = 1 \quad (39)$$

A convergência quase certa é mais forte do que a convergência em probabilidade. Então, se uma sequência converge quase certamente, essa mesma sequência também converge em probabilidade. Uma consequência do conceito de convergência quase certa é a Lei Forte dos Grandes Números, enunciada a seguir. Esta lei é a análoga mais forte da Lei Fraca dos Grandes Números, indicando a convergência para uma constante.

Lei 2.2 (Lei Forte dos Grandes Números) *Sejam X_1, X_2, \dots, X_n variáveis aleatórias iid, com $E(X_i) = \mu$ e $Var(X_i) = \sigma^2 < \infty$. Defina-se a média aritmética, Eq. (1). Então para cada $\varepsilon > 0$, define-se a Eq. (40).*

$$P\left(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \varepsilon\right) = 1 \quad (40)$$

Isto significa que a média amostral converge para uma constante, que é a média da população. Tanto para a Lei Fraca como para a Lei Forte dos Grandes Números foi feita a suposição de uma variância finita. Esta suposição é desejável, mas ambas as leis se mantêm sem ela.

Definição 2.5 (Convergência em Distribuição) *Uma sequência de variáveis aleatórias iid, X_1, X_2, \dots, X_n , converge em distribuição para uma variável aleatória X , se em todos os pontos x onde $F_{X_n}(x)$ é contínua, tem-se Eq. (41).*

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (41)$$

Observe que na verdade são as funções de distribuição acumulada que convergem em distribuição e não as variáveis aleatórias. Como consequência deste tipo de convergência, tem-se os teoremas apresentados a seguir.

Teorema 2.1 *Uma sequência de variáveis aleatórias, X_1, X_2, \dots, X_n , converge em probabilidade para uma constante μ se, e somente se, a sequência também convergir em distribuição para μ , conforme mostra a Eq. (42).*

$$(P|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0 \text{ para todo } \varepsilon > 0. \quad (42)$$

A Eq. (42) é equivalente a Eq. (43).

$$P(X_n \leq x) \rightarrow \begin{cases} 0, & \text{se } x < \mu \\ 1, & \text{se } x > \mu \end{cases} \quad (43)$$

O Teorema do Limite Central usa a convergência em distribuição para mostrar que uma soma de variáveis aleatórias independentes e convenientemente padronizadas converge para um modelo Normal.

Teorema 2.2 (Teorema do Limite Central) *Suponha uma sequência de variáveis aleatórias iid, X_1, X_2, \dots, X_n , com $E(x_i) = \mu$ e $Var(X_i) = \sigma^2 > 0$, ambos finitos. Definida a Eq. (1). Seja $F_n(x)$ denotada como função de distribuição acumulada (fda) de $\frac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n}$. Então, para qualquer x no intervalo $-\infty < x < \infty$, $\frac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n}$ tem distribuição normal padrão limite, conforme Eq. (44).*

$$\lim_{n \rightarrow \infty} F_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \quad (44)$$

Este teorema fornece uma aproximação, mas não existe uma maneira automática de saber o quanto esta aproximação é satisfatória. A suposição de variância finita é necessária à convergência para a normalidade, que se origina da soma de pequenos distúrbios.

O teorema do limite central tem algumas formulações específicas, como as citadas por MAGALHÃES (2006).

Teorema 2.3 (Teorema do Limite Central de Moivre-Laplace) *Seja X_1, X_2, \dots, X_n variáveis independentes, seguindo o modelo Bernoulli com parâmetro p . Assim, $\mu = E(X_n) = p$ e $\sigma^2 = Var(X_n) = p(1 - p)$. Para $S_n = X_1 + X_2 + \dots + X_n$, temos, pelo teorema anterior, que a Eq. (44) pode ser escrita conforme a Eq. (45).*

$$\frac{S_n - np}{\sqrt{np(1 - p)}} \xrightarrow{d} N(0, 1) \quad (45)$$

Teorema 2.4 (Teorema do Limite Central de Liapunov) *Seja X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes, com esperança μ_n e variância σ_n^2 , com $\sigma_n^2 < \infty$ e pelo menos um dos σ_n^2 s maior que zero. Sejam $S_n = X_1 + X_2 + \dots + X_n$ e $s_n^2 = \sum_{k=1}^n \sigma_k^2$. Se a condição de Liapunov estiver satisfeita, ou seja, se existir $\delta > 0$ tal que a Eq. (46) exista, então a Eq. (47) representa o limite central dessa sequência.*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2 + \delta}} \sum_{k=1}^n E(|X_k - \mu_k|^{2 + \delta}) = 0 \quad (46)$$

$$\frac{S_n - E(S_n)}{s_n} \xrightarrow{d} N(0, 1) \quad (47)$$

Teorema 2.5 (Teorema do Limite Central de Lindeberg-Feller) *Seja X_1, X_2, \dots, X_n uma seqüência de variáveis aleatórias independentes, com distribuição F_n , esperança μ_n e variância σ_n^2 , com $\sigma_n^2 < \infty$ e pelo menos um dos σ_n^2 s maior que zero. Sejam $S_n = X_1 + X_2 + \dots + X_n$ e $s_n^2 = \sum_{k=1}^n \sigma_k^2$. Então a Eq. (48) representa o limite central dessa seqüência se, e somente se, vale a condição de Lindeberg na Eq. (49).*

$$\frac{S_n - E(S_n)}{s_n} \xrightarrow{d} N(0, 1) \text{ e } \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{\sigma_k}{s_n} = 0 \quad (48)$$

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \frac{1}{s^2} \sum_{k=1}^n \int_{\mu_k - \varepsilon s_n}^{\mu_k + \varepsilon s_n} (x - \mu_k)^2 dF_k(x) = 1 \quad (49)$$

Diante do exposto, conclui-se que:

- os conceitos de convergência descrevem diferentes modos pelos quais a distribuição de X_n converge para alguma distribuição limite a medida que o tamanho da amostra se torna grande;
- as Leis dos Grandes Números dizem que a média amostral converge em probabilidade ao valor esperado, $E(X_i) = \mu$, o que significa que \bar{X}_n está perto de μ com elevada probabilidade;
- o Teorema do Limite Central diz que a média padronizada converge em distribuição a uma distribuição normal, indicando que a distribuição que descreve o comportamento desta medida amostral se aproxima a uma distribuição normal, melhorando esta aproximação para valores elevados de n .

2.4 Distribuições Amostrais

Na Inferência Estatística frequentemente tomam-se decisões sobre uma população a partir de informações obtidas em uma pequena parcela desta, ou seja, em uma amostra, fazendo uso de estatísticas. Uma estatística é qualquer função das observações de uma amostra, ou seja, qualquer quantidade calculada a partir de dados obtidos em uma amostra.

Seja X uma variável aleatória que segue algum modelo probabilístico. Diz-se que um conjunto de observações X_1, X_2, \dots, X_n é uma amostra aleatória se os X_i são iid, com a mesma distribuição de X .

Considere que foram extraídas todas as amostras possíveis de tamanho n dessa população, empregando o mesmo plano de amostragem e sendo calculada uma estatística para cada uma das amostras. Os valores desta estatística formam uma nova população e sua distribuição de probabilidade recebe o nome de distribuição amostral. Lembre que qualquer função de variáveis aleatórias é também uma variável aleatória, podendo seu comportamento ser descrito por algum modelo probabilístico.

Qualquer medida descritiva, tais como a média aritmética e a mediana, é uma estatística quando calculada a partir de dados amostrais. Segundo ANGELINI; MILONE (1993b), no estudo do parâmetro populacional média, representado por μ , a estatística correspondente a ele, chamada de média amostral, representada por \bar{X} , está associada à Distribuição Amostral das Médias.

Para analisar essa distribuição, é preciso considerar uma população X , com os parâmetros de média populacional, $\mu = E(X)$, e variância populacional, $\sigma^2 = Var(X)$, conhecidos. Ao retirar todas as possíveis amostras aleatórias simples de tamanho n dessa população, e calcular a média \bar{X} para cada uma dessas amostras, obtém-se a distribuição amostral das médias, que apresentará algumas propriedades, apresentadas sob a forma de dois teoremas, definidos a partir do Teorema do Limite Central.

Teorema 2.6 *Seja X uma variável aleatória com média μ e variância σ^2 , e seja X_1, X_2, \dots, X_n , uma amostra aleatória simples de X . Então, o valor esperado e a variância são, respectivamente, $E(\bar{X}) = \mu$ e $Var(\bar{X}) = \frac{\sigma^2}{n}$. (BUSSAB; MORETTIN, 2005).*

Teorema 2.7 *Seja X uma variável aleatória com média μ e variância σ^2 . Para amostras aleatórias, X_1, X_2, \dots, X_n , retiradas desta população. a distribuição amostral da média \bar{X} aproxima-se para um n suficientemente grande, de uma distribuição normal com valor esperado e variância, respectivamente, iguais a $E(\bar{X}) = \mu$ e $Var(\bar{X}) = \frac{\sigma^2}{n}$. (BUSSAB; MORETTIN, 2005).*

De acordo estes teoremas, a distribuição de \bar{X} aproxima-se de uma distribuição normal quando n tende ao infinito e esta convergência depende da distribuição da população da qual a amostra é retirada. Se a distribuição for normal, a convergência é rápida. Em caso contrário, a convergência é tanto mais lenta quanto mais afastada de uma distribuição normal estiver a população. Alguns autores, entre os quais BUSSAB; MORETTIN (2005), consideram que esta aproximação satisfatória a partir de 30 observações.

Considere novamente uma população X , com os parâmetros de média populacional, $\mu = E(X)$, e variância populacional, $\sigma^2 = Var(X)$, conhecidos. Ao retirar todas as possíveis amostras aleatórias simples de tamanho n dessa população, e calcular a mediana md para cada uma dessas amostras, obtém-se a distribuição amostral da mediana. Segundo BUSSAB; MORETTIN (2005), a distribuição amostral desta estatística, quando

obtida de amostras de uma população normal com parâmetros conhecidos, também segue aproximadamente uma distribuição normal com valor esperado e variância dados respectivamente por $E(md) = \mu$ e $Var(md) = \frac{\pi\sigma^2}{2n}$.

As estatísticas e suas respectivas distribuições amostrais podem ser utilizadas para estimar parâmetros populacionais.

2.5 Estimadores

Segundo BOLFARINE; SANDOVAL (2010), os estimadores são responsáveis por determinar os melhores parâmetros θ , para avaliar uma amostra de uma variável aleatória X (população), por meio de propriedades estatísticas. Os parâmetros são quantidades, em geral desconhecidas, de uma população, sobre as quais tem-se interesse.

Os estimadores são obtidos através de uma fórmula dos valores amostrais de $f(X) = \{X_1, X_2, \dots, X_n\}$, denotado por $\hat{\theta}(d_s)$ ou $\hat{\theta}$. As principais qualidades procuradas nas amostragens são: pequenos vieses (vícios) e pequenas variâncias.

As características base da distribuição de $\hat{\theta}$ são sua média e sua variância, conforme Eqs. (50) e (51), respectivamente.

$$\mu[\hat{\theta}] = E[\hat{\theta}] \quad (50)$$

$$\sigma^2[\hat{\theta}] = E(\hat{\theta} - \mu[\hat{\theta}])^2 \quad (51)$$

O desvio padrão de $\hat{\theta}$, é calculado pela Eq. (52) e é denominado "erro padrão" de $\hat{\theta}$.

$$\sigma[\hat{\theta}] = \sqrt{Var(\hat{\theta})} \quad (52)$$

O erro amostral representa a diferença entre o valor do estimador $\hat{\theta}$ e o parâmetro a ser estimado θ . Só pode ser calculado, se e somente se, o parâmetro θ for improvável de ser encontrado, através da Eq. (53).

$$\varepsilon = \hat{\theta} - \theta \quad (53)$$

O viés ou vício de um parâmetro representa a diferença entre a média da distribuição de $\hat{\theta}$ e o parâmetro a ser estimado θ . Este valor é fixo por estimador, podendo ser zero ou não, calculado pela Eq. (54).

$$Viés(\hat{\theta}) = E[\hat{\theta}] - \theta \quad (54)$$

O erro quadrático médio é uma variância que mede a dispersão do estimador em torno do verdadeiro parâmetro, ao invés de medir em torno da média, calculado pela Eq. (55).

$$EQM[\hat{\theta}] = E_A[\hat{\theta} - \theta]^2 \quad (55)$$

A relação entre $EQM[\hat{\theta}]$ e $\sigma^2[\hat{\theta}]$ pode ser verificada conforme Eq. (56).

$$EQM_A[\hat{\theta}] = \sigma^2[\hat{\theta}] + Viés[\hat{\theta}] \quad (56)$$

Propriedade 2.8 (Viés) *Um estimador $\hat{\theta}$ é dito não-tendencioso, segundo um plano amostral A , se a Eq. (57) é válida.*

$$E_A[\hat{\theta}] = \theta \quad (57)$$

Caso o estimador seja viciado, o viés do estimador $\hat{\theta}$, segundo o plano amostral A , é dado pela Eq. 58.

$$Viés[\hat{\theta}] = E_A[\hat{\theta} - \theta] = E_A[\hat{\theta}] - \theta \quad (58)$$

A não-tendenciosidade é uma qualidade desejável de um estimador, porém não é suficiente como critério, pois toda a média ponderada dos valores amostrais é um estimador não-tendencioso da sua média populacional.

Propriedade 2.9 (Precisão ou Eficiência) *Um estimador não viciado é considerado o mais eficiente entre seus semelhantes, se a sua variância $\sigma^2[\hat{\theta}]$ for menor a variância de qualquer outro estimador não-enviesado de θ .*

Propriedade 2.10 (Consistência) *Um estimador $\hat{\theta}$ é consistente, se a medida que o tamanho de amostra aumenta, seu valor esperado converge para o parâmetro de interesse e sua variância converge para zero, conforme Eqs. (59) e (60), respectivamente.*

$$\lim_{n \rightarrow \infty} E[\hat{\theta}] = \theta \quad (59)$$

$$\lim_{n \rightarrow \infty} \sigma^2[\hat{\theta}] = 0 \quad (60)$$

Propriedade 2.11 (Validade ou acurácia) *Dados dois estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$ de um mesmo parâmetro θ , dizemos que $\hat{\theta}_1$ é mais acurado que $\hat{\theta}_2$ se a Eq. 61 se verifica.*

$$EQM[\hat{\theta}_1] < EQM[\hat{\theta}_2] \quad (61)$$

A eficiência relativa de $\hat{\theta}_1$ em relação a $\hat{\theta}_2$ é definida pela Eq. 62.

$$\frac{EQM[\hat{\theta}_1]}{EQM[\hat{\theta}_2]} \quad (62)$$

Na escolha dos estimadores é preferível escolher um estimador não-tendencioso a um tendencioso. Entretanto, se tivermos que escolher, entre um estimador tendencioso que tenha sua distribuição concentrada na vizinhança do verdadeiro valor θ e um não-tendencioso com grande variância, podemos escolher o tendencioso.

A inferência estatística trabalha com vários métodos de estimação de parâmetros, os que mais se destacam são: método dos momentos, método da máxima verossimilhança, método dos momentos-L e método dos mínimos quadrados.

O método da máxima verossimilhança é considerado o método de estimação mais eficiente, pois produz estimadores de menor variância, mas para alguns casos, a maior eficiência é apenas assintótica, o que faz com que a aplicação em amostras de pequeno tamanho produzam estimadores de qualidade comparável ou inferior aos outros métodos. Os estimadores desse método são consistentes, suficientes e assintoticamente sem viés, entretanto, para amostras finitas podem apresentar um viés que pode ser corrigido. De maneira geral, esse método busca, a partir de um conjunto de dados amostrais e um modelo estatístico, estimar os valores dos diferentes parâmetros do modelo estatístico de maneira que maximizem a probabilidade dos dados observados, ou seja, busca parâmetros que maximizem a função de verossimilhança.

O método do momentos é baseado nos momentos teóricos e amostrais das variáveis aleatórias envolvidas e é o método de estimação mais simples e de estimadores, em geral, de qualidade inferior e menos eficiente do que os de outros métodos, principalmente em distribuições com três ou mais parâmetros.

O método dos momentos-L produz estimadores de parâmetros comparáveis aos produzidos pelo método de máxima verossimilhança, com a vantagem de menor esforço computacional para a solução de sistemas de equações. Frequentemente, para amostras pequenas, esses estimadores são mais acurados do que os estimadores de máxima verossimilhança. Esse método consiste em igualar os momentos-L populacionais aos momentos-L amostrais, o resultado dessa operação produzirá as estimativas dos parâmetros da distribuição de probabilidade.

O método dos mínimos quadrados consiste em minimizar o quadrado das diferenças entre os valores observados de uma amostra e seus respectivos valores esperados, essas diferenças são chamadas de resíduos. O método procura um estimador que minimiza a soma dos quadrados dos resíduos da regressão, de forma que maximize o grau de ajuste do modelo aos dados observados. O método dos mínimos quadrados apresenta alguns requisitos: que o fator imprevisível (erro) seja distribuído aleatoriamente e que essa distribuição seja normal e independente, e que o modelo seja linear nos parâmetros, ou seja, que as variáveis apresentam uma relação linear entre si.

2.6 Lógica Fuzzy

Outro conceito utilizado nesse estudo foi o de lógica *fuzzy*. A palavra *fuzzy* vem do inglês que significa algo vago ou mal definido, razão pela qual esta lógica também é conhecida pelo nome de lógica nebulosa. Segundo GUIMARÃES (2008), a lógica fuzzy pode ser definida como uma lógica que suporta os modos de raciocínio aproximado, como as pessoas estão naturalmente acostumadas a trabalhar. Seu princípio fundamental é o princípio da dualidade, que estabelece que dois eventos opostos podem coexistir.

SHAW; SIMOES (2007) afirmam que nas áreas onde é necessário lidar com a imprecisão, a subjetividade e o desconhecimento, essa teoria tem demonstrado grande capacidade de aplicação, ajudando profissionais a produzir modelos condizentes com as necessidades e realidades, mesmo com dados que apresentem alto grau de incerteza. A lógica fuzzy é baseada no princípio de que o pensamento humano se estrutura a partir de classes de objetos e não de números, logo se aplica corretamente na situação de transformar dados qualitativos em quantitativos e vice-versa.

Diferentemente da lógica clássica, onde os valores possíveis são o puramente verdadeiro e o puramente falso, na lógica *fuzzy* podem ser considerados vários graus de verdade. Ao optar trabalhar com a lógica *fuzzy* ao invés de trabalhar com a lógica clássica, amplia-se o contradomínio da função característica, que na lógica clássica é o conjunto $\{0, 1\}$, para o intervalo $[0, 1]$, contemplando os infinitos valores entre zero e um.

Segundo BARROS; BASSANEZI (2006), um subconjunto *fuzzy* F em um universo U é definido por uma função de pertinência, conforme a Eq. 63.

$$\mu F(x) : U \rightarrow [0, 1] \quad (63)$$

Algumas definições são importantes na lógica *fuzzy*: suporte, core, α -*corte* (α -nível) e números *fuzzy*.

O suporte de um subconjunto *fuzzy* F no conjunto universo U é o conjunto clássico que contém os elementos de U que têm grau de pertinência maior que zero, conforme Eq. 64.

$$suppF = \{x \in U : \mu F(x) > 0\} \quad (64)$$

O core de um conjunto *fuzzy* F consiste em todos os elementos que pertencem ao universo de discurso cujo grau de pertinência é igual a 1, conforme definido na Eq. 65.

$$Core(F) = \{x \in U / \mu F(x) = 1\} \quad (65)$$

Para um subconjunto *fuzzy* F de U e $\alpha \in [0, 1]$. O α -nível do conjunto F é o subconjunto clássico de U definido pela Eq. 66.

$$[F]^\alpha = \{x \in U : \mu F(x) \geq \alpha\} \text{ para } 0 < \alpha \leq 1 \quad (66)$$

De acordo com VIANA (2017), os números *fuzzy* são um tipo especial de conjuntos *fuzzy* e possuem uma relevância na resolução de problemas que envolvem imprecisões de dados. De maneira geral, em muito problemas, os números utilizados provém de informações imprecisas, envolvendo valores numéricos. Esses números contém uma carga de informações que precisam ser tratados de acordo com a lógica *fuzzy*, como os números *fuzzy*, que possuem uma relevância na resolução de problemas que envolvem uma imprecisão de dados. Um subconjunto *fuzzy* é chamado de número *fuzzy* quando o conjunto universo no qual $\mu F(x)$ está definida é o conjunto dos números reais e satisfaz às condições:

- i) todos os α -níveis de F são não vazios, com $0 \leq \alpha \leq 1$;
- ii) todos os α -níveis de F são intervalos fechados do conjunto dos números reais;
- iii) $\text{supp}F = \{x \in R : \mu F(x) > 0\}$ é limitado.

Usualmente, os números *fuzzy* mais utilizados são trapezoidais e triangulares, pois tornam mais simples o desempenho computacional. Segundo BRIAO (2013), os números *fuzzy* trapezoidais são mais gerais que os números *fuzzy* triangulares, sendo que todo número *fuzzy* triangular é um número *fuzzy* trapezoidal.

Um número *fuzzy* F é dito trapezoidal se sua função de pertinência é descrita conforme Eq. (67).

$$\mu F(x) = \begin{cases} \frac{x-a}{b-a}, & \text{se } a \leq x < b \\ 1, & \text{se } b \leq x \leq c \\ \frac{d-x}{d-c}, & \text{se } c < x \leq d \\ 0, & \text{caso contrário} \end{cases} \quad (67)$$

Esse número *fuzzy* tem a forma de um trapézio, conforme apresentado na Fig. 6. Na Eq. (67) a , b , c e d são a composição dos quatro vértices de coordenadas, $(a, 0)$, $(b, 1)$, $(c, 1)$ e $(d, 0)$, que formam o trapézio.

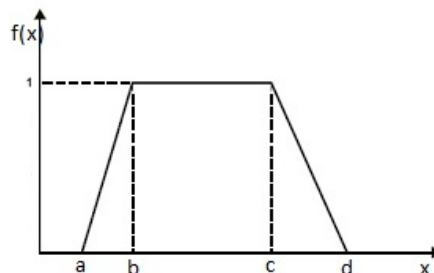


Figura 6: Número *fuzzy* trapezoidal: representação gráfica

Um número *fuzzy* F é dito triangular se sua função de pertinência é descrita conforme Eq. (68).

$$\mu F(X) = \begin{cases} \frac{x-a}{m-a}, & \text{se } a < x \leq m \\ \frac{b-x}{b-m}, & \text{se } m < x \leq b \\ 0, & \text{caso contrário} \end{cases} \quad (68)$$

Esse número *fuzzy* tem a forma de um triângulo, conforme apresentado na Fig. 7. Na Eq. (68), a , m e b são a composição dos três vértices de coordenadas, $(a, 0)$, $(m, 1)$ e $(b, 0)$, que formam o triângulo.

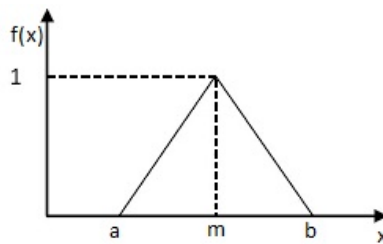


Figura 7: Número *fuzzy* triangular: representação gráfica

Ao realizar uma coleta de dados dificilmente uma informação é obtida como um número *fuzzy* (a, m, b) , razão pela qual métodos matemáticos que possam gerá-los são bastante importantes.

2.7 MCheng

A medida utilizada para determinar a tendência central de um conjunto de valores *crisp* para ser associada ao *core* de um número *fuzzy* triangular é utilizada pelo método citado e descrito em CHENG (2005), SILVESTRE (2014) e MATTOS et al. (2015).

O método sugere a geração de um número *fuzzy* (a, m, b) a partir de um conjunto de valores *crisp*.

Seja o conjunto com n valores discretos x_i , onde $i = 1, 2, \dots, n$. Inicialmente são calculadas as distâncias entre esses valores discretos de acordo com a Eq. (69).

$$d_{ij} = |x_i - x_j| \quad (69)$$

Estas são organizadas em uma matriz D , com i linhas e j colunas, onde $j = 0, 1, 2, \dots, n$ e suas respectivas proximidades médias ao centro dos valores, conforme Eq. (70).

$$\bar{d}_i = \frac{\sum_{j=1}^n d_{ij}}{n-1} \quad (70)$$

O grau de importância de cada x_i , com a comparação entre os x_i , com base em suas distâncias médias é fornecido pela Eq. (71), os valores obtidos são armazenados numa matriz P .

$$p_{ij} = \frac{\bar{d}_j}{d_i} \quad (71)$$

Os pesos w_j associado a cada x_i são calculados conforme Eq. (72).

$$w_j = \frac{1}{\sum_{i=1}^n p_{ij}} \quad (72)$$

Calcula-se o valor do centro m do número *fuzzy* de acordo com a Eq. (73).

$$m = \sum_{i=1}^n w_i \cdot x_i \quad (73)$$

Estima-se a variância, conforme a Eq (74).

$$\hat{S}^2 = \frac{1}{2} \left\{ \frac{\sum_i^n (m - x_i)^2}{\sum_{i(x_i < m)} w_i} + \frac{\sum_i^n (m - x_i)^2}{\sum_{i(x_i > m)} w_i} \right\} \quad (74)$$

Encontra-se ρ , razão da distribuição dos dados em torno de m , obtendo como a média ponderada dos valores menores que m e a média ponderada dos valores maiores que m , o que é feito, respectivamente, pelas Eqs. (75) e (76).

$$g^l = \frac{\sum_i^n w_i \cdot g_i}{w_i} \quad \text{para } x_i < m \quad (75)$$

$$g^r = \frac{\sum_i^n w_i \cdot g_i}{w_i} \quad \text{para } x_i > m \quad (76)$$

Estes valores são utilizados para encontrar a razão de distribuição dos dados da esquerda para a direita, o que é calculado pela Eq. (77).

$$\hat{\rho} = \frac{m - g^l}{g^r - m} \quad (77)$$

Finalmente, os limites a e b são encontrados pelas Eqs. (78) e (79).

$$a = m - \sqrt{\frac{12\rho^2 \cdot s^2}{1 + \rho^2}} \quad (78)$$

$$b = m + \sqrt{\frac{12s^2}{1 + \rho^2}} \quad (79)$$

Com os parâmetros, m , a e b , o número *fuzzy* triangular é definido conforme Eq. (80) onde, $0 \leq x \leq G$, sendo G o maior valor possível para gerar g . O número *fuzzy* triangular (μ) é denotado como (a, m, b) .

$$\mu(x; a, m, b) = \begin{cases} 1 - \frac{m-g}{m-a}, & a \leq g \leq m \\ 1 - \frac{g-m}{b-m}, & m < g \leq b \\ 0, & \text{outros valores} \end{cases} \quad (80)$$

3 METODOLOGIA

Primeiramente foi realizado um estudo sobre os assuntos abordados no trabalho. Esse estudo utilizou alguns autores além dos citados no trabalho dentre os quais: BARBETTA; REIS; BORNIA (2010), CASELLA; BERGER (2010), CHEN; PHAM (2000), DELGADO; VERDEGAY; VILA (1988), DEVORE (2006), KLIR; YUAN (1995), KOHAGURA (2007), MAGALHÃES; LIMA (2008), MATTOS; COELHO (2015), MONTGOMERY; RUNGER (2009), PORTAL ACTION. SOFTWARE DE ESTATÍSTICA ONLINE. (2017), ROSA (2012), ZADEH (1973), ZADEH (1965), ZADEH (1975a), ZADEH (1975b), ZADEH (1975c).

A análise comparativa das medidas média aritmética, mediana e mCheng inicia com a verificação para a medida *mCheng* de algumas propriedades matemáticas verificadas na média aritmética e mediana.

Em sequência, na análise comparativa das distribuições amostrais destas medidas, são realizados experimentos cujos resultados são gerados por simulação a partir da definição de alguns parâmetros que determinam algumas propriedades da população (assimetria e variabilidade) e da amostra (tamanho), que podem interferir em algumas propriedades das distribuições amostrais. Essas simulações foram realizadas no software SciLab e os programas desenvolvidos encontram-se em apêndice.

Para representar uma população com distribuição simétrica foi utilizada uma distribuição normal com média $\mu = 500$. Neste contexto foram geradas 1000 amostras para cada um dos seis experimentos descritos a seguir:

Experimento 1: desvio padrão pequeno ($\sigma = 10$) e amostra de tamanho muito pequeno ($n = 5$), conforme Tab. 1.

Tabela 1: Experimento 1: simulação

Amostra	Dados					Média	Mediana	mCheng
1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	\bar{X}_1	md_1	$mCheng_1$
2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	\bar{X}_2	md_2	$mCheng_2$
3	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	\bar{X}_3	md_3	$mCheng_3$
\vdots			\ddots			\vdots	\vdots	\vdots
1000	x_{10001}	x_{10002}	x_{10003}	x_{10004}	x_{10005}	\bar{X}_{1000}	md_{1000}	$mCheng_{1000}$

Experimento 2: desvio padrão grande ($\sigma = 150$) e amostra de tamanho muito pequeno ($n = 5$).

Experimento 3: desvio padrão pequeno ($\sigma = 10$) e amostra de tamanho pequeno ($n = 30$).

Experimento 4: desvio padrão grande ($\sigma = 150$) e amostra de tamanho pequeno ($n = 30$).

Experimento 5: desvio padrão pequeno ($\sigma = 10$) e amostra de tamanho grande ($n = 100$).

Experimento 6: desvio padrão grande ($\sigma = 150$) e amostra de tamanho grande ($n = 100$).

Para representar uma população com distribuição assimétrica foi utilizada uma distribuição exponencial. Neste contexto foram geradas 1000 amostras para cada um dos seis experimentos descritos a seguir.

Experimento 7: desvio padrão pequeno ($\sigma = 50$) e amostra de tamanho muito pequeno ($n = 5$).

Experimento 8: desvio padrão grande ($\sigma = 500$) e amostra de tamanho muito pequeno ($n = 5$).

Experimento 9: desvio padrão pequeno ($\sigma = 50$) e amostra de tamanho pequeno ($n = 30$).

Experimento 10: desvio padrão grande ($\sigma = 500$) e amostra de tamanho pequeno ($n = 30$).

Experimento 11: desvio padrão pequeno ($\sigma = 50$) e amostra de tamanho grande ($n = 100$).

Experimento 12: desvio padrão grande ($\sigma = 500$) e amostra de tamanho grande ($n = 100$).

Para caracterizar as distribuições amostrais das medidas estudadas, em cada experimento, foi realizada uma análise exploratória dos dados com o objetivo de identificar tendência central, variabilidade, assimetria, curtose, presença de valor fora do padrão e normalidade, o que foi feito por métodos analíticos e métodos gráficos. Foram calculadas as medidas descritivas usuais (mínimo, máximo, média, mediana, desvio padrão, primeiro e terceiro quartil, desvio interquartil, assimetria e curtose com os respectivos erro padrão), que permitem avaliar tendência central, variabilidade, assimetria e curtose e alguns gráficos (histograma, probabilidade normal e boxplot), que permitem avaliar assimetria, presença de lacunas e possíveis valores fora do padrão. Para avaliar a normalidade foram realizados testes de hipóteses Kolmogorov-Smirnov Lilliefors e Shapiro-Wilk, que avaliam a hipótese nula de que os dados provêm de uma distribuição normal.

Posteriormente, foi realizada uma análise com as três amostras pareadas, através de um teste F com blocos, que verifica a hipótese nula de que os dados provêm de distribuições com a mesma média, sendo verificadas suas suposições teóricas:

- 1 - independência;
- 2 - normalidade; e
- 3 - homogeneidade de variâncias.

A primeira suposição foi considerada como verdadeira, pelo fato dos dados terem sido gerados aleatoriamente por um simulador computacional. As suposições 2 e 3 foram analisadas em conjunto. A suposição 2 foi verificada pelos testes de hipóteses já mencionados (Kolmogorov-Smirnov Lilliefors e Shapiro-Wilk), enquanto a suposição 3 pelo teste Levene que avalia a hipótese nula de que os dados provém de distribuições com a mesma variância (homogeneidade). Quando a suposição 2 foi confirmada e a suposição 3 não, foi analisada a curtose dos resíduos. Por curtose entende-se a propriedade associada ao maior ou menor achatamento das distribuições dos dados, sendo denominada leptocúrtica quando a distribuição for menos achatada do que uma distribuição normal.

Se os resíduos apresentaram curtose não leptocúrtica, o teste F com blocos está validado. Se os resíduos apresentaram curtose leptocúrtica, foi utilizado o teste de Welch e, neste caso, também o teste não paramétrico de Friedman para confirmação dos resultados.

O teste de Welch verifica a mesma hipótese nula do teste F, no caso de não existência de homogeneidade entre as variâncias. Porém ao utilizar esse teste é perdida a informação dos blocos, razão pela qual também foi aplicado o teste de Friedman que é a alternativa não paramétrica para o teste F com blocos quando as suposições de normalidade ou homogeneidade entre variâncias não se verifica. Este teste, entretanto, faz a comparação das distribuições pela mediana.

Em vista disto, ao não se verificar a suposição 2, este mesmo teste foi aplicado. Em todas essas situações ao serem encontradas evidências de que os dados provém de distribuições com tendência central diferentes, foram utilizados testes de comparações múltiplas para identificar quais grupos são diferentes entre si. Para as comparações múltiplas foi utilizado o teste de postos de Wilcoxon, tendo em vista que essas evidências só foram encontradas em distribuições com evidências de não normalidade. Este teste é equivalente ao teste de Friedman, mas compara apenas duas distribuições.

Essa análise pode ser melhor visualizada no fluxograma da Fig. 8.

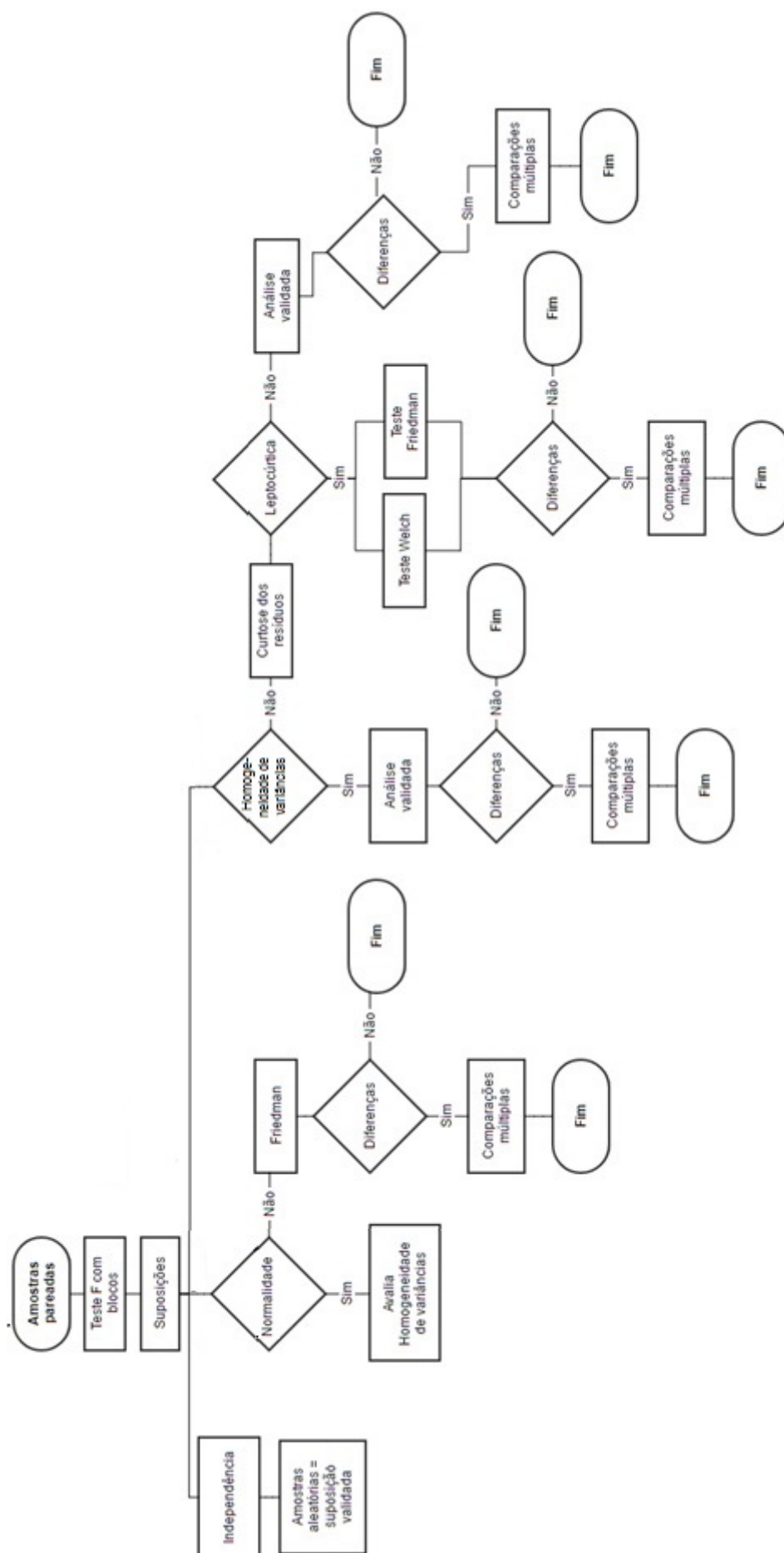


Figura 8: Fluxograma representando a análise realizada

4 RESULTADOS

No início do cálculo da medida *mCheng* foi constatada uma primeira limitação. Na etapa inicial de seu cálculo, ela encontra a diferença entre cada valor e todos os demais, conforme Eq. (69). Logo em seguida, estes valores são somados e divididos por $n - 1$, conforme Eq. (70). E esses resultados são utilizados para encontrar o valor de p_{ij} conforme Eq. (71). Quando o conjunto for formado por valores idênticos, a Eq. (71) leva a uma indeterminação matemática. Para o uso desta medida como *core* do número *fuzzy* triangular, isto não tem importância porque, se não existe variabilidade, a utilização da métrica *fuzzy* não é necessária.

4.1 Verificação de algumas propriedades da medida *mCheng*

Essa seção traz uma verificação das propriedades verificadas na média e mediana para a medida *mCheng*.

A propriedade 2.1 da média diz que a soma algébrica dos desvios tomados em relação esta medida é nula. Portanto,

$$\sum_{i=1}^n (x_i - \bar{X}) = 0 \quad (81)$$

Para a medida *mCheng*, tem-se:

$$\sum_{i=1}^n (x_i - mCheng) = 0 \quad (82)$$

Portanto:

$$\sum_{i=1}^n (x_i - \sum_{i=1}^n (w_i \cdot x_i)) = 0 \quad (83)$$

$$\sum_{i=1}^n (x_i) - \sum_{i=1}^n \sum_{i=1}^n (w_i \cdot x_i) = 0 \quad (84)$$

$$\sum_{i=1}^n (x_i) - n \cdot \sum_{i=1}^n (w_i \cdot x_i) = 0 \quad (85)$$

$$\sum_{i=1}^n (x_i) = n \cdot \sum_{i=1}^n (w_i \cdot x_i) \quad (86)$$

$$\frac{\sum_{i=1}^n (x_i)}{n} = \sum_{i=1}^n (w_i \cdot x_i) \quad (87)$$

$$\bar{X} = \sum_{i=1}^n (w_i \cdot x_i) \quad (88)$$

A Eq. 88 só se verifica se todos os w_i forem iguais a $\frac{1}{n}$, portanto, não se pode generalizar e a propriedade 2.1 não é válida para a medida *mCheng*.

A propriedade 2.2 da média, que é similar a propriedade 2.6 da mediana, diz que somando ou subtraindo uma constante de todos os valores de uma variável a média ou mediana do conjunto fica aumentada ou diminuída dessa constante.

Para o cálculo dos pesos w_i na medida mCheng, tem-se

$$d_{ij} = |x_i - x_j| \quad (89)$$

Somando uma constante tem-se:

$$d_{ij} = |x_i + c - (x_j + c)| \quad (90)$$

$$d_{ij} = |x_i + c - x_j - c| \quad (91)$$

$$d_{ij} = |x_i - x_j| \quad (92)$$

Como $|x_i - x_j|$ é o valor inicial de d_{ij} antes da soma da constante c a cada um dos elementos, o valor do peso w_i não é alterado pela adição de uma constante aos elementos do conjunto. Para a medida mCheng tem-se:

$$m = \sum_{i=1}^n w_i \cdot (x_i + c) \quad (93)$$

$$m = \sum_{i=1}^n (w_i \cdot x_i + w_i \cdot c) \quad (94)$$

$$m = \sum_{i=1}^n w_i \cdot x_i + \sum_{i=1}^n w_i \cdot c \quad (95)$$

$$m = \sum_{i=1}^n w_i \cdot x_i + c \sum_{i=1}^n w_i \quad (96)$$

$$\text{Como } \sum_{i=1}^n w_i = 1 \quad (97)$$

$$m = c + \sum_{i=1}^n w_i \cdot x_i \quad (98)$$

De acordo com o demonstrado, as propriedades 2.2 da média aritmética e 2.6 da mediana são válidas para a medida *mCheng*.

Da mesma forma, podemos verificar a propriedade 2.3 da média, similar à propriedade 2.7 da mediana.

Para o cálculo dos pesos na medida *mCheng*, tem-se:

$$d_{ij} = |x_i \cdot c - x_j \cdot c| \quad (99)$$

$$d_{ij} = |(x_i - x_j) \cdot c| \quad (100)$$

$$d_{ij} = c \cdot |x_i - x_j| \quad (101)$$

$$d_{ij} = c \cdot d_{ij} \quad (102)$$

$$\bar{d}_i = \frac{\sum_{j=1}^n c \cdot d_{ij}}{n - 1} \quad (103)$$

$$\bar{d}_i = c \cdot \frac{\sum_{j=1}^n d_{ij}}{n - 1} \quad (104)$$

$$\bar{d}_i = c \cdot \bar{d}_i \quad (105)$$

$$p_{ij} = \frac{c \cdot \bar{d}_j}{c \cdot \bar{d}_i} \quad (106)$$

$$p_{ij} = \frac{\bar{d}_j}{\bar{d}_i} \quad (107)$$

Como $\frac{\bar{d}_j}{\bar{d}_i}$ é o valor inicial de p_{ij} antes da multiplicação pela constante c a cada um

dos elementos, o valor do peso w_i não é alterado pela multiplicação de uma constante aos elementos do conjunto. Para a medida *mCheng* tem-se:

$$m = \sum_{i=1}^n w_i \cdot x_i \cdot c \quad (108)$$

$$m = c \cdot \sum_{i=1}^n w_i \cdot x_i \quad (109)$$

De acordo com o demonstrado, as propriedades 2.3 da média aritmética e 2.7 da mediana são válidas para a medida *mCheng*.

A propriedade 2.4 da média, que diz que a soma dos quadrados dos afastamentos dos valores em relação a essa medida é um ponto de mínimo, não se verifica para a medida *mCheng*, conforme demonstrado a seguir.

Seja:

$$S = \sum_{i=1}^n (x_i - a)^2 \quad (110)$$

Para encontrar um ponto de mínimo, tem-se:

$$\frac{dS}{da} = 2 \sum_{i=1}^n (x_i - a)(-1) \quad (111)$$

$$2 \left(\sum_{i=1}^n (x_i) - n \cdot a \right) \cdot (-1) = 0 \quad (112)$$

$$\sum_{i=1}^n = n \cdot a \quad (113)$$

$$a = \frac{\sum_{i=1}^n x_i}{n} = \bar{X} \quad (114)$$

Como $\bar{X} \neq mCheng$ a propriedade 2.4 não se verifica para a medida *mCheng*, a não ser que os mesmos pesos sejam considerados para todos os valores no somatório dos afastamentos.

A propriedade 2.5, que diz que a mediana não é influenciada pelos valores extremos, não se verifica para a medida *mCheng*, pois cada valor do conjunto é considerado no cálculo embora com diferentes pesos.

As propriedades que foram verificadas e que não foram verificadas para a medida *mCheng* podem ser melhor visualizadas na Tab. 2.

Através da Tab. 2 pode-se evidenciar que a medida *mCheng* possui as propriedades 2.2 e 2.3 da média aritmética e 2.6 e 2.7 da mediana. A verificação de tais propriedades na

Tabela 2: Propriedades verificadas e não verificadas para a medida $mCheng$

Propriedades		Verificada	Não verificada
2.1	A soma algébrica dos desvios tomados em relação a $mCheng$ é nula.		X
2.2	Somando ou subtraindo uma constante (c) de todos os valores de uma variável, a $mCheng$ do conjunto fica aumentada ou diminuída dessa constante.	X	
2.3	Multiplicando ou dividindo todos os valores de uma variável por uma constante (c), a $mCheng$ do conjunto fica multiplicada ou dividida por essa constante.	X	
2.4	A soma dos quadrados dos afastamentos contados a partir da $mCheng$ é um mínimo.		X
2.5	A $mCheng$ não é influenciada pelos valores extremos. Somando ou subtraindo uma constante (c) de todos os valores de uma variável, a $mCheng$ do conjunto fica aumentada ou diminuída dessa constante.		X
2.6	Multiplicando ou dividindo todos os valores de uma variável por uma constante (c), a $mCheng$ do conjunto fica multiplicada ou dividida por essa constante.	X	
2.7	Multiplicando ou dividindo todos os valores de uma variável por uma constante (c), a $mCheng$ do conjunto fica multiplicada ou dividida por essa constante.	X	

medida $mCheng$ traz uma confiabilidade maior a medida como estimador da tendência central de conjuntos de dados.

4.2 Análise dos resultados dos experimentos

4.2.1 Visão geral

Em relação a tendência central (média) e variabilidade (desvio padrão), os achados para os seis experimentos com dados provenientes de uma distribuição normal podem ser visualizados nas Figs. 9 e 10, respectivamente.

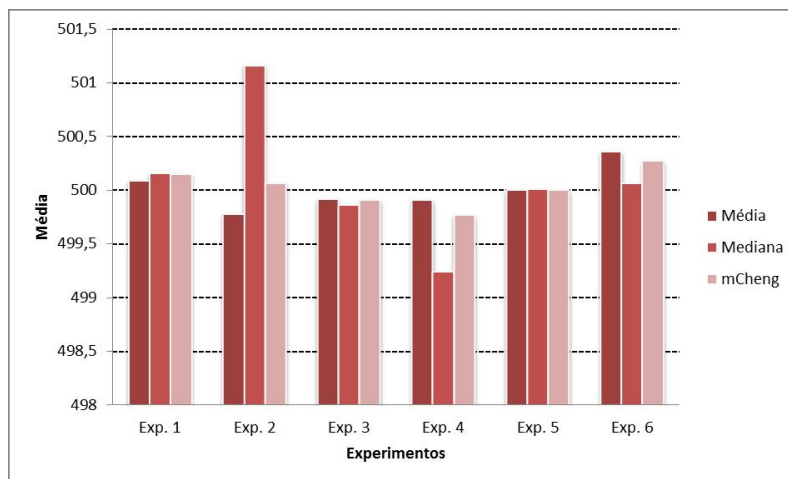


Figura 9: Média: experimentos nº 1 à nº 6

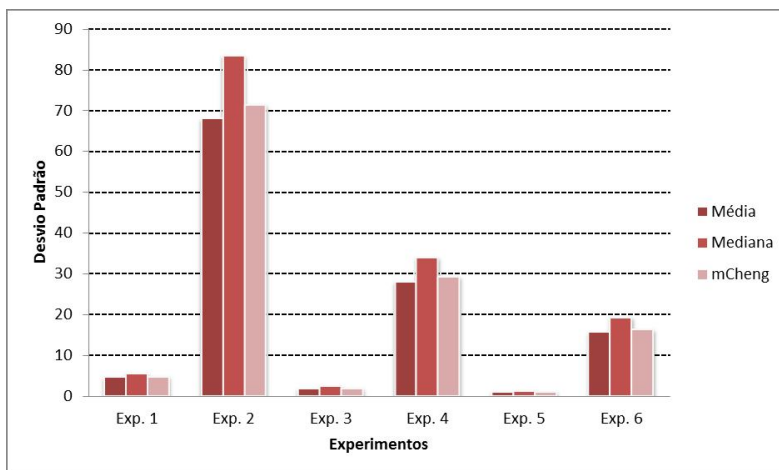


Figura 10: Desvio padrão: experimentos nº 1 à nº 6

Na Fig. 9 constata-se que nos experimentos 1, 3 e 5 realizados com dados provenientes de uma distribuição normal com desvio padrão pequeno parece existir menos diferença entre as medidas de tendência central do que entre as medidas de tendência central obtidas nos experimentos 2, 4 e 6 que usaram dados provenientes também de uma distribuição normal mas com desvio padrão grande. Neste mesmo gráfico constata-se também que o resultado da média da distribuição amostral da medida *mCheng* fica sempre entre este mesmo resultado das distribuições amostrais da média e da mediana.

A Fig. 10 evidencia que estas mesmas propriedades são válidas para a dispersão das três distribuições amostrais estudadas. Salienta-se que a menor variabilidade ocorreu sempre na distribuição amostral da média. Também fica evidenciado que esta variabilidade é inversamente proporcional ao tamanho das amostras, sendo mais evidente quando os dados são provenientes de distribuição normal com desvio padrão grande.

Os achados para os seis experimentos com dados provenientes de uma distribuição exponencial podem ser visualizados nas Figs. 11 e 12.

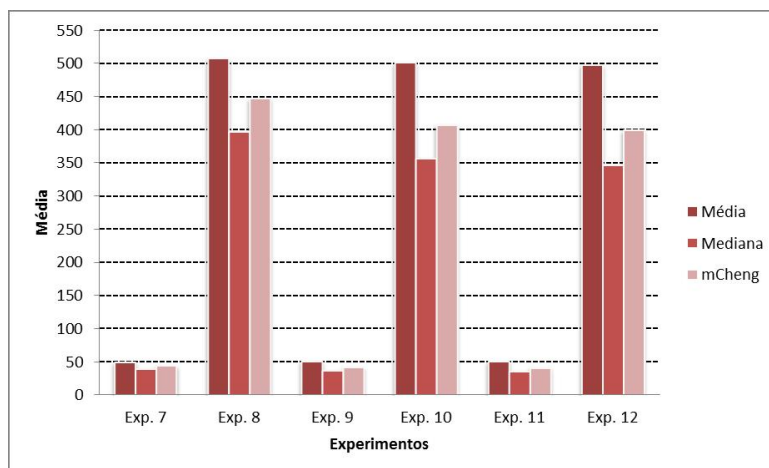


Figura 11: Média: experimentos nº 7 à nº 12

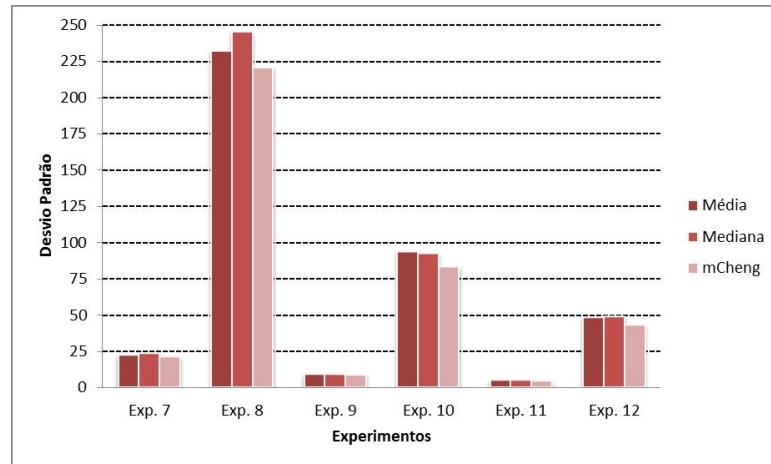


Figura 12: Desvio padrão: experimentos n° 7 à n° 12

Na Fig. 11, constata-se que as propriedades encontradas para a tendência central das distribuições amostrais obtidas para dados provenientes de uma distribuição exponencial são as mesmas encontradas para dados provenientes de uma distribuição normal (Fig. 9). O mesmo acontecendo em relação a variabilidade conforme mostrado na Fig. 12. A única diferença reside no fato de que a menor variabilidade ocorreu nas distribuições amostrais da medida *mCheng*.

4.2.2 Distribuição Normal

4.2.2.1 Experimento n° 1 e Experimento n° 2

No experimento n° 1 $X_1 \sim N(500, 100)$ e no experimento n° 2 $X_2 \sim N(500, 22500)$. Em ambos os experimentos $n = 5$. Os resultados da análise estatística nestes experimentos estão apresentados na Tab. 3.

No experimento n° 1, a distribuição amostral das médias apresentou uma levíssima assimetria negativa, não significativa e uma levíssima tendência a curtose leptocúrtica, também não significativa, compatíveis com uma distribuição normal. As médias amostrais variaram entre 485,161 e 514,738, concentrando-se em torno da média 500,088 com desvio padrão de 4,612. A distribuição amostral da mediana apresentou propriedades similares em relação a assimetria e curtose, podendo ser considerada compatível com uma distribuição normal. Sua média (500,153) ficou um pouco mais afastada do valor esperado (500), e sua variabilidade foi também um pouco maior, pois os resultados das medianas amostrais variaram entre 481,216 e 515,042, apresentando desvio padrão de 5,485. Observa-se, entretanto, que a mediana das medianas amostrais está um pouco mais próxima da média desta medida do que as respectivas medidas das médias amostrais. A distribuição amostral da medida *mCheng* apresenta comportamento similar em relação à normalidade. Seus resultados variaram entre 483,772 e 513,403, concentrando-se em torno da média 500,146 e com desvio padrão de 4,775.

Tabela 3: Análise estatística: experimento n° 1 e experimento n° 2

Medidas		Experimento n° 1			Experimento n° 2		
		Média	Mediana	mCheng	Média	Mediana	mCheng
Mínimo		485,161	481,216	483,722	252,774	275,673	258,923
Máximo		514,738	515,042	513,403	733,868	758,911	741,696
Média		500,088	500,153	500,146	499,780	501,159	500,065
Mediana		500,119	500,140	500,046	501,986	503,072	500,569
Desvio Padrão		4,612	5,485	4,775	68,012	83,503	71,476
Q1		497,073	496,508	496,910	452,721	443,955	452,284
Q3		503,109	503,914	503,342	547,089	558,354	549,243
Desvio interquartílico		6,036	7,406	6,432	94,368	114,399	96,959
Assimetria		-0,013	-0,101	-0,041	-0,051	-0,001	-0,015
Curtose		0,053	0,022	0,008	-0,104	-0,152	-0,116
Kolmogorov-Smirnov Lilliefors	Dn	0,021	0,019	0,018	0,021	0,017	0,019
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,200	0,200	0,200	0,200	0,200	0,200
Shapiro-Wilk	W	0,999	0,998	0,999	0,999	0,998	0,999
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,684	0,468	0,822	0,659	0,487	0,868
Levene	F*	16,699			19,718		
	gl1	2			2		
	gl2	2997			2997		
	valor p	<0,001			<0,001		
Welch	Fc	0,055			0,086		
	gl1	2			2		
	gl2	1987,917			1984,405		
	valor p	0,946			0,917		

Obs.: Erro padrão da assimetria: 0,077; Erro padrão da curtose: 0,155; $n = 5$

Portanto, as medidas descritivas apresentadas na Tab. 3 mostram que as propriedades das distribuição amostrais analisadas são bastante similares. Ainda, as propriedades da distribuição amostral (tendência central, dispersão, assimetria e curtose) da medida mCheng parecem ser mais próximas das propriedades da distribuição amostral da média do que as propriedades da distribuição amostral da mediana. Salienta-se que a mediana da distribuição amostral mCheng foi a medida mais próxima da média da população seguida da média da distribuição amostral das médias.

Ainda, de acordo com a Tab. 3, as distribuições amostrais das três medidas estudadas no experimento n° 2, apresentaram propriedades similares as distribuições encontradas no experimento n° 1, diferenciando-se em relação:

- curtose - as distribuições caracterizaram-se como levemente platicúrticas, embora não significativas;
- a tendência central da medida mCheng representada pela média e pela mediana ficou mais próxima da média da população do que as mesmas medidas das

distribuições amostrais da média e da mediana;

- a média da distribuição amostral da medida mCheng foi a medida de tendência central mais próxima da média da população, seguida da média das distribuições amostrais das médias.

O experimento n° 2, se diferencia do experimento n° 1 em relação à uma variabilidade maior.

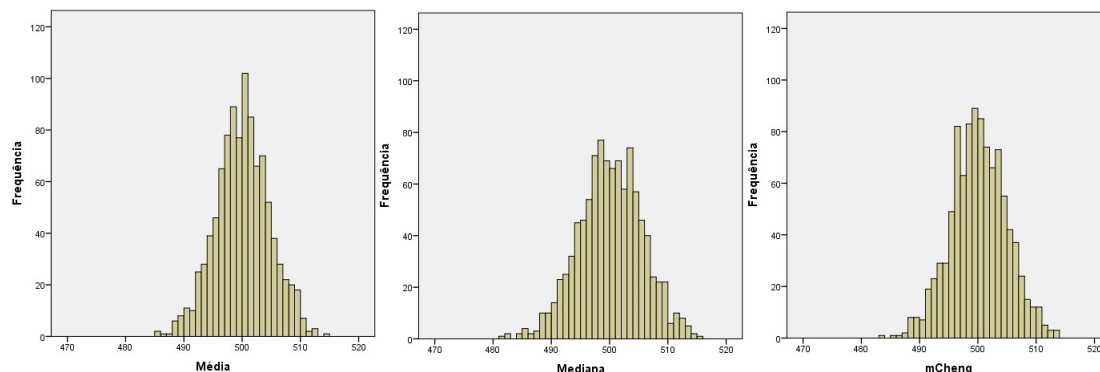


Figura 13: Histogramas: experimento n° 1

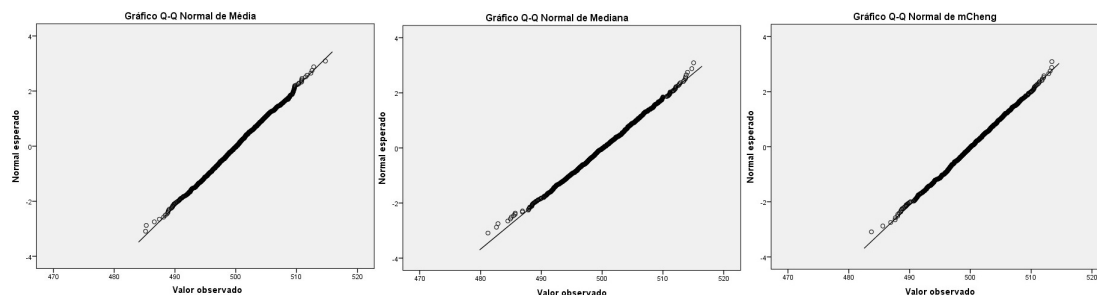


Figura 14: Gráficos de probabilidade normal: experimento n° 1

Analisando os histogramas apresentados na Fig. 13, percebe-se que as distribuições das três medidas tendem a uma distribuição normal, o mesmo acontecendo em relação aos gráficos de probabilidade normal, apresentados na Fig. 14. Também não se evidenciam lacunas, nem possíveis valores fora do padrão. Ainda, nestes últimos gráficos percebe-se que os resultados estão um pouco mais afastados da reta nas extremidades. Este afastamento parece ser maior para valores da mediana do que para valores da média e da mCheng.

Os boxplots das três medidas, apresentados na Fig. 15, são similares, sugerindo simetria e evidenciando que a maior variabilidade ocorre na distribuição amostral das medianas, enquanto a menor, na distribuição amostral das médias, propriedade esperada de um bom estimador. Por estes gráficos, todas as distribuições amostrais apresentam alguns possíveis *outliers*, tanto superiores como inferiores. Observa-se ainda que a mediana amostral possui o menor possível outlier inferior e o maior possível outlier superior.

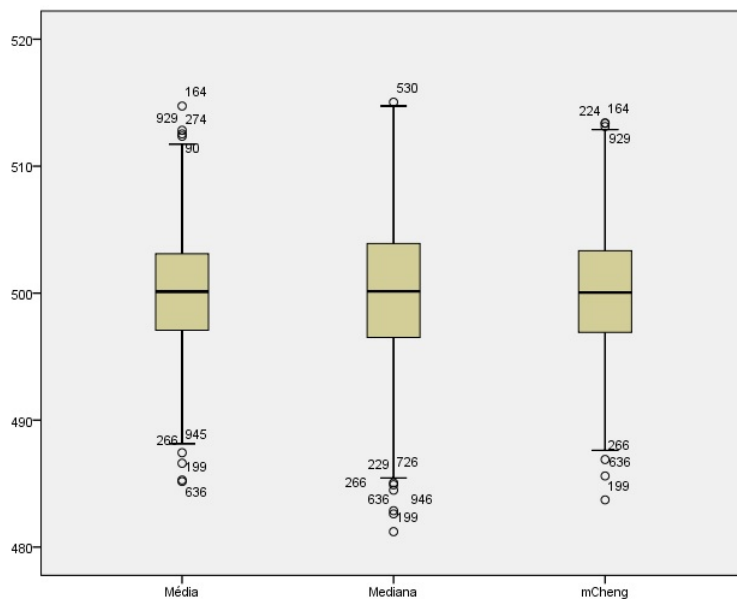


Figura 15: Boxplots: experimento n° 1

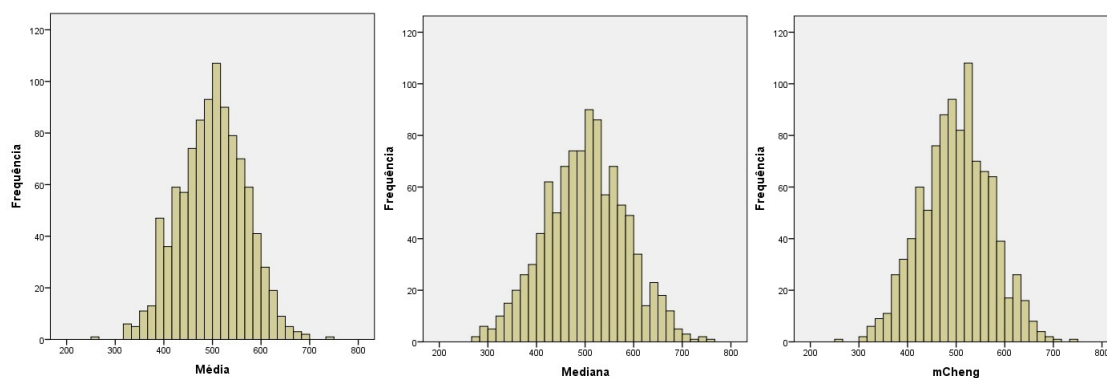


Figura 16: Histogramas: experimento n° 2

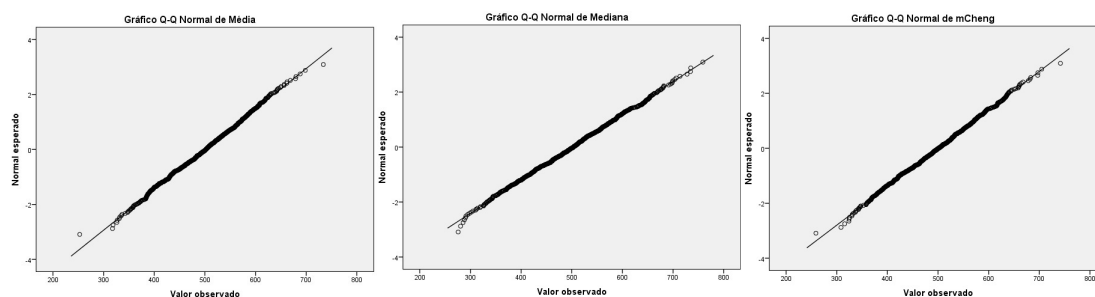


Figura 17: Gráficos de probabilidade normal: experimento n° 2

Os histogramas apresentados na Fig. 16, os gráficos de probabilidade normal apresentados na Fig. 17 e os boxplots apresentados na Fig. 18 foram similares aos encontrados no experimento n° 1. Apenas, nota-se que a condição de possíveis outliers inferiores diminuíram, além disso nos histogramas das medidas média e mCheng foram encontradas lacunas.

No experimento n° 1 a análise dos testes de normalidade foi realizada pelos testes de

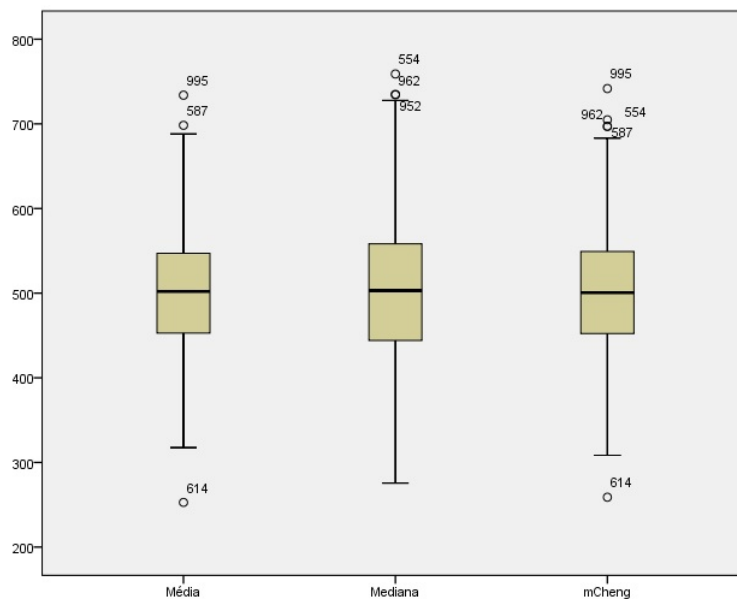


Figura 18: Boxplots: experimento n° 2

Kolmogorov-Smirnov Lilliefors e Shapiro-Wilk, que verificam a hipótese nula de que os dados seguem uma distribuição normal, obtendo-se os resultados apresentados na Tab. 3. Consta-se que, em ambos os testes, não foram encontradas evidências de que os dados não seguem uma distribuição normal, concordando com a análise gráfica.

O experimento n° 2 apresentou resultados similares ao experimento n° 1 para os testes de normalidade, apresentados na Tab. 3, ou seja não foram encontradas evidências de não normalidade nas três medidas.

O teste Levene, aplicado aos dados do experimento n° 1 encontrou os resultados apresentados na Tab. 3, encontrando evidências de que existe heterogeneidade entre as variâncias. A distribuição amostral das médias apresentou menor variabilidade, enquanto a distribuição amostral das medianas apresentou a maior variabilidade. Evidencia-se aqui a proximidade da variabilidade da distribuição amostral da mCheng à variabilidade amostral das médias.

Semelhante ao experimento n° 1, o teste Levene aplicado aos dados do experimento n° 2, apresentado na Tab. 3, encontrando evidências de que existe heterogeneidade entre as variâncias.

O teste Welch, aplicado aos dados do experimento n° 1 encontrou os resultados apresentados na Tab. 3, não encontrando evidências de diferença entre as médias das distribuições amostrais das três medidas. Chama-se a atenção para o fato, entretanto, de que a média da distribuição amostral das médias está mais próxima da média da distribuição amostral da mCheng do que da média da distribuição amostral das medianas. O teste não paramétrico de Friedman ($\chi^2 = 3,962$; $gl = 2$; $valor\ p = 0,138$) apresentou o mesmo resultado que o teste de Welch, portanto, compreende-se que não há diferença na centralidade das distribuições amostrais das três medidas.

Similar ao Experimento nº 1, o teste Welch aplicado aos dados do experimento nº 2, apresenta os resultados na Tab. 3, não encontrando evidências de diferença entre as três médias das distribuições amostrais. O teste não paramétrico de Friedman ($\chi^2 = 1,554$; $gl = 2$; $valor p = 0,460$) apresentou o mesmo resultado que o teste de Welch, portanto, compreende-se que não há diferença na centralidade das distribuições amostrais das três medidas.

4.2.2.2 Experimento nº 3 e Experimento nº 4

No experimento nº 3 $X_3 \sim N(500, 100)$ e no experimento nº 4 $X_4 \sim N(500, 22500)$, em ambos os experimentos $n = 30$. Os resultados da análise estatística realizada para os experimentos nº 3 e nº 4 estão apresentados na Tab. 4.

Tabela 4: Análise estatística: experimento nº 3 e experimento nº 4

Medidas		Experimento nº 3			Experimento nº 4		
		Média	Mediana	mCheng	Média	Mediana	mCheng
Mínimo		494,715	492,702	494,495	408,057	392,148	399,714
Máximo		505,647	507,248	505,689	599,745	631,961	609,659
Média		499,914	499,859	499,907	499,907	499,241	499,767
Mediana		500,033	499,855	499,987	500,145	500,427	500,334
Desvio Padrão		1,819	2,368	1,918	28,014	34,016	29,315
Q1		498,665	498,243	498,625	481,218	475,391	479,444
Q3		501,124	501,427	501,172	519,129	523,602	520,602
Desvio interquartílico		2,464	3,184	2,547	37,911	48,211	41,158
Assimetria		-0,073	-0,060	-0,086	-0,043	0,013	-0,049
Curtose		-0,213	-0,089	-0,154	0,142	-0,050	0,134
Kolmogorov-	Dn	0,028	0,027	0,021	0,017	0,022	0,025
Smirnov	gl	1000	1000	1000	1000	1000	1000
Lilliefors	valor p	0,058	0,077	0,200	0,200	0,200	0,131
Shapiro-Wilk	W	0,998	0,999	0,998	0,998	0,998	0,998
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,161	0,557	0,268	0,516	0,516	0,165
Levene	F*	32,701			21,425		
	gl1	2			2		
	gl2	2997			2997		
	valor p	<0,001			<0,001		
Welch	Fc	0,253			0,120		
	gl1	2			2		
	gl2	1976,183			1985,708		
	valor p	0,776			0,887		

Obs.: Erro padrão da assimetria: 0,077; Erro padrão da curtose: 0,155; $n = 30$

De acordo com a Tab. 4, as distribuições amostrais das três medidas estudadas no experimento nº 3 apresentaram propriedades similares as distribuições encontradas no experimento nº 1, diferenciando-se em relação:

- curtose - as distribuições caracterizaram-se como levemente platicúrticas, embora

não significativas.

Ainda, de acordo com a Tab. 4, as distribuições amostrais das três medidas estudadas no experimento nº 4 apresentaram propriedades similares as distribuições encontradas no experimento nº 1, diferenciando-se em relação:

- assimetria - a distribuição amostral da mediana apresentou uma leve assimetria positiva;
- curtose - a mediana apresentou uma leve tendência platicúrtica.
- a média da distribuição amostral da média foi a medida de tendência central mais próxima da média da população, seguida da mediana dessa mesma distribuição amostral.

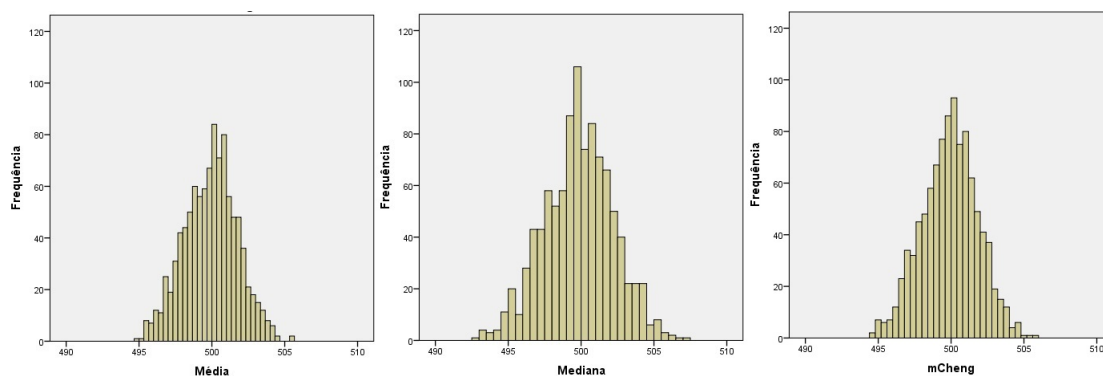


Figura 19: Histogramas: experimento nº 3

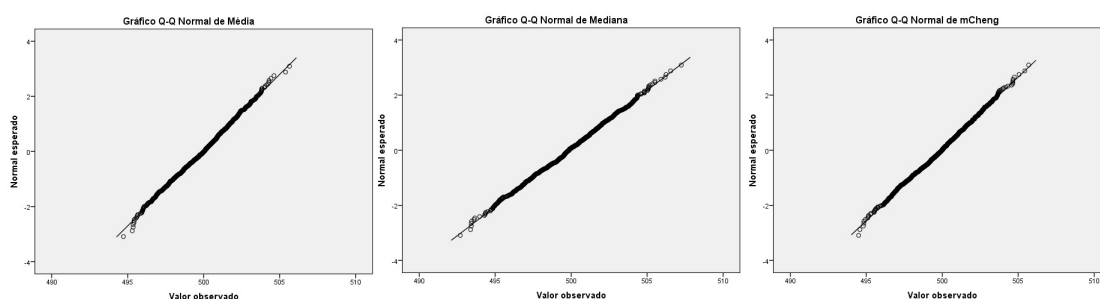


Figura 20: Gráficos de probabilidade normal: experimento nº 3

Os histogramas apresentados na Fig. 19, os gráficos de probabilidade normal apresentados na Fig. 20 e os boxplots apresentados na Fig. 21 foram similares aos encontrados no experimento nº 1. Apenas, no histograma representativo da distribuição amostral das médias apresentou uma lacuna.

Os histogramas apresentados na Fig. 22, os gráficos de probabilidade normal apresentados na Fig. 23 e os boxplots apresentados na Fig. 24 foram similares aos encontrados

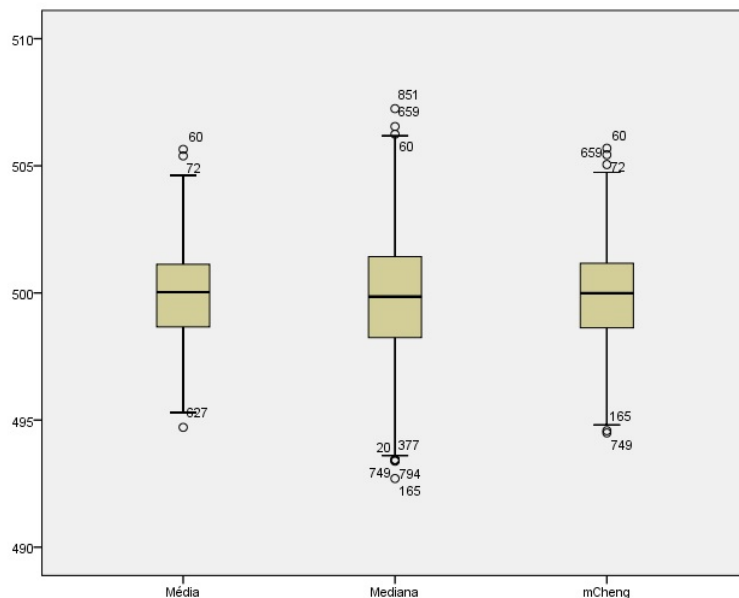


Figura 21: Boxplots: experimento n° 3

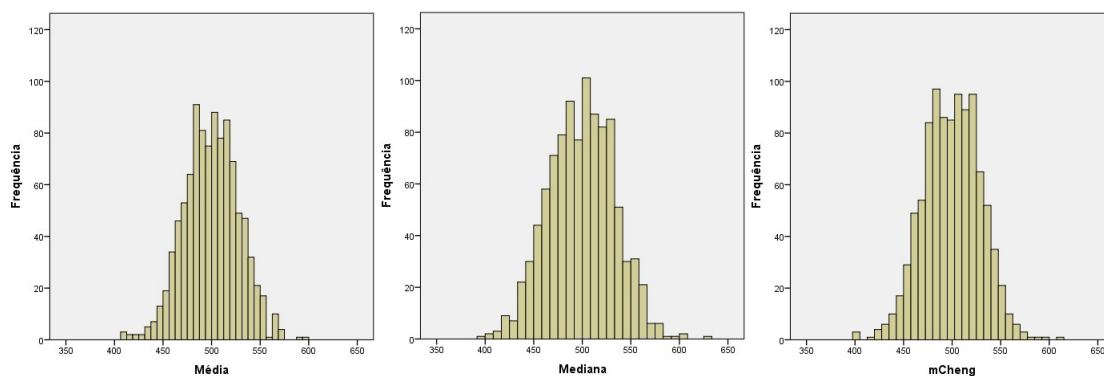


Figura 22: Histogramas: experimento n° 4

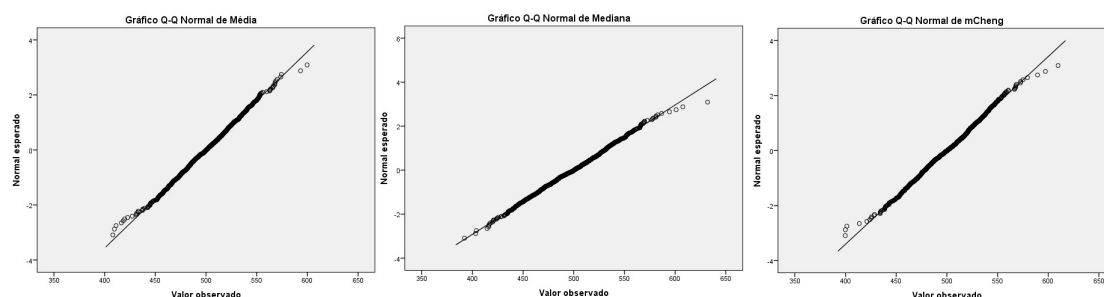


Figura 23: Gráficos de probabilidade normal: experimento n° 4

no experimento n° 1. Apenas, nos histogramas representativos das três medidas foram encontradas lacunas superiores, a mCheng apresenta uma lacuna inferior.

Similar ao experimento n° 1, os resultados dos testes de normalidade dos experimentos n° 3 e n° 4, apresentados na Tab. 4, foram similares, ou seja não foram encontradas evidências de não normalidade nas três medidas.

Semelhante ao experimento n° 1, o teste Levene aplicado aos dados dos experimentos

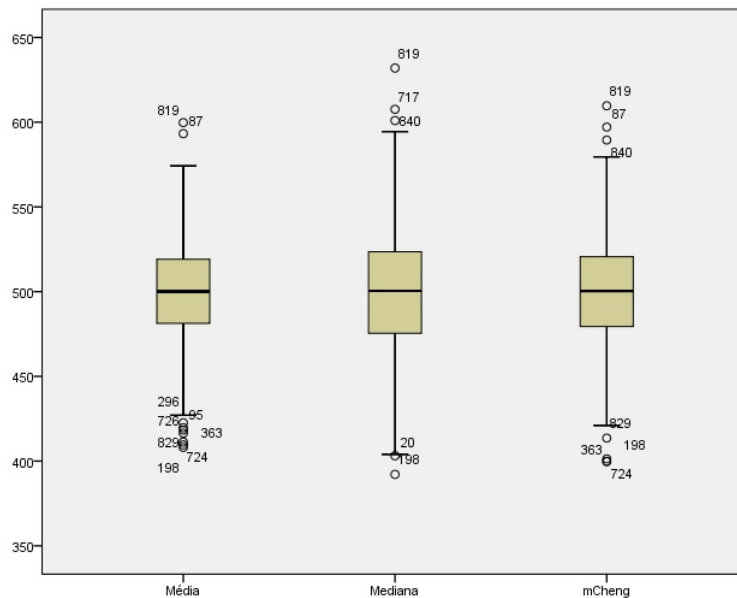


Figura 24: Boxplots: experimento n° 4

n° 3 e n° 4, apresentado na Tab. 4, encontrando evidências de que existe heterogeneidade entre as variâncias.

Similar ao Experimento n° 1, o teste Welch aplicado aos dados do experimento n° 3, apresenta os resultados na Tab. 4, não encontrando evidências de diferença entre as três médias das distribuições amostrais. O teste não paramétrico de Friedman ($\chi^2 = 3,614$; $gl = 2$; $valor\ p = 0,164$) apresentou o mesmo resultado que o teste de Welch, portanto, compreende-se que não há diferença na centralidade das distribuições amostrais das três medidas.

Similar ao Experimento n° 1, o teste Welch aplicado aos dados do experimento n° 4, apresenta os resultados na Tab. 4, não encontrando evidências de diferença entre as três médias das distribuições amostrais. O teste não paramétrico de Friedman ($\chi^2 = 1,568$; $gl = 2$; $valor\ p = 0,457$) apresentou o mesmo resultado que o teste de Welch, portanto, compreende-se que não há diferença na centralidade das distribuições amostrais das três medidas.

4.2.2.3 Experimento n° 5 e Experimento n° 6

No experimento n° 5 $X_5 \sim N(500, 100)$ e no experimento n° 6 $X_6 \sim N(500, 22500)$, em ambos os experimentos $n = 100$. Os resultados da análise estatística realizada para os experimentos n° 5 e n° 6 estão apresentados na Tab. 5.

De acordo com a Tab. 5, as distribuições amostrais das três medidas estudadas no experimento n° 5 apresentaram propriedades similares as distribuições encontradas no experimento n° 1, diferenciando-se em relação:

- assimetria - as distribuições apresentam assimetria positiva, embora não significati-

Tabela 5: Análise estatística: experimento nº 5 e experimento nº 6

Medidas		Experimento nº 5			Experimento nº 6		
		Média	Mediana	mCheng	Média	Mediana	mCheng
Mínimo		497,289	496,349	496,926	450,653	443,246	451,248
Máximo		503,394	503,998	503,496	553,718	556,356	544,195
Média		500,001	500,008	499,999	500,355	500,064	500,271
Mediana		500,028	499,985	499,986	500,031	499,849	500,055
Desvio Padrão		0,992	1,252	1,044	15,642	19,288	16,340
Q1		499,331	499,131	499,274	489,481	487,556	489,568
Q3		500,664	500,836	500,689	510,934	512,746	511,114
Desvio interquartílico		1,333	1,705	1,415	21,453	25,190	21,546
Assimetria		0,009	0,016	0,081	-0,022	0,007	-0,038
Curtose		-0,077	-0,100	-0,100	-0,121	-0,077	-0,175
Kolmogorov-Smirnov	Dn	0,014	0,018	0,015	0,014	0,018	0,016
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,200	0,200	0,200	0,200	0,200	0,200
Shapiro-Wilk	W	0,999	0,999	0,999	0,999	0,998	0,998
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,653	0,845	0,699	0,856	0,387	0,410
Levene	F*	27,443			18,822		
	gl1	2			2		
	gl2	2997			2997		
	valor p	<0,001			<0,001		
Welch	Fc	0,015			0,070		
	gl1	2			2		
	gl2	1980,774			1983,698		
	valor p	0,985			0,932		

Obs.: Erro padrão da assimetria: 0,077; Erro padrão da curtose: 0,155; $n = 100$

vas;

- curtose - as distribuições caracterizaram-se como levemente platicúrticas, embora não significativas;
- a média da distribuição amostral das médias e da medida mCheng foram as mais próximas da média da população, seguidas da média das distribuições amostrais da mediana.

De acordo com a Tab. 5, as distribuições amostrais das três medidas estudadas no experimento nº 6 apresentaram propriedades similares as distribuições encontradas no experimento nº 1, diferenciando-se em relação:

- assimetria - a distribuição amostral da mediana apresentou uma leve assimetria positiva;
- curtose - as distribuições caracterizaram-se como levemente platicúrticas, embora não significativas;

- a mediana da distribuição amostral das médias foi a medida de tendência central mais próxima da média da população, seguida da mediana da medida mCheng.

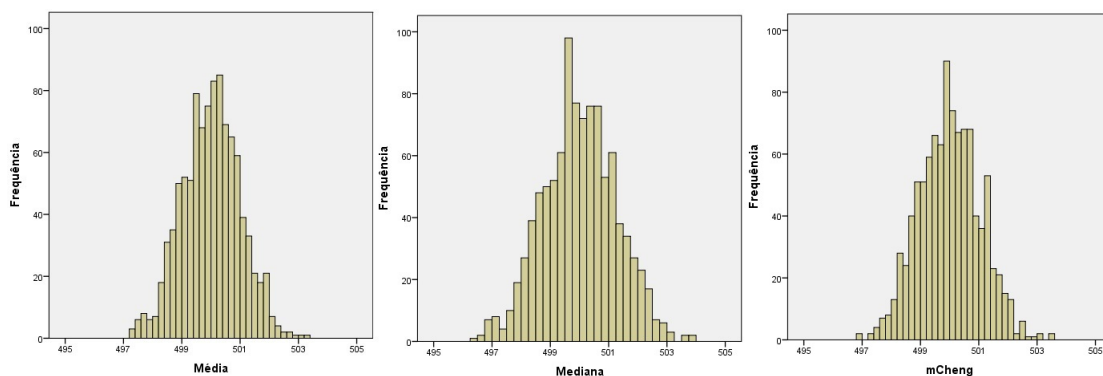


Figura 25: Histogramas: experimento nº 5

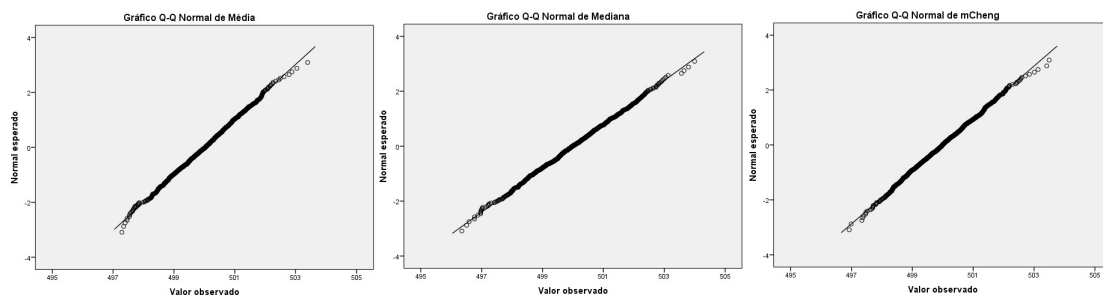


Figura 26: Gráficos de probabilidade normal: experimento nº 5

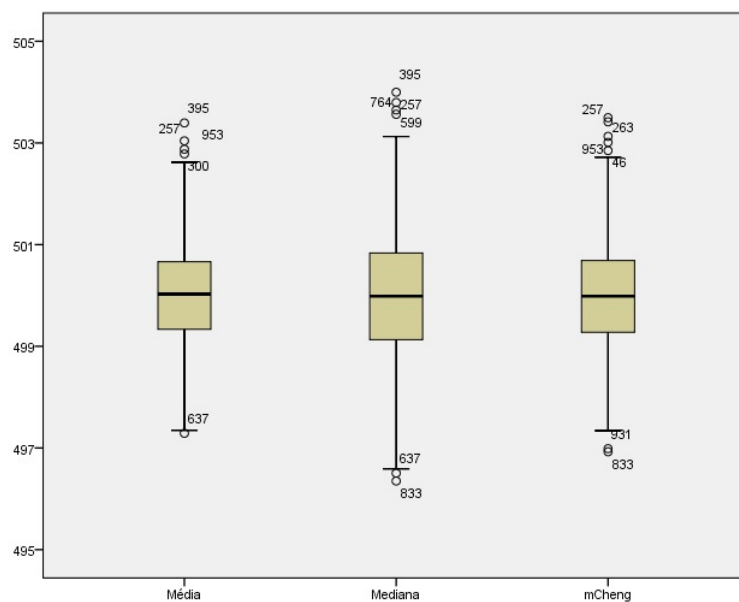


Figura 27: Boxplots: experimento nº 5

Os histogramas apresentados na Fig. 25, os gráficos de probabilidade normal apresentados na Fig. 26 e os boxplots apresentados na Fig. 27 foram similares aos encontrados

no experimento n^o 1. Apenas, nota-se nos histogramas, lacunas superiores para a mediana e a mCheng, e lacuna inferior na mCheng.

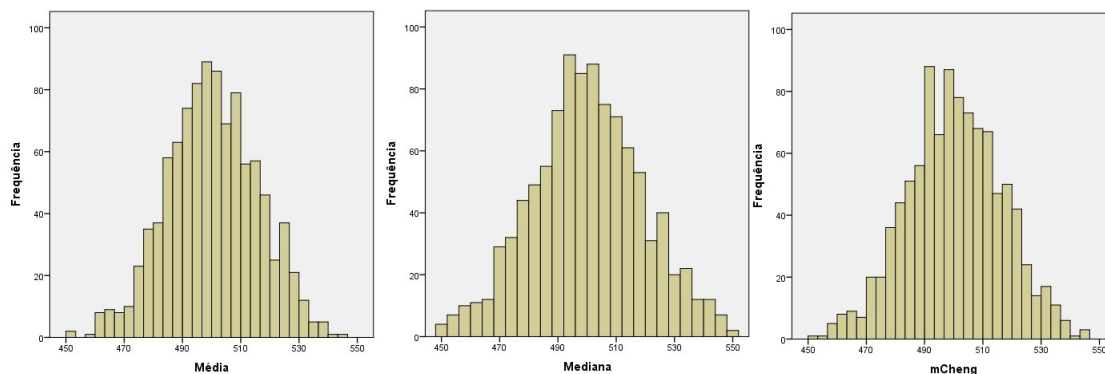


Figura 28: Histogramas: experimento n^o 6

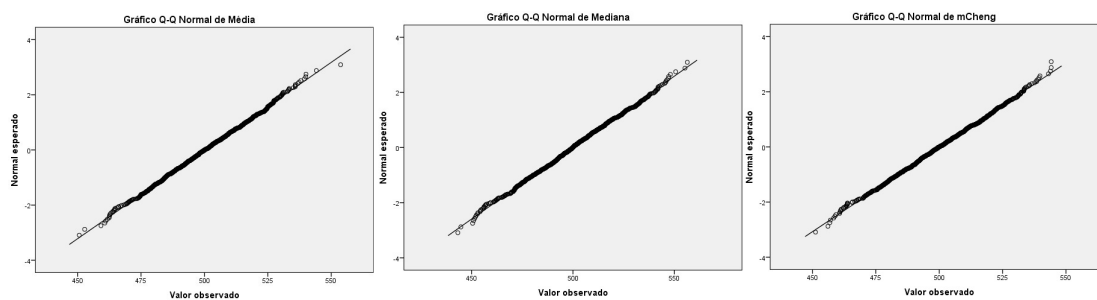


Figura 29: Gráficos de probabilidade normal: experimento n^o 6

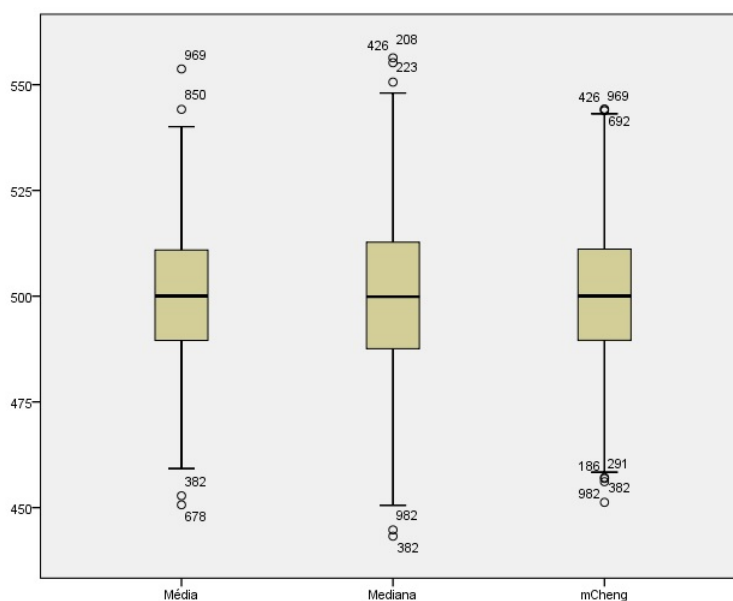


Figura 30: Boxplots: experimento n^o 6

Os histogramas apresentados na Fig. 28, os gráficos de probabilidade normal apresentados na Fig. 29 e os boxplots apresentados na Fig. 30 foram similares aos encontrados

no experimento nº 1. Apenas, nota-se que o histograma da média apresenta uma lacuna inferior.

Similar ao experimento nº 1, os resultados dos testes de normalidade dos experimentos nº 5 e nº 6, apresentados na Tab. 5, foram similares, ou seja não foram encontradas evidências de não normalidade nas três medidas.

Semelhante ao experimento nº 1, o teste Levene aplicado aos dados dos experimentos nº 5 e nº 6, apresentado na Tab. 5, encontrando evidências de que existe heterogeneidade entre as variâncias.

Similar ao Experimento nº 1, o teste Welch aplicado aos dados do experimento nº 5, apresenta os resultados na Tab. 5, não encontrando evidências de diferença entre as três médias das distribuições amostrais. O teste não paramétrico de Friedman ($\chi^2 = 1,950$; $gl = 2$; $valor p = 0,377$) apresentou o mesmo resultado que o teste de Welch, portanto, compreende-se que não há diferença na centralidade das distribuições amostrais das três medidas.

Similar ao Experimento nº 1, o teste Welch aplicado aos dados do experimento nº 6, apresenta os resultados na Tab. 5, não encontrando evidências de diferença entre as três médias das distribuições amostrais. O teste não paramétrico de Friedman ($\chi^2 = 1,814$; $gl = 2$; $valor p = 0,404$) apresentou o mesmo resultado que o teste de Welch, portanto, compreende-se que não há diferença na centralidade das distribuições amostrais das três medidas.

4.2.3 Distribuição Exponencial

4.2.3.1 Experimento nº 7 e Experimento nº 8

No experimento nº 7 $X_7 \sim Exp(0,02)$ e no experimento nº 8 $X_8 \sim Exp(0,002)$, em ambos os experimentos $n = 5$. Os resultados da análise estatística realizada para os experimentos nº 7 e nº 8 estão apresentados na Tab. 6.

No experimento nº 7, as médias amostrais variaram de 6,119 até 163,184 em torno de uma média 49,192 com um desvio padrão 22,396. As medianas amostrais tiveram o valor mínimo de 2,705 e o valor máximo de 171,737 em torno da média 38,311 com desvio padrão de 23,391. A mCheng teve como valor mínimo 4,696 e valor máximo 154,388 com um média de 43,223 e desvio padrão de 21,067. A média amostral apresentou a média mais próxima do valor esperado (50), seguida da mediana da média. Em relação a assimetria e curtose as três medidas apresentaram problemas com assimetria e curtose acentuadas. A assimetria foi positiva para as três medidas o mesmo verificado nos dados, e a curva leptocúrtica. A média apresentou o menor valor para o coeficiente de assimetria e curtose. A medida mCheng apresentou o menor valor para o desvio padrão enquanto a mediana o maior.

De acordo com a Tab. 6, as distribuições amostrais das três medidas estudadas no experimento nº 8 apresentaram propriedades similares as distribuições encontradas no

Tabela 6: Análise estatística: experimento nº 7 e experimento nº 8

Medidas		Experimento nº 7			Experimento nº 8		
		Média	Mediana	mCheng	Média	Mediana	mCheng
Mínimo		6,119	2,705	4,696	70,808	10,431	44,522
Máximo		163,184	171,737	154,388	1571,338	1661,122	1583,409
Média		49,192	38,311	43,223	507,059	396,826	446,443
Mediana		45,967	32,885	39,730	477,019	345,861	413,386
Desvio Padrão		22,396	23,391	21,067	232,438	245,122	220,702
Q1		32,838	21,269	28,168	338,167	210,505	283,855
Q3		60,465	50,239	53,593	639,266	518,552	565,592
Desvio interquartílico		27,627	28,970	25,425	301,099	308,047	281,737
Assimetria		1,041	1,254	1,169	0,803	1,204	0,964
Curtose		1,656	2,107	2,097	0,907	1,953	1,368
Kolmogorov-Smirnov	Dn	0,074	0,106	0,084	0,061	0,099	0,073
	gl	1000	1000	1000	1000	1000	1000
	valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Shapiro-Wilk	W	0,944	0,914	0,931	0,963	0,918	0,948
	gl	1000	1000	1000	1000	1000	1000
	valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Friedman	χ^2	581,432			490,014		
	gl	2			2		
	valor p	<0,001			<0,001		

Obs.: Erro padrão da assimetria: 0,077; Erro padrão da curtose: 0,155; $n = 5$

experimento nº 7.

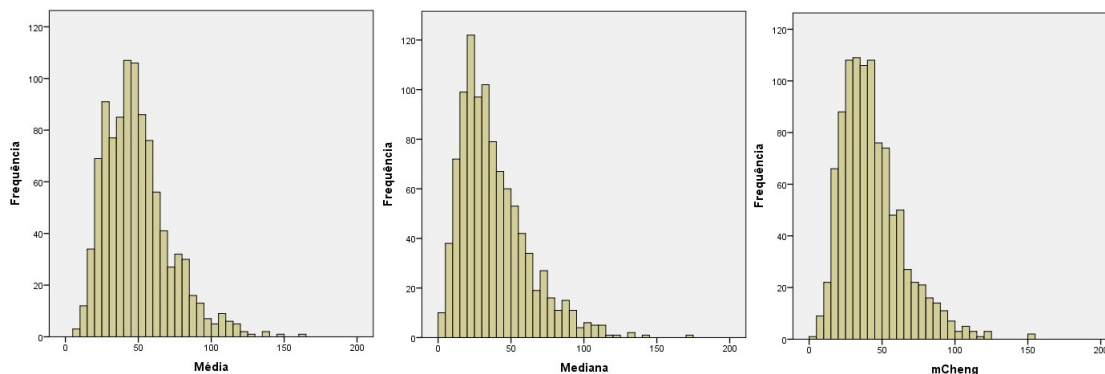


Figura 31: Histogramas: experimento nº 7

Os histogramas, apresentados na Fig. 31, mostram uma assimetria acentuada para as três medidas estudadas. Também sugerem a presença de lacunas superiores nas três medidas. Os gráficos de probabilidade normal, Fig. 32, estão bem afastados da reta nas extremidades, se aproximando no centro. Os boxplots, apresentados na Fig. 33, sugerem assimetria para as três medidas, com vários possíveis valores fora do padrão superiores.

Os histogramas apresentados na Fig. 34, os gráficos de probabilidade normal apresentados na Fig. 35 e os boxplots apresentados na Fig. 36 foram similares aos encontrados no experimento nº 7.

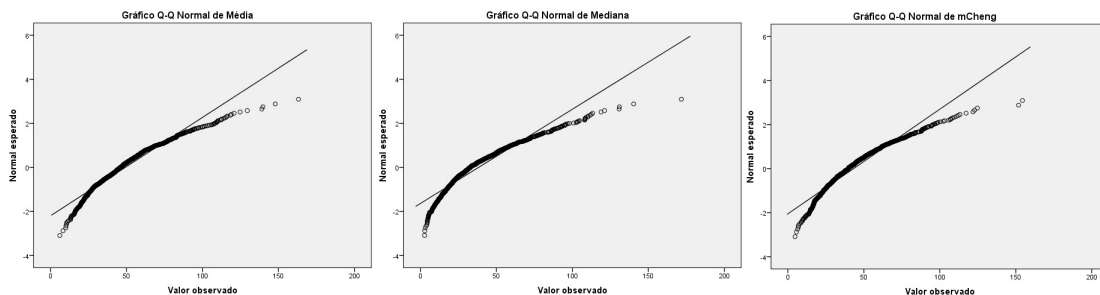


Figura 32: Gráficos de probabilidade normal: experimento nº 7

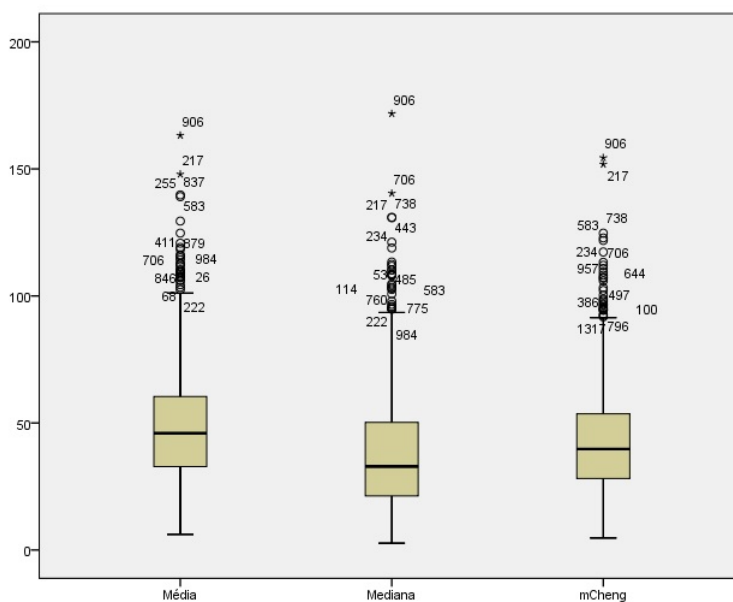


Figura 33: Boxplots: experimento nº 7

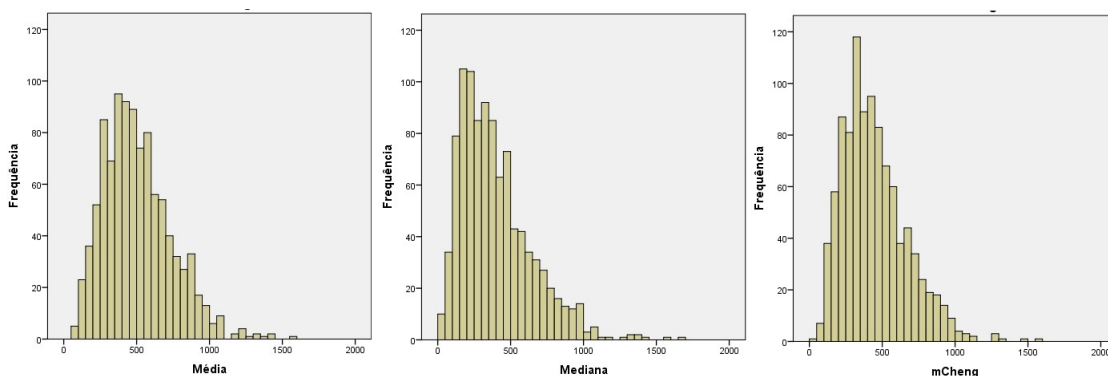


Figura 34: Histogramas: experimento nº 8

A análise foi complementada com os testes de normalidade de Kolmogorov-Smirnov e Shapiro-Wilk, apresentados na Tab. 6, onde ambos os testes mostram evidências para se rejeitar a hipótese nula, ou seja, foram encontradas evidências de que as distribuições estudadas não sejam normais.

Similar ao experimento nº 7, nos resultados dos testes de normalidade do experi-

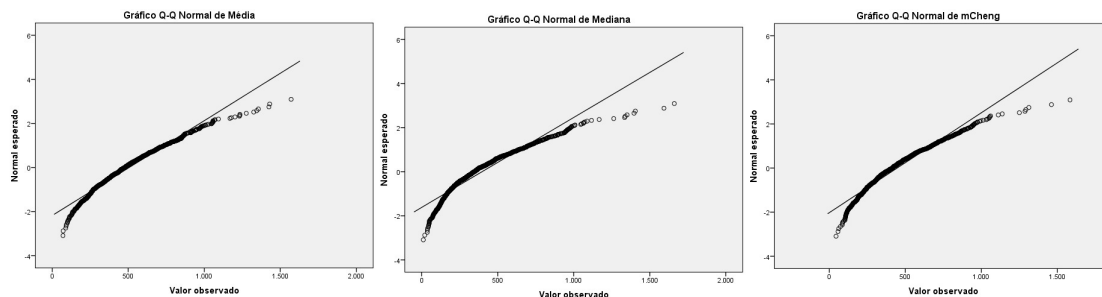


Figura 35: Gráficos de probabilidade normal: experimento nº 8

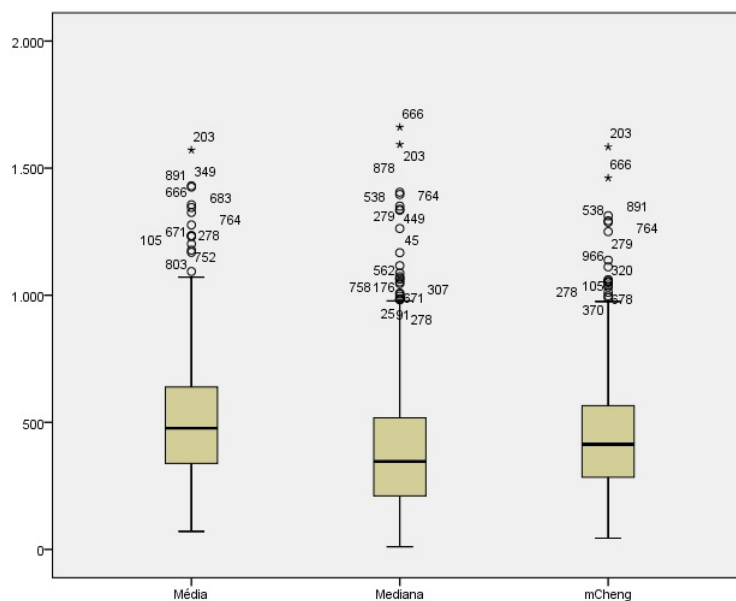


Figura 36: Boxplots: experimento nº 8

mento nº 8, apresentados na Tab. 6, foram encontradas evidências de não normalidade nas distribuições amostrais das três medidas.

Tabela 7: Comparações múltiplas pelo teste de postos de Wilcoxon: experimentos nº 7 e nº 8

Medidas	Experimento nº 7			Experimento nº 8		
	Média-Mediana	Média- <i>mCheng</i>	Mediana- <i>mCheng</i>	Média-Mediana	Média- <i>mCheng</i>	Mediana- <i>mCheng</i>
Z	-19,403	-22,176	-14,942	-18,640	-21,712	-14,153
valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

O teste não paramétrico de Friedman aplicado aos dados do experimento nº 7, apresentou os resultados mostrados na Tab. 6. Portanto, pode-se compreender que existe diferença entre as medianas das distribuições amostrais das três medidas. Na Tab. 7 são apresentados os resultados das comparações múltiplas entre as três medidas realizado pelo teste de postos de Wilcoxon, concluindo-se que as médias das distribuições amostrais das três medidas são diferentes entre si. A menor média é da distribuição amostral da medi-

ana, seguida da distribuição amostral da mCheng. A média da distribuição amostral da média foi a maior.

Para o experimento nº 8 o teste não paramétrico de Friedman, cujos resultados estão na Tab. 6, e as comparações múltiplas realizadas pelo teste de postos de Wilcoxon, apresentados na Tab. 7 apresentaram resultados semelhantes ao experimento nº 7, ou seja, foram encontradas evidências que as medianas das distribuições amostrais das três medidas são diferentes entre si.

4.2.3.2 Experimento nº 9 e Experimento nº 10

No experimento nº 9 $X_9 \sim Exp(0,02)$ e no experimento nº 10 $X_{10} \sim Exp(0,002)$, em ambos os experimentos $n = 30$. Os resultados da análise estatística realizada para os experimentos nº 9 e nº 10 estão apresentados na Tab. 8.

Tabela 8: Análise estatística: experimento nº 9 e experimento nº 10

Medidas		Experimento nº 9			Experimento nº 10		
		Média	Mediana	mCheng	Média	Mediana	mCheng
Mínimo		30,233	12,431	22,016	267,797	133,099	214,608
Máximo		80,523	77,664	71,568	894,406	718,988	792,608
Média		50,433	35,653	40,917	500,328	356,541	406,951
Mediana		49,736	34,881	40,321	491,698	348,619	398,896
Desvio Padrão		9,069	8,941	8,126	93,740	92,475	83,140
Q1		43,734	29,257	35,151	434,685	288,185	349,384
Q3		56,290	41,187	46,078	556,694	410,319	455,108
Desvio interquartílico		12,556	11,930	10,927	122,009	122,134	105,724
Assimetria		0,399	0,482	0,420	0,504	0,661	0,605
Curtose		-0,124	0,346	0,067	0,304	0,768	0,709
Kolmogorov-	Dn	0,043	0,044	0,044	0,051	0,042	0,048
Smirnov	gl	1000	1000	1000	1000	1000	1000
Lilliefors	valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Shapiro-Wilk	W	0,988	0,986	0,988	0,985	0,976	0,980
	gl	1000	1000	1000	1000	1000	1000
	valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
Friedman	χ^2	1807,934			1785,266		
	gl	2			2		
	valor p	<0,001			<0,001		

Obs.: Erro padrão da assimetria: 0,077; Erro padrão da curtose: 0,155; $n = 30$

De acordo com a Tab. 8, as distribuições amostrais das três medidas estudadas no experimento nº 9 apresentaram propriedades similares as das distribuições encontradas no experimento nº 7, diferenciando-se em relação:

- curtose - a média apresenta uma tendência a curva platicúrtica e a mCheng uma tendência a curva leptocúrtica, não acentuada. O menor valor é para a medida mCheng;

- desvio padrão - a média apresenta o maior valor.
- a mediana da distribuição amostral das médias foi a medida de tendência central mais próxima da média da população, seguida da média das médias.

De acordo com a Tab. 8, as distribuições amostrais das três medidas estudadas no experimento nº 10 apresentaram propriedades similares as distribuições encontradas no experimento nº 7, diferenciando-se em relação:

- curtose - a média não apresenta curtose acentuada;
- desvio padrão - a média apresenta o maior valor.

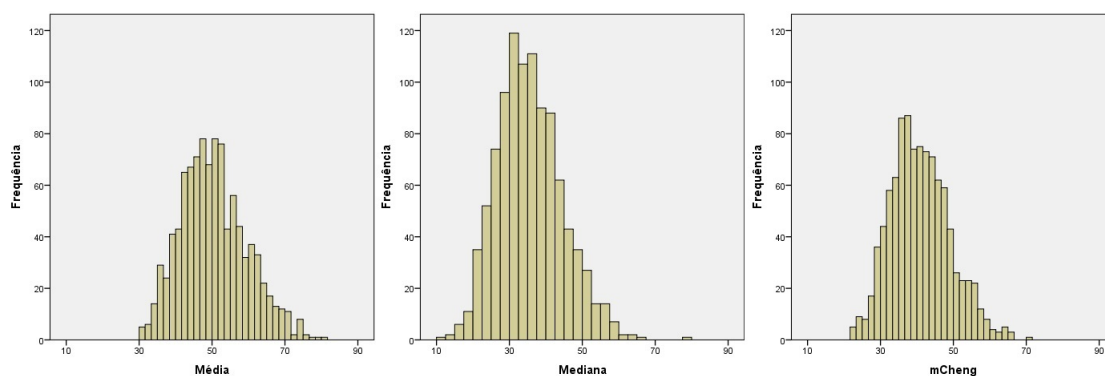


Figura 37: Histogramas: experimento nº 9

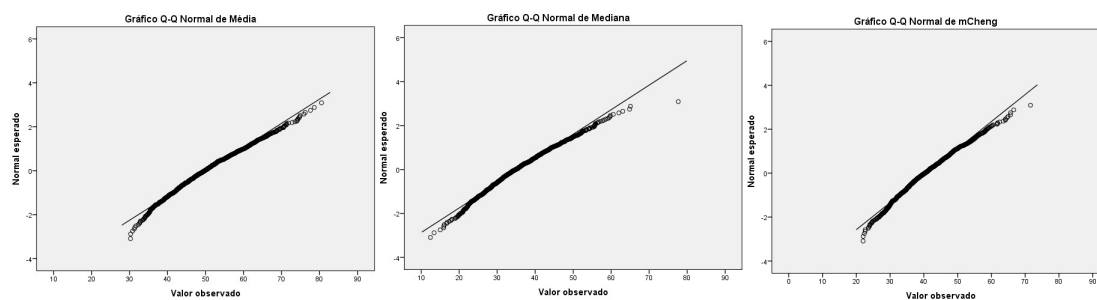


Figura 38: Gráficos de probabilidade normal: experimento nº 9

Os histogramas apresentados na Fig. 37, os gráficos de probabilidade normal apresentados na Fig. 38 e os boxplots apresentados na Fig. 39 foram similares aos encontrados no experimento nº 7. Embora, nota-se que o histograma da média não apresenta lacunas, e os gráficos de probabilidade normal se aproximam mais da reta normal, com afastamento nas extremidades.

Os histogramas apresentados na Fig. 40, os gráficos de probabilidade normal apresentados na Fig. 41 e os boxplots apresentados na Fig. 42 foram similares aos encontrados no experimento nº 7. Apenas, nota-se que o histograma da mediana não apresenta lacunas, e os gráficos apresenta, resultados próximos aos do experimento nº 9.

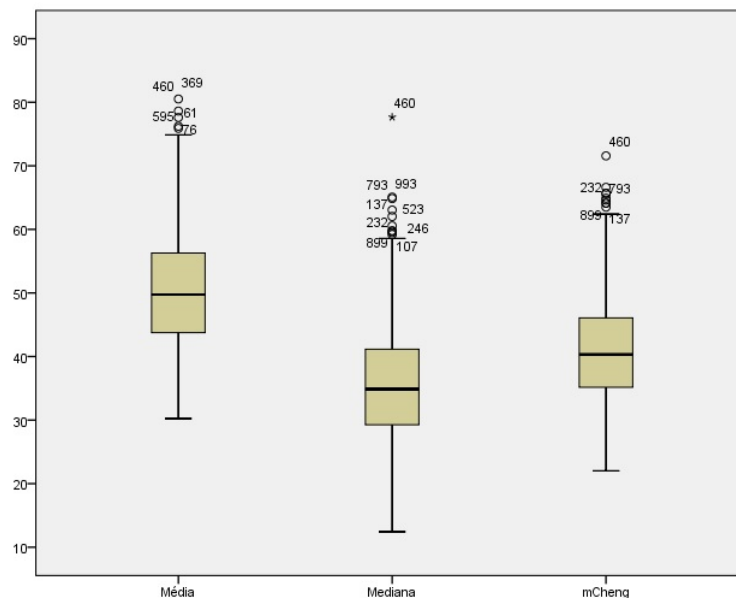


Figura 39: Boxplots: experimento n° 9

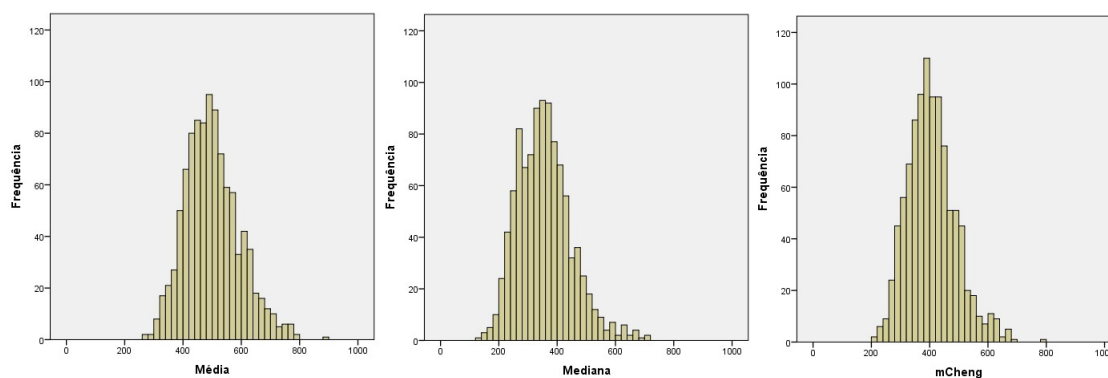


Figura 40: Histogramas: experimento n° 10

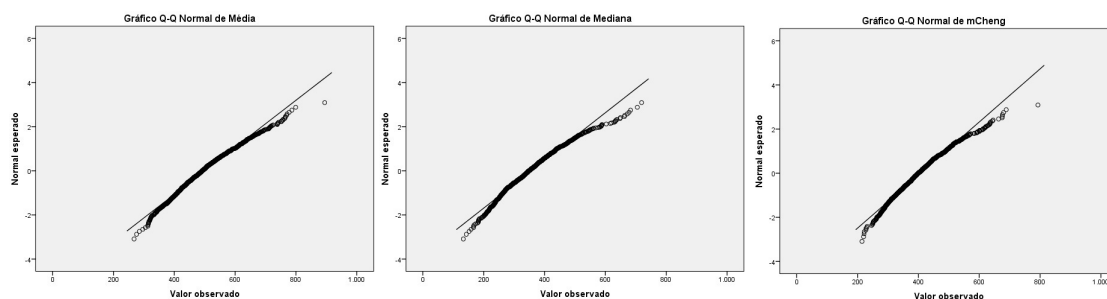


Figura 41: Gráficos de probabilidade normal: experimento n° 10

Similar ao experimento n° 7, nos resultados dos testes de normalidade dos experimentos n° 9 e n° 10, apresentados na Tab. 8, foram encontradas evidências de não normalidade nas distribuições amostrais das três medidas.

De acordo com o teste não paramétrico de Friedman, apresentado na Tab. 8, e as comparações múltiplas pelo teste de postos de Wilcoxon, apresentado na Tab. 9, os experimentos n° 9 e n° 10 apresentaram resultados semelhantes ao experimento n° 7, ou

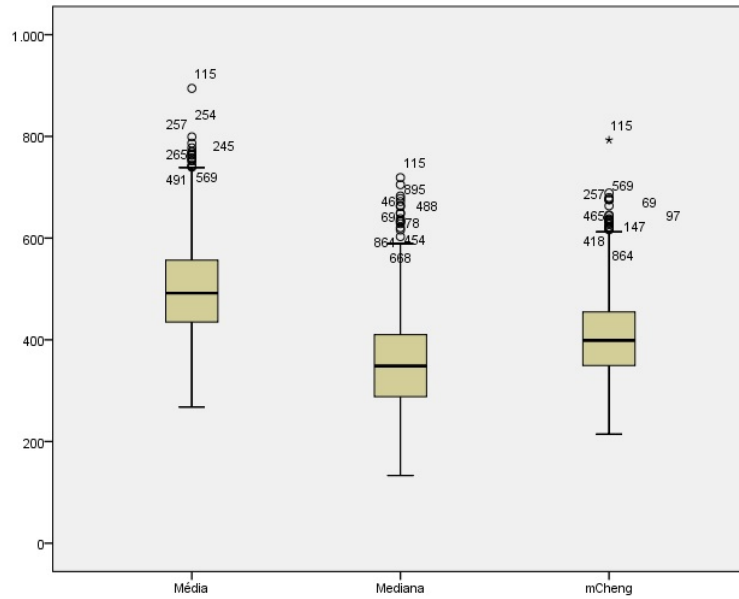


Figura 42: Boxplots: experimento nº 10

Tabela 9: Comparações múltiplas pelo teste de postos de Wilcoxon: experimentos nº 9 e nº 10

Medidas	Experimento nº 9			Experimento nº 10		
	Média-Mediana	Média-mCheng	Mediana-mCheng	Média-Mediana	Média-mCheng	Mediana-mCheng
Z	-27,380	-27,393	-25,743	-27,359	-27,393	-25,427
Valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

seja, foram encontradas evidências de que as medianas das três distribuições amostrais são diferentes.

4.2.3.3 Experimento nº 11 e Experimento nº 12

No experimento nº 11 $X_{11} \sim Exp(0, 02)$ e no experimento nº 12 $X_{12} \sim Exp(0, 002)$, em ambos os experimentos $n = 100$. Os resultados da análise estatística realizada para os experimentos nº 11 e nº 12 estão apresentados na Tab. 10.

De acordo com a Tab. 10, as distribuições amostrais das três medidas estudadas apresentaram propriedades similares as distribuições encontradas no experimento nº 7, diferenciando-se em relação:

- assimetria - não acentuada para a média;
- curtose - não acentuada para a média e a mCheng;
- desvio padrão - a média apresenta o maior valor;
- a mediana da distribuição amostral das médias foi a medida de tendência central mais próxima da média da população, seguida da média das médias.

Tabela 10: Análise estatística: experimento nº 11 e experimento nº 12

Medidas		Experimento nº 11			Experimento nº 12		
		Média	Mediana	mCheng	Média	Mediana	mCheng
Mínimo		35,548	21,521	27,506	346,132	228,604	274,452
Máximo		70,701	55,739	60,157	655,838	601,391	546,198
Média		50,078	34,765	40,108	497,503	346,234	398,611
Mediana		50,030	34,627	40,049	496,697	344,584	395,468
Desvio Padrão		5,032	4,973	4,455	48,318	48,816	42,936
Q1		46,756	31,336	37,037	463,832	311,193	369,356
Q3		53,445	37,802	43,090	528,979	378,092	427,029
Desvio interquartílico		6,689	6,466	6,053	65,147	66,899	57,673
Assimetria		0,034	0,356	0,163	0,152	0,374	0,236
Curtose		0,187	0,472	0,220	0,030	0,393	-0,006
Kolmogorov-Smirnov	Dn	0,015	0,024	0,020	0,022	0,027	0,031
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,200	0,146	0,200	0,200	0,089	0,023
Shapiro-Wilk	W	0,999	0,992	0,997	0,998	0,990	0,996
	gl	1000	1000	1000	1000	1000	1000
	valor p	0,583	<0,001	0,068	0,285	<0,001	0,007
Friedman	χ^2	1976,288			1962,614		
	gl	2			2		
	valor p	<0,001			<0,001		

Obs.: Erro padrão da assimetria: 0,077; Erro padrão da curtose: 0,155; $n = 100$

De acordo com a Tab. 10, as distribuições amostrais das três medidas estudadas apresentaram propriedades similares as distribuições encontradas no experimento nº 7, diferenciando-se em relação:

- assimetria - não acentuada para a média;
- curtose - não acentuada para média e mCheng e tendência a curva platicúrtica e menor valor para a mCheng.

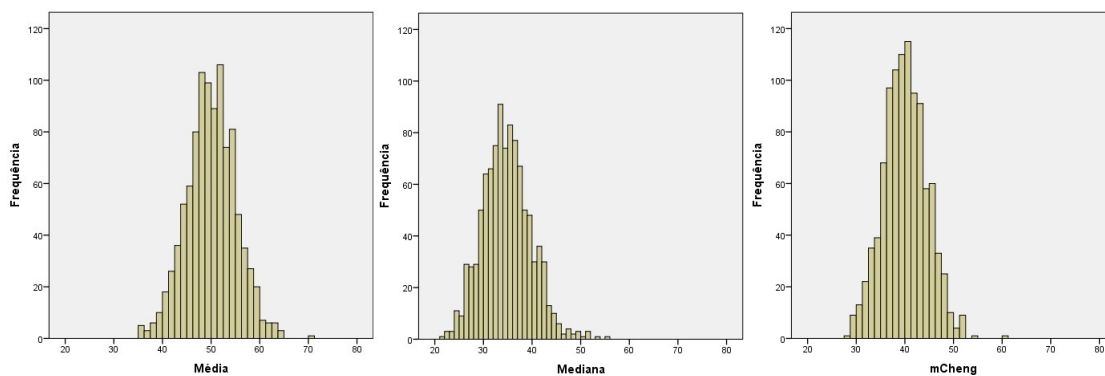


Figura 43: Histogramas: experimento nº 11

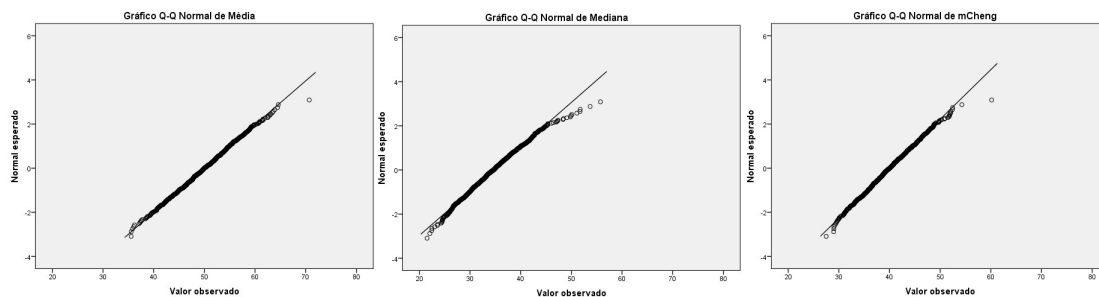


Figura 44: Gráficos de probabilidade normal: experimento nº 11

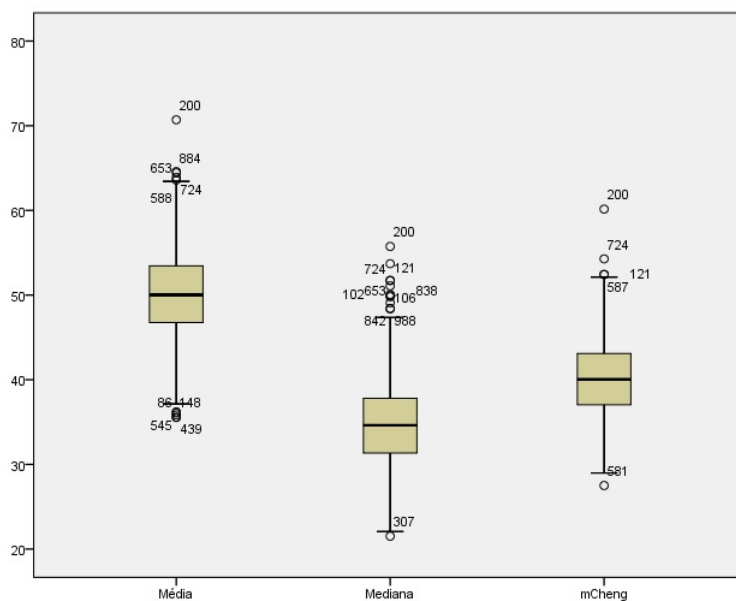


Figura 45: Boxplots: experimento nº 11

Os histogramas apresentados na Fig. 43, os gráficos de probabilidade normal apresentados na Fig. 44 e os boxplots apresentados na Fig. 45 foram similares aos encontrados no experimento nº 7. Apenas, nota-se que os gráficos de probabilidade normal tiveram resultados similares aos encontrados no experimento nº 9, os boxplots também apresentaram possíveis outliers inferiores.

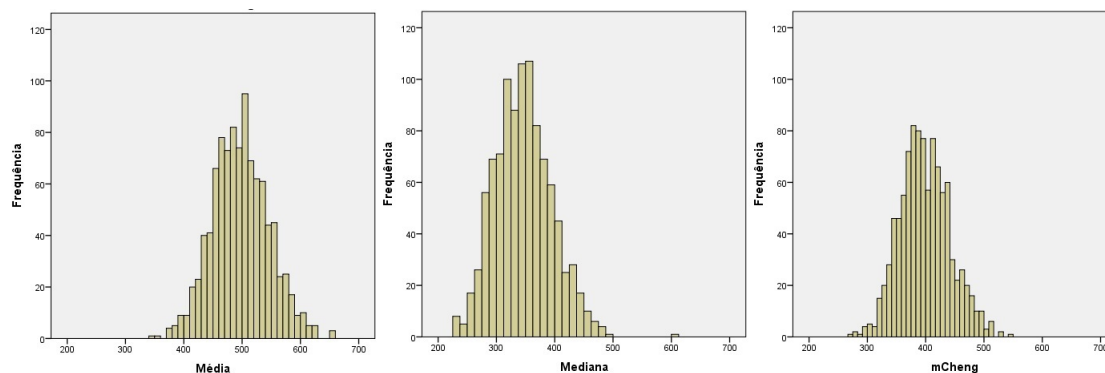


Figura 46: Histogramas: experimento nº 12

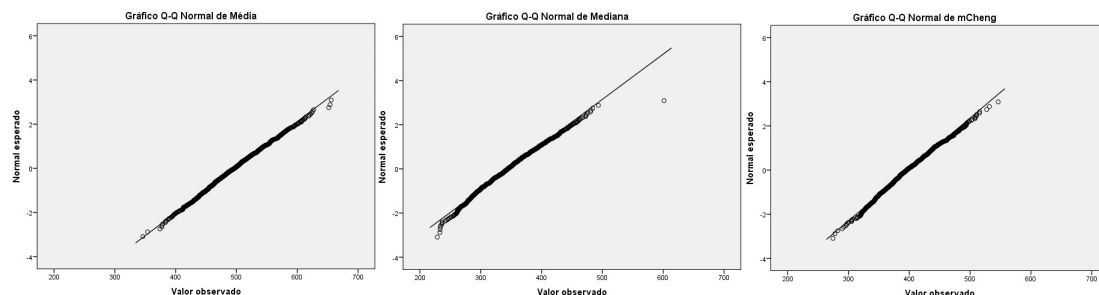


Figura 47: Gráficos de probabilidade normal: experimento nº 12

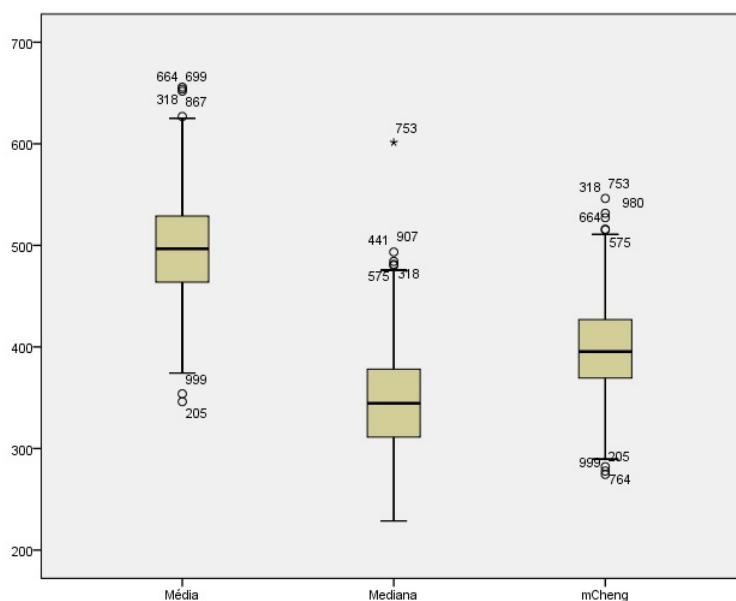


Figura 48: Boxplots: experimento nº 12

Os histogramas apresentados na Fig. 46, os gráficos de probabilidade normal apresentados na Fig. 47 e os boxplots apresentados na Fig. 48 foram similares aos encontrados no experimento nº 7. Apenas, nota-se que o histograma da média apresenta lacuna inferior e os gráficos de probabilidade normal similar ao experimento nº 9. Ainda os boxplots apresentam possíveis valores fora do padrão inferiores para a média e a mCheng.

Os resultados dos testes de normalidade do experimento nº 11, apresentados na Tab. 10, para o teste de Kolmogorov-Smirnov não foram encontradas evidências de não normalidade nas três medidas, no teste de Shapiro-Wilk a média e a mCheng não apresentam evidências de não normalidade e para a mediana foram encontradas evidências de não normalidade.

Os resultados dos testes de normalidade do experimento nº 12, apresentados na Tab. 10, para o teste de Kolmogorov-Smirnov não foram encontradas evidências de não normalidade para a média e a mediana, para a medida mCheng foram encontradas evidências de que a distribuição não seja normal. No teste de Shapiro-Wilk a média não apresenta evidências de não normalidade e para a mediana e a mCheng foram encontradas

evidências de que as distribuições não sejam normais.

Tabela 11: Comparações múltiplas pelo teste de postos de Wilcoxon: experimentos nº 11 e nº 12

Medidas	Experimento nº 11			Experimento nº 12		
	Média-Mediana	Média-mCheng	Mediana-mCheng	Média-Mediana	Média-mCheng	Mediana-mCheng
Z	-27,393	-27,393	-27,348	-27,393	-27,393	-27,295
Valor p	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

De acordo com o teste não paramétrico de Friedman na Tab. 10 e as comparações múltiplas pelo teste de postos de Wilcoxon na Tab. 11, os experimento nº 11 e nº 12 apresentaram resultados semelhantes ao experimento nº 7, ou seja, as medianas das distribuições amostrais das três medidas são diferentes entre si.

5 CONSIDERAÇÕES FINAIS

A realização desse estudo abordou conceitos de estatística, probabilidade, programação e lógica *fuzzy*, tendo como objetivo conhecer a distribuição amostral da medida *mCheng* que estima o centro de um número *fuzzy* triangular no método proposto por CHENG (2005), caracterizando-se como um estudo interdisciplinar.

Seu objetivo geral, conhecer a distribuição amostral da medida de tendência central *mCheng*, foi atingido.

Inicialmente foi detectada uma limitação na operacionalização do método que leva a uma indeterminação matemática quando não existe variabilidade. Entretanto essa limitação deixa de ser importante quando se trabalha com a lógica *fuzzy*.

Em sequência, na avaliação das propriedades matemáticas da média aritmética e da mediana, foi constatado que a medida *mCheng* não possui 3 delas, listadas a seguir:

- A soma algébrica dos desvios tomados em relação a medida *mCheng* não é nula, ao contrário do que acontece com a média aritmética;
- A soma dos quadrados dos afastamentos contados a partir da medida *mCheng* não é um mínimo, ao contrário do que acontece com a média aritmética;
- A medida *mCheng* é influenciada pelos valores extremos, ao contrário do que acontece com a mediana.

Este fato pode limitar o uso desta medida em algumas análises inferenciais.

Nos experimentos nº 1 ao nº 6, em que as amostras foram retiradas aleatoriamente de uma população com distribuição normal, a distribuição da *mCheng* apresentou distribuição aproximadamente normal com parâmetros muito similares aos das demais distribuições amostrais estudadas. As medidas-resumo da distribuição amostral das médias ficaram mais próximas das medidas-resumo da medida *mCheng* do que da mediana. Não foram encontradas evidências de diferença entre as medidas de tendência central das distribuições amostrais, o mesmo não acontecendo em relação a variabilidade. Foram encontradas evidências de heterogeneidade entre as variâncias, sendo que a distribuição amostral da média aquela que apresenta a menor variabilidade. Estas propriedades fazem

com que a média seja o melhor estimador para o *core* do número *fuzzy* triangular, em amostras extraídas de populações com distribuição normal. Além disso, essa medida é bem simples e tem excelentes propriedades matemáticas.

Nos experimentos nº 7 ao nº 12, as amostras foram retiradas aleatoriamente de uma população com distribuição exponencial, apresentando uma assimetria positiva acentuada, o que já é um indicativo de que a média não teria um bom desempenho para encontrar o *core* do número *fuzzy* triangular. Nas amostras com tamanho $n = 5$ e $n = 30$, nenhuma das distribuições amostrais foi considerada próxima a distribuição normal. Ao comparar a tendência central das distribuições amostrais foram encontradas evidências de diferença, sendo a menor, a tendência central das medianas, e a maior, a tendência central das médias. Também foram encontradas evidências de diferença entre as variabilidades das distribuições amostrais, sendo a menor, a da distribuição amostral da *mCheng*. Nas amostras de tamanho $n = 100$ as distribuições amostrais da média se aproximaram da distribuição normal. A distribuição amostral das medianas, pelo teste Shapiro-Wilk, apresentou indícios de não normalidade, enquanto a distribuição amostral do *mCheng* apresentou evidências de não normalidade por ambos os testes no experimento em que a população apresentava maior variabilidade. Em relação a variabilidade, a distribuição da *mCheng* foi a de menor dispersão. Esses achados indicam que, em amostras de até 30 elementos extraídas de população com distribuição exponencial, a medida *mCheng* parece ser mais adequada para encontrar o *core* do número *fuzzy* triangular. Este procedimento, entretanto, pode ser o indicado para amostras de tamanho maior que 30 mas menor que 100, pois as simulações foram executadas apenas para amostras com 5, 30 e 100 elementos. Mais estudos são necessários para definir o tamanho da amostra limite para utilização da medida *mCheng* para caracterizar a tendência central de dados provenientes de uma distribuição exponencial.

Além disso, chama a atenção o fato de que a distribuição amostral da média de amostras de até 30 elementos, extraídas de população com distribuição exponencial apresentaram distribuição bastante diferenciada de uma distribuição normal, pois de acordo com COSTA (2012), MORETTIN (2010), LEME (1974) e DOWNING; CLARK (1998), este modelo teórico poderia ser utilizado para estudar o comportamento de médias obtidas em amostras com pelo menos 30 elementos, independentemente do tipo de distribuição na população. Na distribuição amostral de amostras com 100 elementos, esta premissa parece válida, mas sugere-se utilizar um limite maior que 30 para aplicação da regra.

Em distribuições assimétricas, como é o caso da distribuição exponencial, a tendência central, considerada como o valor em torno do qual existe maior concentração de dados, é melhor representada pela mediana, pois a média aritmética se desloca no sentido da cauda mais longa e se afasta do ponto que se deseja encontrar. Entretanto, a distribuição amostral da mediana apresenta maior variabilidade, o que não é desejável. A medida *mCheng* consegue se aproximar mais do ponto com maior concentração do que a média aritmética

e consegue apresentar menor variabilidade que a mediana e que a própria média.

PINHEIRO et al. (2012) gerou dados por simulação a partir de uma distribuição exponencial com média e desvio padrão igual a 3, gerando amostras de tamanho 1 até 20. Os autores mostraram que a medida que o tamanho da amostra cresce a forma do histograma se aproxima cada vez mais de uma curva normal. Encontrando que a partir de $n = 5$ uma aproximação bastante satisfatória, discordando portanto, dos achados deste estudo. Chama-se a atenção para o fato, entretanto, que a média e o desvio padrão da distribuição exponencial utilizada por esses autores, foi bem menor do que a média e o desvio padrão utilizado nesse estudo. Evidencia-se aqui a necessidade do desenvolvimento de estudos para avaliar a influência do desvio padrão da população na convergência para a distribuição normal da distribuição amostral das medidas de tendência central.

SILVESTRE (2014) desenvolveu estudo aplicando o método Cheng para geração de número *fuzzy* a partir de amostras de tamanho 5. Seus dados, obtidos empiricamente, apresentaram evidências de não normalidade. Os achados deste estudo mostram que a determinação do *core* do número *fuzzy* pela medida *mCheng* foi adequada. Este estudo poderia ser complementado com sua replicação para conhecer e avaliar as distribuições amostrais das medidas de variabilidade usadas na determinação dos extremos do número *fuzzy* triangular, a e b .

Como trabalhos futuros sugere-se ainda replicar o estudo aumentando a quantidade de amostras simuladas.

Os achados deste estudo são importantes porque o conhecimento do comportamento das distribuições amostrais das diversas estatísticas podem dar maior eficiência aos resultados da estatística inferencial.

REFERÊNCIAS

- ANGELINI, F.; MILONE, G. **Estatística geral**. São Paulo: Atlas, 1993. v.1.
- ANGELINI, F.; MILONE, G. **Estatística geral**. São Paulo: Atlas, 1993. v.2.
- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. **Estatística para cursos de Engenharia e Informática**. São Paulo: Editora Atlas, 2010.
- BARROS, L. C. de; BASSANEZI, R. C. Tópicos de lógica fuzzy e biomatemática. **Coleção IMECC: textos didáticos**, Universidade Estadual de Campinas, v.5, 2006.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. Rio de Janeiro: SBM, 2010.
- BRIAO, S. L. **Modelagem em programação linear para resolução de jogos fuzzy intervalares**. 2013. Dissertação de Mestrado. Programa de Pós-Graduação em Modelagem Computacional — Universidade Federal do Rio Grande.
- BUSSAB, W. d. O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 2005.
- CAMPOS FILHO, P. **Método para apoio à decisão na verificação da sustentabilidade de uma unidade de conservação, usando lógica fuzzy**. 2004. Tese de Doutorado. Programa de Pós-Graduação em Engenharia de Produção — Universidade Federal de Santa Catarina.
- CASELLA, G.; BERGER, R. L. **Inferência estatística**. São Paulo: [s.n.], 2010.
- CASTRO, L. S. V. d. **Pontos de estatística**. São Paulo: Científica, 1970.
- CHEN, G.; PHAM, T. T. **Introduction to Fuzzy Sets, Fuzzy logic and Fuzzy Control Systems**. New York: CRC Press, 2000.
- CHENG, C.-B. Fuzzy process control: construction of control charts with fuzzy numbers. **Fuzzy sets and systems**, Amsterdã, v.154, n.2, p.287–303, 2005.
- COSTA, G. G. d. O. **Curso de estatística inferencial e probabilidades: teoria e prática**. São Paulo: Atlas, 2012.

CRESPO, A. A. **Estatística fácil**. São Paulo: Saraiva, 2009.

DELGADO, M.; VERDEGAY, J. L.; VILA, M. A. A procedure for ranking fuzzy numbers using fuzzy relations. **Fuzzy sets and systems**, Amsterdã, v.26, n.1, p.49–62, 1988.

DEVORE, J. L. **Probabilidade e Estatística para Engenharia e Ciências**. São Paulo: Pioneira Thomson Learning, 2006.

DODGE, Y. **The Oxford dictionary of statistical terms**. Oxford: Oxford University Press, 2003.

DOWNING, D.; CLARK, J. **Estatística aplicada**. São Paulo: Saraiva, 1998.

GUIMARÃES, R. O. **Avaliação dos riscos de violações de conformidade de tensão em sistemas de distribuição, utilizando método probabilístico e conjuntos fuzzy**. 2008. Tese de Doutorado. Programa de Pós-Graduação em Engenharia — Universidade de São Paulo.

KLIR, G. J.; YUAN, B. **Fuzzy Sets and Fuzzy Logic: Theory and Applications**. New Jersey: Prentice Hall, 1995.

KOHAGURA, T. Lógica fuzzy e suas aplicações. **Departamento de Computação, Universidade Estadual de Londrina**, Londrina, 2007. Trabalho de Conclusão de Curso. Curso de Ciências da Computação.

LEME, R. A. d. S. **Curso de estatística: elementos**. Rio de Janeiro: Livros Técnicos e Científicos, 1974.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. São Paulo: Edusp, 2006.

MAGALHÃES, M. N.; LIMA, A. C. P. de. **Noções de probabilidade e estatística**. [S.l.]: IME-USP, 2008.

MATTOS, V. L. D.; COELHO, L. C. Propriedades de controladores fuzzy: um estudo de caso. **Scientia Plena**, São Cristóvão, v.11, p.1–7, 2015.

MATTOS, V. L. D. d.; KONRATH, A. C.; AZAMBUJA, A. M. V. d. **Introdução a estatística - aplicações em ciências exatas**. Rio de Janeiro: LTC, 2017.

MATTOS, V. L. D. de et al. Avaliação de um método de transformação de valores crisp em valores fuzzy. **Revista Gepros**, Bauru, v.10, n.4, 2015.

MONTGOMERY, D.; RUNGER, C. **Estatística Aplicada e Probabilidade para Engenheiros**. Rio de Janeiro: Editora LTC, 2009.

MORETTIN, L. G. **Estatística básica: probabilidade e inferência**. São Paulo: Pearson Prentice Hall, 2010.

PINHEIRO, J. I. et al. **Probabilidade e Estatística**. Rio de Janeiro: Elsevier, 2012.

PORTAL Action. Software de Estatística Online. Acesso em: 10 jun 2017.

ROSA, S. E. **Modelo Fuzzy de avaliação de risco para o reconhecimento de Cursos de Graduação**. 2012. Dissertação de Mestrado. Programa de Pós-Graduação em Modelagem Computacional — Universidade Federal do Rio Grande.

SHAW, I. S.; SIMOES, M. G. Controle e modelagem fuzzy. **Blucher**, São Paulo, 2007.

SILVESTRE, I. B. M. **Gráficos de controle: aspectos teóricos e práticos a partir da ótica da lógica formal e da lógica fuzzy**. 2014. Dissertação de Mestrado. Programa de Pós-Graduação em Modelagem Computacional — Universidade Federal do Rio Grande.

UPTON, G.; COOK, I. **A dictionary of statistics**. Oxford: Oxford University Press, 2008.

VIANA, K. D. **Modelo hierárquico coppe-cosenza fuzzy: adaptação e análise de seu desempenho para avaliação de satisfação de clientes**. 2017. Dissertação de Mestrado. Programa de Pós-Graduação em Modelagem Computacional — Universidade Federal do Rio Grande.

WASSERMAN, L. **All of statistics: a concise course in statistical inference**. Dordrecht: Springer Science & Business Media, 2003.

WEISBERG, H. F. Central tendency and variability. **Sage University Paper Series on Quantitative Applications in the Social Sciences**, Newbury Park, CA, n.07–083, 1992.

ZADEH, L. A. Fuzzy sets. **Information and Control**, Amsterdã, v.8, n.3, p.338–353, 1965.

ZADEH, L. A. Outline of a new approach to the analysis of complex systems and decision processes. **IEEE Transactions on Systems, Man, and Cybernetics**, Systems, Man, and Cybernetics Society, v.3, p.28–44, 1973.

ZADEH, L. A. The Concept of a Linguistic Variable and its Applications to approximate Reasoning-I. **Information Science**, Amsterdã, v.8, n.3, p.199–249, 1975.

ZADEH, L. A. The Concept of a Linguistic Variable and its Applications to approximate Reasoning-II. **Information Science**, Amsterdã, v.8, p.301–357, 1975.

ZADEH, L. A. The Concept of a Linguistic Variable and its Applications to approximate Reasoning-III. **Information Science**, Amsterdã, v.9, p.43–80, 1975.

ANEXO A EXPRESSÕES MATEMÁTICAS

- Média:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (115)$$

- Mediana:

$$md(X) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ ímpar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ par} \end{cases} \quad (116)$$

- Desvio Padrão:

$$s = \sqrt{\frac{\sum_{i=1}^n x_i - \bar{X}}{n}} \quad (117)$$

- Primeiro Quartil:

$$Q1 = \begin{cases} x_{(\frac{n+1}{4})}, & \text{se } n \text{ ímpar} \\ x_{(\frac{n+2}{4})}, & \text{se } n \text{ par} \end{cases} \quad (118)$$

- Terceiro Quartil:

$$Q3 = \begin{cases} x_{(\frac{3(n+1)}{4})}, & \text{se } n \text{ ímpar} \\ x_{(\frac{3n+2}{4})}, & \text{se } n \text{ par} \end{cases} \quad (119)$$

- Assimetria:

$$As = \frac{n \sum_{i=1}^n (x_i - \bar{X})^3}{(n-1)(n-2)s^3} \quad (120)$$

- Erro Padrão da Assimetria:

$$EPAs = \sqrt{\frac{6n(n-1)}{(n-2)(n-1)(n+3)}} \quad (121)$$

- Curtose:

$$Curtose = \frac{n(n+1) \sum_{i=1}^n (x_i - \bar{X})^4 - 3s^4(n-1)}{(n-1)(n-2)(n-3)s^4} \quad (122)$$

- Erro Padrão da Curtose:

$$EPC_{urtose} = 2 \cdot \sqrt{\frac{6n(n-1)}{(n-2)(n-1)(n+3)}} \cdot \sqrt{\frac{n^2-1}{(n-3)(n+5)}} \quad (123)$$

- Estatística de teste Kolmogorov-Smirnov Lilliefors:

$$D_n = \sup_x |F(x) - F_n(x)| \quad (124)$$

- Estatística de teste Shapiro-Wilk:

$$W = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (125)$$

- Estatística de teste Levene:

$$F_* = \frac{\sum_{i=1}^k \frac{n_i(\bar{z}_i - \bar{z})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2}{\sum_{i=1}^k (n_i - 1)}} \quad (126)$$

- Estatística de teste Welch:

$$F_C = \frac{\sum_{i=1}^k w_i \frac{(\bar{y}_i - \bar{y}^*)^2}{k-1}}{1 + \frac{2(k-2)\Omega}{k^2-1}} \sim F(v_1, v_2) \quad (127)$$

- Estatística de teste Friedman:

$$\chi^2 = \frac{12b}{k(k+1)} \sum_{j=1}^k \left(\frac{R_j}{b} - \frac{k+1}{2} \right)^2 = \left[\frac{12}{bk(k+1)} \sum_{j=1}^k R_j^2 \right] - 3b(k+1) \quad (128)$$

APÊNDICE A PROGRAMAS CONSTRUÍDOS PARA AS SIMULAÇÕES

A.1 Simulações com distribuição normal

```
1 //Programa para as simulações com distribuição normal
2 //Entra com o dado da média populacional
3 mi=input("Entre com o valor da média populacional: ");
4 //Programa para as simulações com distribuição normal
5 //Entra com o dado da média populacional
6 mi=input("Entre com o valor da média populacional: ");
7 //Entra com os valores do desvio padrão um e dois
8 sd_um=input("Entre com o valor do desvio padrão um: ");
9 sd_dois=input("Entre com o valor do desvio padrão dois: ");
10 //Entra com o tamanho da amostra um, dois e três
11 n_um=input("Entre com o tamanho da amostra um: ");
12 n_dois=input("Entre com o tamanho da amostra dois: ");
13 n_tres=input("Entre com o tamanho da amostra três: ");
14 //Define o número de amostras para cada condição experimental
15 amostra=input("Entre com o número de amostra para cada condição
    experimental: ");
16 //Condições experimentais e Geração das amostras
17 //Condição 1 (sd_um,n_um) - matriz A
18 A=grand(amostra,n_um,"nor",mi,sd_um);
19 //Condição 2 (sd_dois,n_um) - matriz B
20 B=grand(amostra,n_um,"nor",mi,sd_dois);
21 //Condição 3 (sd_um,n_dois) - matriz C
22 C=grand(amostra,n_dois,"nor",mi,sd_um);
23 //Condição 4 (sd_dois,n_dois) - matriz D
24 D=grand(amostra,n_dois,"nor",mi,sd_dois);
25 //Condição 5 (sd_um,n_tres) - matriz E
26 E=grand(amostra,n_tres,"nor",mi,sd_um);
27 //Condição 6 (sd_dois,n_tres) - matriz F
```

```
28 F=grand(amostra,n_tres,"nor",mi,sd_dois);
29 //Cálculo da média de A
30 media_a=mean(A,'c');
31 //Cálculo da mediana de A
32 mediana_a=median(A,'c');
33 //Cálculo da média de B
34 media_b=mean(B,'c');
35 //Cálculo da mediana de B
36 mediana_b=median(B,'c');
37 //Cálculo da média de C
38 media_c=mean(C,'c');
39 //Cálculo da mediana de C
40 mediana_c=median(C,'c');
41 //Cálculo da média de D
42 media_d=mean(D,'c');
43 //Cálculo da mediana de D
44 mediana_d=median(D,'c');
45 //Cálculo da média de E
46 media_e=mean(E,'c');
47 //Cálculo da mediana de E
48 mediana_e=median(E,'c');
49 //Cálculo da média de F
50 media_f=mean(F,'c');
51 //Cálculo da mediana de F
52 mediana_f=median(F,'c');
53 //Cálculo do "m" do Método de Cheng de A
54 //Lê cada linha da matriz como vetor g
55 for i=1:1:amostra
56     for j=1:1:n_um
57         g(j)=A(i,j);
58     end
59 //Calcula o tamanho do vetor
60     tam=length(g);
61 //Cria a matriz N
62     for r=1:1:tam
63         for s=1:1:tam
64             N(r,s)=abs(g(r)-g(s));
65         end
66     end
67 //Cálculo do vetor d_barra
68     for r=1:1:tam
```

```

69     di(r)=0;
70     for s=1:1:tam
71         di(r)=N(r,s)+di(r);
72     end
73     di(r)=(di(r))/(tam-1);
74 end
75 //Cria a matriz P
76     for r=1:1:tam
77         for s=1:1:tam
78             P(r,s)=di(s)/di(r);
79         end
80     end
81 //Encontrar os pesos wi
82     for r=1:1:tam
83         n=0;
84         for s=1:1:tam
85             n=n+P(s,r);
86         end
87         w(r)=1/n;
88     end
89 //Encontrar o valor mcheng
90     mcheng=0;
91     for r=1:1:tam
92         mcheng=w(r)*g(r)+mcheng;
93     end
94     mcheng_vet_a(i)=mcheng;
95 end
96 //Cálculo do "m" do Método de Cheng de B
97 for i=1:1:amostra
98     for j=1:1:n_um
99         g(j)=B(i,j);
100    end
101    tam=length(g);
102    for r=1:1:tam
103        for s=1:1:tam
104            N(r,s)=abs(g(r)-g(s));
105        end
106    end
107    for r=1:1:tam
108        di(r)=0;
109        for s=1:1:tam

```

```

110         di(r)=N(r,s)+di(r);
111     end
112     di(r)=(di(r))/(tam-1);
113 end
114 for r=1:1:tam
115     for s=1:1:tam
116         P(r,s)=di(s)/di(r);
117     end
118 end
119 for r=1:1:tam
120     n=0;
121     for s=1:1:tam
122         n=n+P(s,r);
123     end
124     w(r)=1/n;
125 end
126 mcheng=0;
127 for r=1:1:tam
128     mcheng=w(r)*g(r)+mcheng;
129 end
130 mcheng_vet_b(i)=mcheng;
131 end
132 //Cálculo do "m" do Método de Cheng de C
133 for i=1:1:amostra
134     for j=1:1:n_dois
135         g(j)=C(i,j);
136     end
137     tam=length(g);
138     for r=1:1:tam
139         for s=1:1:tam
140             N(r,s)=abs(g(r)-g(s));
141         end
142     end
143     for r=1:1:tam
144         di(r)=0;
145         for s=1:1:tam
146             di(r)=N(r,s)+di(r);
147         end
148         di(r)=(di(r))/(tam-1);
149     end
150     for r=1:1:tam

```

```

151     for s=1:1:tam
152         P(r,s)=di(s)/di(r);
153     end
154 end
155 for r=1:1:tam
156     n=0;
157     for s=1:1:tam
158         n=n+P(s,r);
159     end
160     w(r)=1/n;
161 end
162 mcheng=0;
163 for r=1:1:tam
164     mcheng=w(r)*g(r)+mcheng;
165 end
166 mcheng_vet_c(i)=mcheng;
167 end
168 //Cálculo do "m" do Método de Cheng de D
169 for i=1:1:amostra
170     for j=1:1:n_dois
171         g(j)=D(i,j);
172     end
173     tam=length(g);
174     for r=1:1:tam
175         for s=1:1:tam
176             N(r,s)=abs(g(r)-g(s));
177         end
178     end
179     for r=1:1:tam
180         di(r)=0;
181         for s=1:1:tam
182             di(r)=N(r,s)+di(r);
183         end
184         di(r)=(di(r))/(tam-1);
185     end
186     for r=1:1:tam
187         for s=1:1:tam
188             P(r,s)=di(s)/di(r);
189         end
190     end
191     for r=1:1:tam

```



```

192     n=0;
193     for s=1:1:tam
194         n=n+P(s,r);
195     end
196     w(r)=1/n;
197 end
198 mcheng=0;
199 for r=1:1:tam
200     mcheng=w(r)*g(r)+mcheng;
201 end
202 mcheng_vet_d(i)=mcheng;
203 end
204 //Cálculo do "m" do Método de Cheng de E
205 for i=1:1:amostra
206     for j=1:1:n_tres
207         g(j)=E(i,j);
208     end
209     tam=length(g);
210     for r=1:1:tam
211         for s=1:1:tam
212             N(r,s)=abs(g(r)-g(s));
213         end
214     end
215     for r=1:1:tam
216         di(r)=0;
217         for s=1:1:tam
218             di(r)=N(r,s)+di(r);
219         end
220         di(r)=(di(r))/(tam-1);
221     end
222     for r=1:1:tam
223         for s=1:1:tam
224             P(r,s)=di(s)/di(r);
225         end
226     end
227     for r=1:1:tam
228         n=0;
229         for s=1:1:tam
230             n=n+P(s,r);
231         end
232         w(r)=1/n;

```

```
233     end
234     mcheng=0;
235     for r=1:1:tam
236         mcheng=w(r)*g(r)+mcheng;
237     end
238     mcheng_vet_e(i)=mcheng;
239 end
240 //Cálculo do "m" do Método de Cheng de F
241 for i=1:1:amostra
242     for j=1:1:n_tres
243         g(j)=F(i,j);
244     end
245     tam=length(g);
246     for r=1:1:tam
247         for s=1:1:tam
248             N(r,s)=abs(g(r)-g(s));
249         end
250     end
251     for r=1:1:tam
252         di(r)=0;
253         for s=1:1:tam
254             di(r)=N(r,s)+di(r);
255         end
256         di(r)=(di(r))/(tam-1);
257     end
258     for r=1:1:tam
259         for s=1:1:tam
260             P(r,s)=di(s)/di(r);
261         end
262     end
263     for r=1:1:tam
264         n=0;
265         for s=1:1:tam
266             n=n+P(s,r);
267         end
268         w(r)=1/n;
269     end
270     mcheng=0;
271     for r=1:1:tam
272         mcheng=w(r)*g(r)+mcheng;
273     end
```

```

274     mcheng_vet_f(i)=mcheng;
275 end
276 //Resultado de A
277 resultado_a=[media_a,mediana_a,mcheng_vet_a];
278 //Resultado de B
279 resultado_b=[media_b,mediana_b,mcheng_vet_b];
280 //Resultado de C
281 resultado_c=[media_c,mediana_c,mcheng_vet_c];
282 //Resultado de D
283 resultado_d=[media_d,mediana_d,mcheng_vet_d];
284 //Resultado de E
285 resultado_e=[media_e,mediana_e,mcheng_vet_e];
286 //Resultado de F
287 resultado_f=[media_f,mediana_f,mcheng_vet_f];

```

A.2 Simulações com distribuição exponencial

```

1 //Programa para as simulações com distribuição exponencial
2 //Entra com os dados da médias populacionais
3 mi_um=input("Entre com o valor da média populacional um: ");
4 mi_dois=input("Entre com o valor da média populacional dois: ");
5 //Entra com o tamanho da amostra um, dois e três
6 n_um=input("Entre com o tamanho da amostra um: ");
7 n_dois=input("Entre com o tamanho da amostra dois: ");
8 n_tres=input("Entre com o tamanho da amostra três: ");
9 //Define o número de amostras
10 amostra=input("Entre com o número de amostras: ");
11 //Condições experimentais e Geração das amostras
12 //Condição 1 (mi_um,n_um) - matriz A
13 A=grand(amostra,n_um,"exp",mi_um);
14 //Condição 2 (mi_dois,n_um) - matriz B
15 B=grand(amostra,n_um,"exp",mi_dois);
16 //Condição 3 (mi_um,n_dois) - matriz C
17 C=grand(amostra,n_dois,"exp",mi_um);
18 //Condição 4 (mi_dois,n_dois) - matriz D
19 D=grand(amostra,n_dois,"exp",mi_dois);
20 //Condição 5 (mi_um,n_tres) - matriz E
21 E=grand(amostra,n_tres,"exp",mi_um);
22 //Condição 6 (mi_dois,n_tres) - matriz F

```

```
23 F=grand(amostra,n_tres,"exp",mi_dois);
24 //Cálculo da média de A
25 media_a=mean(A,'c');
26 //Cálculo da mediana de A
27 mediana_a=median(A,'c');
28 //Cálculo da média de B
29 media_b=mean(B,'c');
30 //Cálculo da mediana de B
31 mediana_b=median(B,'c');
32 //Cálculo da média de C
33 media_c=mean(C,'c');
34 //Cálculo da mediana de C
35 mediana_c=median(C,'c');
36 //Cálculo da média de D
37 media_d=mean(D,'c');
38 //Cálculo da mediana de D
39 mediana_d=median(D,'c');
40 //Cálculo da média de E
41 media_e=mean(E,'c');
42 //Cálculo da mediana de E
43 mediana_e=median(E,'c');
44 //Cálculo da média de F
45 media_f=mean(F,'c');
46 //Cálculo da mediana de F
47 mediana_f=median(F,'c');
48 //Cálculo do "m" do Método de Cheng de A
49 //Lê cada linha da matriz como vetor g
50 for i=1:1:amostra
51     for j=1:1:n_um
52         g(j)=A(i,j);
53     end
54 //Calcula o tamanho do vetor
55     tam=length(g);
56 //Cria a matriz N
57     for r=1:1:tam
58         for s=1:1:tam
59             N(r,s)=abs(g(r)-g(s));
60         end
61     end
62 //Cálculo do vetor d_barra
63     for r=1:1:tam
```

```

64     di(r)=0;
65     for s=1:1:tam
66         di(r)=N(r,s)+di(r);
67     end
68     di(r)=(di(r))/(tam-1);
69 end
70 //Cria a matriz P
71     for r=1:1:tam
72         for s=1:1:tam
73             P(r,s)=di(s)/di(r);
74         end
75     end
76 //Encontrar os pesos wi
77     for r=1:1:tam
78         n=0;
79         for s=1:1:tam
80             n=n+P(s,r);
81         end
82         w(r)=1/n;
83     end
84 //Encontrar o valor mcheng
85     mcheng=0;
86     for r=1:1:tam
87         mcheng=w(r)*g(r)+mcheng;
88     end
89     mcheng_vet_a(i)=mcheng;
90 end
91 //Cálculo do "m" do Método de Cheng de B
92 for i=1:1:amostra
93     for j=1:1:n_um
94         g(j)=B(i,j);
95     end
96     tam=length(g);
97     for r=1:1:tam
98         for s=1:1:tam
99             N(r,s)=abs(g(r)-g(s));
100        end
101    end
102    for r=1:1:tam
103        di(r)=0;
104        for s=1:1:tam

```

```

105         di(r)=N(r,s)+di(r);
106     end
107     di(r)=(di(r))/(tam-1);
108 end
109 for r=1:1:tam
110     for s=1:1:tam
111         P(r,s)=di(s)/di(r);
112     end
113 end
114 for r=1:1:tam
115     n=0;
116     for s=1:1:tam
117         n=n+P(s,r);
118     end
119     w(r)=1/n;
120 end
121 mcheng=0;
122 for r=1:1:tam
123     mcheng=w(r)*g(r)+mcheng;
124 end
125 mcheng_vet_b(i)=mcheng;
126 end
127 //Cálculo do "m" do Método de Cheng de C
128 for i=1:1:amostra
129     for j=1:1:n_dois
130         g(j)=C(i,j);
131     end
132     tam=length(g);
133     for r=1:1:tam
134         for s=1:1:tam
135             N(r,s)=abs(g(r)-g(s));
136         end
137     end
138     for r=1:1:tam
139         di(r)=0;
140         for s=1:1:tam
141             di(r)=N(r,s)+di(r);
142         end
143         di(r)=(di(r))/(tam-1);
144     end
145     for r=1:1:tam

```

```

146     for s=1:1:tam
147         P(r,s)=di(s)/di(r);
148     end
149 end
150 for r=1:1:tam
151     n=0;
152     for s=1:1:tam
153         n=n+P(s,r);
154     end
155     w(r)=1/n;
156 end
157 mcheng=0;
158 for r=1:1:tam
159     mcheng=w(r)*g(r)+mcheng;
160 end
161 mcheng_vet_c(i)=mcheng;
162 end
163 //Cálculo do "m" do Método de Cheng de D
164 for i=1:1:amostra
165     for j=1:1:n_dois
166         g(j)=D(i,j);
167     end
168     tam=length(g);
169     for r=1:1:tam
170         for s=1:1:tam
171             N(r,s)=abs(g(r)-g(s));
172         end
173     end
174     for r=1:1:tam
175         di(r)=0;
176         for s=1:1:tam
177             di(r)=N(r,s)+di(r);
178         end
179         di(r)=(di(r))/(tam-1);
180     end
181     for r=1:1:tam
182         for s=1:1:tam
183             P(r,s)=di(s)/di(r);
184         end
185     end
186     for r=1:1:tam

```

```

187     n=0;
188     for s=1:1:tam
189         n=n+P(s,r);
190     end
191     w(r)=1/n;
192 end
193 mcheng=0;
194 for r=1:1:tam
195     mcheng=w(r)*g(r)+mcheng;
196 end
197 mcheng_vet_d(i)=mcheng;
198 end
199 //Cálculo do "m" do Método de Cheng de E
200 for i=1:1:amostra
201     for j=1:1:n_tres
202         g(j)=E(i,j);
203     end
204     tam=length(g);
205     for r=1:1:tam
206         for s=1:1:tam
207             N(r,s)=abs(g(r)-g(s));
208         end
209     end
210     for r=1:1:tam
211         di(r)=0;
212         for s=1:1:tam
213             di(r)=N(r,s)+di(r);
214         end
215         di(r)=(di(r))/(tam-1);
216     end
217     for r=1:1:tam
218         for s=1:1:tam
219             P(r,s)=di(s)/di(r);
220         end
221     end
222     for r=1:1:tam
223         n=0;
224         for s=1:1:tam
225             n=n+P(s,r);
226         end
227         w(r)=1/n;

```



```
228     end
229     mcheng=0;
230     for r=1:1:tam
231         mcheng=w(r)*g(r)+mcheng;
232     end
233     mcheng_vet_e(i)=mcheng;
234 end
235 //Cálculo do "m" do Método de Cheng de F
236 for i=1:1:amostra
237     for j=1:1:n_tres
238         g(j)=F(i,j);
239     end
240     tam=length(g);
241     for r=1:1:tam
242         for s=1:1:tam
243             N(r,s)=abs(g(r)-g(s));
244         end
245     end
246     for r=1:1:tam
247         di(r)=0;
248         for s=1:1:tam
249             di(r)=N(r,s)+di(r);
250         end
251         di(r)=(di(r))/(tam-1);
252     end
253     for r=1:1:tam
254         for s=1:1:tam
255             P(r,s)=di(s)/di(r);
256         end
257     end
258     for r=1:1:tam
259         n=0;
260         for s=1:1:tam
261             n=n+P(s,r);
262         end
263         w(r)=1/n;
264     end
265     mcheng=0;
266     for r=1:1:tam
267         mcheng=w(r)*g(r)+mcheng;
268     end
```

```
269     mcheng_vet_f(i)=mcheng;
270 end
271 //Resultado de A
272 resultado_a=[media_a,mediana_a,mcheng_vet_a];
273 //Resultado de B
274 resultado_b=[media_b,mediana_b,mcheng_vet_b];
275 //Resultado de C
276 resultado_c=[media_c,mediana_c,mcheng_vet_c];
277 //Resultado de D
278 resultado_d=[media_d,mediana_d,mcheng_vet_d];
279 //Resultado de E
280 resultado_e=[media_e,mediana_e,mcheng_vet_e];
281 //Resultado de F
282 resultado_f=[media_f,mediana_f,mcheng_vet_f];
```