Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

Programa de Pós-Graduação em Estatística

Modelo de Regressão Elíptico Bivariado Intervalar

Laura Vicuña Torres de Paula

Dissertação de Mestrado

Recife
2015

Laura Vicuña Torres de Paula

Modelo de Regressão Elíptico Bivariado Intervalar

Orientador: Prof. Dr. Francisco José de Azevêdo Cysneiros
Co-Orientadora: Profa. Dra. Renata Maria Cardoso Rodrigues de Souza

Área de Concentração: Estatística Aplicada

Dissertação apresentada ao programa de Pós-graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de **Mestre em Estatística**.

Recife
2015

Catalogação na fonte
Bibliotecária Joana D'Arc Leão Salvador CRB4-532

P324m Paula, Laura Vicuña Torres de.
Modelo de regressão elíptico bivariado intervalar / Laura Vicuña Torres de Paula. – Recife: O Autor, 2015.
80 f.: fig., tab.

Orientador: Francisco José de Azevêdo Cysneiros.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, 2015.
Inclui referências.

1. Estatística aplicada. 2. Análise de regressão. I. Cysneiros, Francisco José de Azevêdo (Orientador). II. Titulo.

519.5 CDD (22. ed.) UFPE-MEI 2015-103

LAURA VICUÑA TORRES DE PAULA

MODELO DE REGRESSÃO ELÍPTICO BIVARIADO INTERVALAR

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 31 de julho de 2015.

BANCA EXAMINADORA

Prof. Dr. Francisco José de Azevêdo Cysneiros
UFPE

Prof. PhD. Getúlio José Amorim do Amaral (Examinador Interno)
UFPE

Prof.ª Dra. Roberta Andrade de Araújo Fagundes (Examinador Externo)
UPE

Dedico esse trabalho

A Deus acima de tudo
À minha mãe Engraçia, por seu apoio incondicional
Aos meus irmãos, Emanuel, Emanuela, Vivianny e Luiz Antônio.

Agradecimentos

Agradeço primeiramente a Deus, por sempre iluminar meu caminho com saúde, sabedoria, determinação e por guiar pessoas especiais para minha vida, que sempre me apoiaram nas minhas escolhas e que estão sempre comigo nos momentos bons e ruins da minha vida.

À minha mãe Engraçia, por ter me dado a oportunidade de estar aqui hoje. Pelo seu esforço para que nunca me faltasse nada, sua força, carinho e dedicação. Por seu apoio incondicional. Agradeço pelos seus sorrisos e lágrimas de alegria e satisfação que estarão sempre gravados em meu coração.

Aos meus tios Lúcia e Luiz, por todos os seus ensinamentos, carinhos e conselhos. Sou muito grata por tudo.

Ao meus orientadores Francisco Cysneiros e Renata Souza, por suas orientações e paciência na elaboração desse trabalho. Sou muito grata pelo apoio, confiança, compreensão, carinho e dedicação.

À grande amiga Roberta Fagundes, que me auxiliou bastante na elaboração desse trabalho, principalmente com os estudos de simulação. Obrigada por tudo.

Agradeço aos meus irmãos Emanuel, Emanuela, Vivianny e Luiz Antônio pelo carinho, disposição e toda força dada. Por serem meus amigos ou amigas e tudo mais que uma pessoa pode ser à outra.

Aos meus sobrinhos lindos, Ana Luzia, José Guilherme e Maria Tereza, por serem crianças adoráveis e carinhosas. Por me terem em seus corações e por me permitirem fazer parte de suas vidas.

Agradeço a todos os professores do programa de pós graduação do Departamento de Estatística, principalmente aqueles que tive o contato em sala de aula, pois pude aprender muito com vocês, não só no estudo mas como pessoa também.

Às funcionários Valeria Bittencourt e Lódino Serbim, que sempre tiveram muita paciência e simpatia, e por sempre me ajudar, quando estava ao seu alcance. Meu muito obrigada.

Aos amigos de mestrado e doutourado Allison, Carolina, Enai, Fernanda, Jennifer, Jonas, Jessica, Luana, Marcio, Pedro, Raquel, Renan (Bacaninha), Rodrigo (Din), Rodrigo (Goaino), Sébastian, Stênio, Terezinha e aos amigos cearenses Jucelino, Hildemar (Calixtinho), Rodney e Yuri. Que estiveram nessa caminhada juntos comigo, sempre me ajudando e incentivando.

As amigas Evelyne, Raphaela e Wanessa que estiveram nessa caminhada de dois anos de mestrado comigo. Agradeço por todos os momentos de alegria, alívio, tristeza, horas de estudo compartilhado e palhaçadas. Espero que a distância não seja um empecilho e que possamos continuar nos comunicando.

As minhas grandes amigas Janaína, Kelly e Raquel que mesmo na distância sempre estiveram comigo, me apoiando, incentivando e dando força com os meus sonhos e objetivos.

À FACEPE, pelo apoio financeiro.

“A educação é a arma mais poderosa que você pode usar para mudar o mundo.”

Nelson Mandela

“A educação tem raízes amargas, mas os seus frutos são doces.”

Aristóteles

Resumo

A análise de dados simbólicos (ADS) é uma abordagem estatística bastante utilizada em grandes bases de dados e tem como característica agregar dados em grupos de interesse. Esses tipos de dados podem ser representados por intervalos, conjuntos de categorias, distribuição de frequência, distribuição de probabilidade, entre outros tipos. Neste trabalho abordaremos dados simbólicos do tipo intervalo que são comumente utilizados em aplicações financeiras, mineração de dados, tráfego de redes, dados confidenciais, etc. Inicialmente, um modelo de regressão elíptico bivariado intervalar que considera a correlação entre os limites inferiores e superiores de uma variável simbólica intervalar foi proposto. Derivamos a função *escore* e a matriz de informação de *Fisher*. O método de máxima verossimilhança foi desenvolvido para estimação dos parâmetros do modelo proposto. Estudos de simulação de Monte Carlo em que avaliamos a sensibilidade do erro de previsão quanto a presença de intervalos *outliers* foram apresentados. Os resultados mostraram que o modelo $t-$Student bivariado intervalar é menos sensível na presença de intervalos *outliers* do que o modelo normal bivariado intervalar. Um conjunto de dados reais foi utilizado para ilustrar a metodologia abordada.

Palavras-chave: Análise de dados simbólicos. Intervalos *outliers*. Modelo de regressão elíptico bivariado intervalar.

Abstract

The symbolic data analysis (SDA) is a statistical approach widely used in large databases and that is characterized by aggregate data into interest groups. These data types may be represented by intervals, sets of categories, frequency distribution, probability distribution, among other types. In this paper we discuss symbolic data of interval type that are commonly used in financial applications, data mining, network traffic, confidential data, etc. First, an interval bivariate elliptical regression model that considers the correlation between the upper and lower limits of an interval symbolic variable was proposed. We derive the score function and the Fisher information matrix. The maximum likelihood method was developed to estimate the parameters of the proposed model. Monte Carlo simulation studies was performed to evaluate the sensitivity of the predictive error for the presence of outliers intervals. The results showed that the interval bivariate t-Student model is less sensitive in presence of outliers intervals than the interval bivariate normal model. A real datasets was used to illustrate the discussed methodology.

Keywords: Symbolic data analysis. Outliers intervals. Interval bivariate elliptical regression model.

Lista de Figuras

Lista de Tabelas

Sumário

CAPÍTULO 1

Introdução

Esse capítulo descreve uma breve fundamentação sobre a análise de dados simbólicos, a motivação e os objetivos em relação ao trabalho proposto e por fim, descrever a estrutura da dissertação.

1.1 Motivação

Nos dados clássicos, as variáveis são classificadas como numérica e/ou categórica e estas podem assumir um único valor para cada indivíduo. Como por exemplo, verificar idade, nível de instrução e renda dos clientes em um determinado banco. Entretanto, pode ocorrer do pesquisador não ter interesse em verificar as observações individuais, mas sim grupos de unidades com características em comum ou observar informações ao longo do tempo ou em outras circunstâncias. Por exemplo, definir um grupo de pessoas pela idade, altura, nível de instrução, instituição de ensino, ou outros. Quando isso acontece, a abordagem clássica torna-se um pouco restrita por não considerar a variabilidade inerente dos dados.

Para lidar com tal situação é usual reduzir os dados considerando medidas de tendência central, como a média, a mediana ou a moda, consequentemente isso levaria a uma perda de informação. Uma outra forma de lidar com esse tipo de problema seria utilizar a análise de dados simbólicos (ADS), que possibilita agregar esses dados em grupos de interesse. Conforme Fagundes (2013), ao usar ADS é possível estudar os grupos considerando a descrição de grupos de indivíduos e as variações dentro desses grupos.

Os dados simbólicos foram tratados inicialmente por Edwin Diday no fim da década de 80 e as primeiras pesquisas sobre esse tipo de dados continham os princípios básicos dessa nova abordagem, Diday (1988), Diday (1989) e Diday & Brito (1989). Esses dados tem como característica expressar intervalos, conjuntos, distribuição de frequências e

distribuição de probabilidades.

Ao longo dos anos Billard & Diday (2003) atentou-se para o elevado crescimento dos dados simbólicos e notou a necessidade de ampliar novas técnicas para tratar esses tipos de dados. A partir disso, Billard & Diday (2006a) e Diday & Noirhomme-Fraiture (2008) introduziram novos conceitos e métodos estatísticos capaz de manusear os dados simbólicos. Atualmente esses métodos são encontrados na literatura, como por exemplo análise de agrupamento para dados simbólicos, análise exploratória, análise de regressão, entre outros.

No contexto de análise de regressão para dados intervalares, Billard & Diday (2000) estende o conceito o modelo de regressão linear clássico (MRLC) para dados simbólicos do tipo intervalo, que utiliza o método do mínimos quadrados para obter as estimativas dos valores médios dos intervalos. Billard & Diday (2002) define uma nova abordagem que considera dois MRLC independentes, um para o limite inferior e outra para o limite superior dos dados intervalares. Neto & de Carvalho (2008) propuseram o método do centro e da amplitude para dados simbólicos intervalares sendo que esta nova representação obteve um comportamento de predição melhor que os propostos por Billard & Diday (2000) e Billard & Diday (2002).

Os trabalhos citados anteriormente não garantem a coerência matemática de que o valor previsto do limite inferior seja menor que o do limite superior. Dessa forma, Neto & de Carvalho (2010) propuseram um novo método de ajustar modelos de regressão para dados simbólicos intervalares, em que incorpora restrições nos valores da amplitude, com a finalidade de assegurar a coerência matemática.

Os primeiros estudos sobre regressão simbólica intervalar não consideravam suposições de distribucionais para os erros. Domingues et al. (2010) propuseram um modelo de regressão para dados intervalares que assumem distribuições de probabilidade simétrica para os erros. Souza et al. (2011) propuseram um modelo de regressão linear logística para dados simbólicos intervalares. Fagundes et al. (2013) propuseram um modelo de regressão robusto para dados simbólicos intervalares em que esse modelo apresentou ser menos sensível a presença de *outliers*. No contexto de regressão não paramétrica, Fagundes et al. (2014) propuseram um modelo regressão *Kernel* intervalar que pode ser ajustado quando as suposições distribucionais dos erros e/ou da forma funcional dos modelos paramétricos não são verificadas.

Todos os trabalhos citados acima, consideram dois modelos de regressão independentes para realizar a análise de regressão intervalar. Entretanto, essa relação de independência pode influenciar de alguma forma o ajustamento. Dessa forma, Neto et al. (2011) propuseram um modelo de regressão bivariado da família exponencial para dados simbólicos intervalar que modela a correlação entre limites inferiores e superiores do intervalo. Contudo, esse modelo de regressão é sensível a presença de intervalos aberrantes.

1.2 Objetivos

O principal objetivo desse trabalho é propor um modelo de regressão elíptico bivariado intervalar que considera os erros do modelo como pertencentes a família de distribuições elípticas.

Este modelo tem como principal característica considerar uma possível relação entre limite inferior e superior da variável simbólica intervalar. Além, de assegurar a coerência matemática de que os valores previstos do limite inferior sejam menores que a do limite superior, ao ajustar o modelo considerando a representação do centro e amplitude dos intervalos.

Um outro objetivo desse trabalho é avaliar a sensibilidade do erro de previsão quanto a presença de intervalos *outliers* no dados.

1.3 Estrutura da dissertação

O presente trabalho é constituído por seis capítulos. No Capítulo 1, uma breve introdução a respeito da motivação e objetivos da dissertação é discutida. No Capítulo 2 são apresentados os conceitos de análise de dados simbólicos, os respectivos tipos de variáveis e as medidas descritivas para variáveis intervalares. No Capítulo 3 introduzimos o modelo de regressão elíptico bivariado intervalar (MREBI) e sua representação centro e amplitude, assim como a função *escore*, matriz de informação de *Fisher* e estimação dos parâmetros. No Capítulo 4 são apresentados estudos de simulações para duas representações estudadas em que verificamos a qualidade de previsão na presença ou não de *outliers*. No Capítulo 5 discutimos a modelagem de um conjunto de dados reais em que a metodologia apresentada é utilizada. Por fim, no Capítulo 6, as principais conclusões desse trabalho são discutidas.

CAPÍTULO 2

Dados simbólicos

Esse capítulo tem como finalidade explanar sobre as principais características de dados simbólicos, as áreas de aplicações, os tipos de variáveis para dados dessa natureza e descrever a estatística descritiva para dados simbólicos intervalares.

2.1 Análise de dados simbólicos (ADS)

Os dados simbólicos são capazes de descrever indivíduos levando em consideração, ou não, imprecisão e incerteza e são classificados como dados mais complexos do que os dados clássicos, visto que eles possuem uma estrutura interna.

Ao analisar dados dessa natureza é possível descrever grupos de indivíduos e a variação dentro desses grupos considerando intervalos, histograma, distribuição de probabilidade e frequência. Eles ainda podem ser usados para representar os limites de um conjunto de possíveis valores de um item ou a variação da extensão de uma variável através da redução de conjuntos de dados, em um número reduzido de pequenos grupos de informação (Domingues et al., 2010). Contudo, essa característica de diminuir o tamanho de grandes conjuntos de dados pode acarretar em um problema de perda de precisão e/ou variabilidade.

Atualmente, a análise de dados dessa natureza vem sendo bastante estudada, uma vez que os trabalhos Diday (1988) e Diday & Brito (1989) tiveram grande impacto e os avanços tecnológicos tiveram um crescimento elevado nos últimos anos. A ADS tem grande relação com a análise multivariada, banco de dados, inteligência artificial, regressão, classificação, entre outros. Nos últimos anos diversos estudos com dados dessa natureza foram propostos e podemos citar como exemplos: Carvalho (1995) propôs a construção de histogramas para ADS; Gordon (2000) propôs um algoritmo iterativo de agrupamento para dados simbólicos que minimiza a soma potencial das descrições dos grupos; Billard &

Diday (2000) propuseram um modelo de regressão para dados intervalares baseado no conceito do modelo de regressão linear clássico (MRLC); Billard & Diday (2006a) propuseram novos conceitos sobre ADS tais conceitos abordam análise descritiva, componente principal, *clustering* e regressão; Carvalho et al. (2007) propuseram agrupamento de dados simbólicos intervalares baseado na distância Hausdorff adaptada; Maia et al. (2008) propuseram um modelo de séries temporais que investiga a previsão de dados simbólicos intervalares; Neto & de Carvalho (2008) propuseram um modelo com nova representação para dados intervalares baseado no centro e na amplitude dos dados intervalares e Carvalho & de Souza (2010) propuseram um método de particionamento para dados simbólicos do tipo de recurso misto usando a euclidiana ao quadrado.

2.2 Tipos de variáveis simbólicas

Variáveis são classificadas nos estudos estatísticos como sendo os valores que assumem determinadas características dentro de uma pesquisa.

Assim, como nos dados clássicos existe uma classificação dos tipos de variáveis para os dados simbólicos. Segundo Bock & Diday (2000), as variáveis simbólicas de dividem em dois grupos: variáveis modais e não modais. Esses tipos de variáveis são definidas com mais detalhes nas próximas seções.

2.2.1 Variável do tipo modal

Define-se Y como sendo uma variável simbólica modal se ela descrever um objeto usando par(c, π), em que c é o conjunto de categorias que a variável pode assumir e π é um vetor de frequência, pesos ou probabilidades que correspondem a cada categoria do conjunto c.

Exemplo: Seja Y os cursos superiores com alunos reprovados na disciplina de estatística em k universidades. Tem-se que para uma determinada universidade t, $Y(t) = \{$(Farmácia, Ciências Contábeis, Administração, Eng. Mecânica); $(0,4; 0,25; 0,25; 0,1)\}$, ou seja, a probabilidade dos alunos do curso de Eng. Mecânica reprovar a disciplina de estatística nessa universidade é de $0,1$, Administração e Ciências Contábeis é de $0,25$ e Farmácia $0,4$. Já em uma universidade u o objeto par é definido como: $Y(u) = \{$(Matemática, Fisíca, Eng. Alimentos, Farmácia); $(0,1; 0,1; 0,3; 0,5)\}$.

2.2.2 Variável do tipo não modal

As variáveis não modais são classificadas como do tipo multivalorada e intervalares, em que a variável multivalorada é dividida em categórica nominal, categórica ordinal e quantitativa. As definições dessas variáveis são dadas a seguir

- **Variável multivalorada nominal:** descreve as categorias de um determinado objeto, em que nessas categorias não existe uma ordenação. **Exemplo:** seja Y o nome da companhia aérea de um determinado grupo k de passageiros, então Y_k = Avianca, Gol, Tam;

- **Variável multivalorada ordinal:** descreve as categorias de um determinado objeto, em que nessas categorias existe uma ordenação. **Exemplo:** seja Y o nível de escolaridade de um grupo k de indivíduos de uma empresa, então Y_k = Ensino Fundamental, Ensino Médio, Graduação, Pós-Graduação;

- **Variável multivalorada quantitativa:** assume um conjunto finito de números reais não ordenados. **Exemplo:** seja Y o número de apartamentos vendidos por uma determinada empresa k em um mês específico, então $Y(k) = \{15, 35, 20, 50, 45\}$;

- **Variável intervalar:** Essa variável tem como característica está definida em um intervalo. Uma variável Y será definida como variável intervalar se para todo $k \in E$ o subconjunto $Y(k) = [a, b]$, em que $a \leq b$. **Exemplo:** Uma empresa de pesquisas está interessada em saber o gasto com impostos de um conjunto de empresas, onde E representa o conjunto de empresas, Y o valor gasto com impostos e k uma empresa qualquer. Assim o $Y(k) = [1500,00; 4500,00]$, ou seja, uma empresa k teve gastos de impostos em um intervalo de R$ $1500,00$ a R$ $4500,00$. Conforme Campos (2008), ao definir Y como sendo uma variável intervalar seria interessante criar histograma afim de visualizar a distribuição de frequências dos valores. Esse tipo de variável é bastante utilizada em aplicações financeiras, mineração de dados, análise de tráfego de redes, aplicações com dados confidencias em que o interesse é apenas conhecer a extensão dos valores, dentre outras. Esse tipo de variável é o objeto deste estudo.

2.2.3 Exemplo de tabelas para dados simbólicos

Dados dessa natureza podem ser apresentados em tabelas, nestas as linhas correspondem aos indivíduos ou grupos de indivíduos e as colunas são as variáveis simbólicas que os descrevem.

Exemplo: Nove países foram divididos em três grupos, a partir desses foram observados na Tabela 2.1 os valores do intervalo do PIB (Produto Interno Bruto) em milhões de dólares e a proporção da população que fala mais de uma língua. Note que a variável PIB é classificada como simbólica intervalar, a variável nome dos países como simbólica multivalorada nominal e a variável mais de uma língua é considerada simbólica modal (representado por uma distribuição de pesos).

Tabela 2.1: Exemplo de Tabela para dados simbólicos

Grupo	PIB	Nome dos países	+ de uma língua
A	[0,31;4,28]	Bélgica, Japão e Suíça	(3/4) Sim, (1/4) Não
B	[0,37;1,98]	Argentina, Brasil e Uruguai	(1/6) Sim, (5/6) Não
C	[0,16;14,02]	EUA, França e Chile	(2/5) Sim, (3/5) Não

Fonte: Salazar (2008)

2.3 Análise descritiva de variáveis simbólicas intervalares

O conceito de estatística descritiva para dados simbólicos é similar aos dos dados clássicos. Carvalho (1994), Carvalho (1995) e Chouakria et al. (1998) propõem métodos para construir histograma para variáveis intervalares. Já Bertrand & Goupil (2000) introduziram métodos para determinar a distribuição de frequência para uma variável simbólica e ampliaram os conceitos de média, mediana e desvio padrão. Em Billard & Diday (2006b) são abordados média, mediana, variância e histograma para dados do tipo não modais. Dessa forma, essa seção é designada a obtenção de média, variância e histograma para variáveis intervalares, visto que o objetivo desse estudo é trabalhar com variáveis simbólicas intervalares. Assim, foi utilizado o conjunto de dados Futebol nessa seção para ilustrar essas medidas.

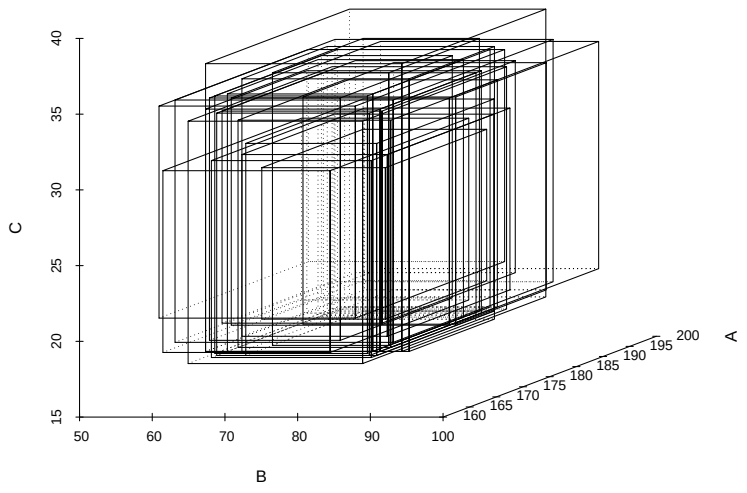
Conjunto de dados Futebol: é composto pelas informações dos jogadores profissionais de 20 times da França, em que cada jogador é descrito pelas variáveis altura, idade e peso. Esse conjunto já foi analisado em Fagundes (2013) e pode ser obtido em `https://www.ceremade.dauphine.fr/touati/foot2.htm`. Esses dados são vistos na Tabela 2.2 e representam os valores intervalares dos times Franceses em cada uma das três variáveis.

A Figura 2.1 interpreta os retângulos dos intervalos das variáveis do conjunto Futebol. Nela é possível notar um comportamento homogêneo dos quadrados e que não há indícios de possíveis *outliers*.

Tabela 2.2: Dados Futebol

Time	Peso (Y)	Altura (X_1)	Idade (X_2)
1	[58,00;85,00]	[164,00;192,00]	[21,00;35,00]
2	[67,00;84,00]	[171,00;190,00]	[20,00;30,00]
3	[65,00;88,00]	[170,00;186,00]	[18,00;36,00]
4	[60,00;83,00]	[162,00;188,00]	[19,00;31,00]
5	[60,00;84,00]	[170,00;189,00]	[18,00;34,00]
6	[67,00;83,00]	[173,00;190,00]	[18,00;36,00]
7	[69,00;90,00]	[176,00;193,00]	[19,00;34,00]
8	[65,00;85,00]	[170,00;193,00]	[19,00;31,00]
9	[63,00;84,00]	[168,00;188,00]	[18,00;34,00]
10	[58,00;88,00]	[167,00;197,00]	[19,00;35,00]
11	[62,00;86,00]	[164,00;191,00]	[18,00;34,00]
12	[62,00;80,00]	[168,00;189,00]	[19,00;35,00]
13	[63,00;85,00]	[167,00;190,00]	[18,00;31,00]
14	[65,00;95,00]	[168,00;196,00]	[20,00;35,00]
15	[63,00;83,00]	[170,00;187,00]	[18,00;35,00]
16	[60,00;87,00]	[170,00;197,00]	[18,00;37,00]
17	[67,00;85,00]	[168,00;190,00]	[18,00;32,00]
18	[62,00;83,00]	[169,00;192,00]	[18,00;35,00]
19	[63,00;84,00]	[172,00;192,00]	[18,00;33,00]
20	[63,00;85,00]	[169,00;194,00]	[20,00;34,00]

Figura 2.1: Gráfico 3D entre as variáveis Peso (B), Altura (A) e Idade (C)



Conforme Billard & Diday (2006b), para obter os valores descritivos das variáveis intervalares desse conjunto de dados é importante definir inicialmente os conceitos de descrição individual e virtual.

2.3.1 Descrição individual e virtual

Descrição individual é um valor de uma variável que pode assumir muitos outros valores, essa descrição individual é denotada por x. Com relação a variáveis intervalares é um intervalo que irá pertencer a um intervalo principal. **Exemplo:** Considerando (T, W) variável intervalar bidimensional, onde $T = [58,00; 85,00]$ representa o peso de um time e $W = [164,00; 192,00]$ representa altura de um time do conjunto Futebol. Então, a descrição individual $x = (x_1, x_2)$ estão contido no retângulo $[58,00; 85,00] \times [164,00; 192,00]$.

De acordo com Billard & Diday (2006b), esses valores são utilizados para obter os histogramas simbólicos. Já que as frequências desse histograma envolve a contagem do número de descrições individuais que tornam verdadeira uma determinada dependência lógica nos dados.

A dependência lógica pode ser representada pela equação (2.1), onde $x \in X$ (X é o conjunto de todas as descrições individuais possíveis presente em uma tabela simbólica) e $A \subseteq D, B \subseteq D$. De acordo com Billard & Diday (2006b), υ retornará um valor binário, ou seja, "0" caso a dependência lógica para x for falsa, ou "1" caso seja verdadeiro.

$$\upsilon : [x \in A] \Rightarrow [x \in B]. \tag{2.1}$$

Descrição virtual de um vetor d é um conjunto de todos os elementos x presentes que satisfazem todas as dependências lógicas em X. Essa descrição é dada pela equação (2.2)

$$vir(d) = x \in D; v(x) = 1, \forall \in V_x. \tag{2.2}$$

Suponha que haja interesse em uma variável $Y_j \equiv Z$ e que o valor observado de um objeto w nessa variável é um intervalo $Z(w) = [a_w, b_w]$, para $w \in E = \{1, \ldots, m\}$, os valores dos vetores de descrição individuais $x \in vir(d_w)$ são distribuídos uniformemente sobre o intervalo $Z(w)$. Portanto, para cada ξ tem-se que:

$$P[x \leq \xi | x \in vir(d_w)] = \begin{cases} 0, & \text{se } \xi \leq a_w; \\ \frac{\xi - a_w}{b_w - a_w}, & \text{se } a_w \leq \xi \leq b_w; \\ 1, & \text{caso contrário} \end{cases} \tag{2.3}$$

O vetor de descrição individual x vai contar com os valores globalmente em $\bigcup_{w \in E} vir(d)_w$, em que cada um desses objetos vai contar com a mesma probabilidade ($p = 1/m$) de ser considerado no experimento. Ao considerarmos as variáveis do conjunto de Futebol, temos que $m = 20$, $w = 1 \ldots, 20$ e $Z_Y(w)$, $Z_{X_1}(w)$ e $Z_{X_2}(w)$ correspondem ao intervalo w de suas respectivas variáveis nesse conjunto, que podem ser observadas na Tabela 2.2.

2.3.2 Funções de distribuição e de densidade empírica para intervalos

A função de distribuição empírica é obtida a partir de um determinado experimento. Para os dados intervalares a função $F_Z(\xi)$ é composta pela distribuição das m distribuições uniformes nos m intervalos $Z(w) = [a_w, b_w]$, em que $w \in E$. Assim, a função empírica intervalar é dada por:

$$\begin{aligned} F_Z(\xi) &= \frac{1}{m}\sum_{w\in E} P[x \leq \xi | x \in vir(d_w)] \\ &= \frac{1}{m}\left\{\sum_{\xi\in Z(w)}\left(\frac{\xi - a_w}{b_w - a_w}\right) + |(w|\xi \geq b_w)|\right\} \end{aligned} \tag{2.4}$$

Exemplo: No conjunto de dados Futebol ao fazer suposições de um $\xi = 76,5$ para variável peso, $\xi = 179,5$ para altura e $\xi = 27,5$ para idade, temos que as funções empíricas dessas variáveis são expressadas por:

$$\begin{aligned} F_Y(76,5) &= \frac{1}{20}\left(\frac{76,5-58}{85-58} + \frac{76,5-67}{84-67} + \frac{76,5-65}{88-65} + \frac{76,5-60}{83-60} + \right. \\ &\quad + \frac{76,5-60}{84-60} + \frac{76,5-67}{83-67} + \frac{76,5-69}{90-69} + \frac{76,5-65}{85-65} + \\ &\quad + \frac{76,5-63}{84-63} + \frac{76,5-58}{88-58} + \frac{76,5-62}{86-62} + \frac{76,5-62}{80-62} + \\ &\quad + \frac{76,5-63}{85-63} + \frac{76,5-65}{95-65} + \frac{76,5-63}{83-63} + \frac{76,5-60}{87-60} + \\ &\quad \left. + \frac{76,5-67}{85-67} + \frac{76,5-62}{83-62} + \frac{76,5-63}{84-63} + \frac{76,5-63}{85-63}\right) \\ &= \frac{1}{20}(12,1018) = 0,6050, \end{aligned}$$

$$\begin{aligned} F_{X_1}(179,5) &= \frac{1}{20}\left(\frac{179,5-164}{192-164} + \frac{179,5-171}{190-171} + \frac{179,5-170}{186-170} + \frac{179,5-162}{188-162} + \right. \\ &\quad + \frac{179,5-170}{189-170} + \frac{179,5-173}{190-173} + \frac{179,5-176}{193-176} + \frac{179,5-170}{193-170} + \\ &\quad + \frac{179,5-168}{188-168} + \frac{179,5-167}{197-167} + \frac{179,5-164}{191-164} + \frac{179,5-168}{189-168} + \\ &\quad + \frac{179,5-167}{190-167} + \frac{179,5-168}{196-168} + \frac{179,5-170}{187-170} + \frac{179,5-170}{197-170} + \\ &\quad \left. + \frac{179,5-168}{190-168} + \frac{179,5-169}{192-169} + \frac{179,5-172}{192-172} + \frac{179,5-169}{194-169}\right) \\ &= \frac{1}{20}(9,5215) = 0,4760, \end{aligned}$$

$$
\begin{aligned}
F_{X_2}(27,5) &= \frac{1}{20}\left(\frac{27,5-21}{35-21}+\frac{27,5-20}{30-20}+\frac{27,5-18}{36-18}+\frac{27,5-19}{31-19}+\right.\\
&\quad \frac{27,5-18}{34-18}+\frac{27,5-18}{36-18}+\frac{27,5-19}{34-19}+\frac{27,5-19}{31-19}+\\
&\quad \frac{27,5-18}{34-18}+\frac{27,5-19}{35-19}+\frac{27,5-18}{34-18}+\frac{27,5-19}{35-19}+\\
&\quad \frac{27,5-18}{31-18}+\frac{27,5-20}{35-20}+\frac{27,5-18}{35-18}+\frac{27,5-18}{37-18}+\\
&\quad \left.\frac{27,5-18}{32-18}+\frac{27,5-18}{35-18}+\frac{27,5-18}{33-18}+\frac{27,5-20}{34-20}\right)\\
&= \frac{1}{20}(11,7929) = 0,5896.
\end{aligned}
$$

A função de densidade empírica existirá se a equação (2.4) poder ser derivada em relação a ξ, em que esta pode ser expressada por:

$$f(\xi) = \frac{1}{m}\sum_{w:\xi\in Z(w)}\frac{1}{b_w - a_w}. \tag{2.5}$$

A equação (2.5) pode ser reescrita de uma outra forma, pois o somatório lida apenas com os objetos w e para $\xi \in Z(w)$. Assim, é possível obter a seguinte forma:

$$f(\xi) = \frac{1}{m}\sum_{w\in E}\frac{I_w(\xi)}{||Z(w)||}, \xi \in \Re, \tag{2.6}$$

em que $I_w(\xi)$ é uma função indicadora e indica se ξ está ou não em $Z(w)$. Então, caso $\xi \in Z(w)$ retornará valor 1 e 0 caso contrário. Já $||Z(w)||$ representa a amplitude do intervalo $Z(w) \in E$, onde $||Z(w)|| = b_w - a_w$.

Exemplo: Utilizando as suposições de que os valores de ξ são equivalentes ao da função de distribuição, tem-se que a função de densidade empírica das variáveis do conjunto de Futebol são definidas por:

$$
\begin{aligned}
f_Y(76,5) &= \frac{1}{20}\left(\frac{1}{85-58}+\frac{1}{84-67}+\frac{1}{88-65}+\frac{1}{83-60}+\right.\\
&\quad +\frac{1}{84-60}+\frac{1}{83-67}+\frac{1}{90-69}+\frac{1}{85-65}+\\
&\quad +\frac{1}{84-63}+\frac{1}{88-58}+\frac{1}{86-62}+\frac{1}{80-62}+\\
&\quad +\frac{1}{85-63}+\frac{1}{95-65}+\frac{1}{83-63}+\frac{1}{87-60}+\\
&\quad \left.+\frac{1}{85-67}+\frac{1}{83-62}+\frac{1}{84-63}+\frac{1}{85-63}\right)\\
&= \frac{1}{20}(0,9248) = 0,0462,
\end{aligned}
$$

$$\begin{aligned} f_{X_1}(179,5) &= \frac{1}{20}\left(\frac{1}{192-164}+\frac{1}{190-171}+\frac{1}{186-170}+\frac{1}{188-162}+\right. \\ &+\frac{1}{189-170}+\frac{1}{190-173}+\frac{1}{193-176}+\frac{1}{193-170}+ \\ &+\frac{1}{188-168}+\frac{1}{197-167}+\frac{1}{191-164}+\frac{1}{189-168}+ \\ &+\frac{1}{190-167}+\frac{1}{196-168}+\frac{1}{187-170}+\frac{1}{197-170}+ \\ &\left.+\frac{1}{190-168}+\frac{1}{192-169}+\frac{1}{192-172}+\frac{1}{194-169}\right) \\ &= \frac{1}{20}(0,9250) = 0,0462, \end{aligned}$$

$$\begin{aligned} f_{X_2}(27,5) &= \frac{1}{20}\left(\frac{1}{35-21}+\frac{1}{30-20}+\frac{1}{36-18}+\frac{1}{31-19}+\right. \\ &\frac{1}{34-18}+\frac{1}{36-18}+\frac{1}{34-19}+\frac{1}{31-19}+ \\ &\frac{1}{34-18}+\frac{1}{35-19}+\frac{1}{34-18}+\frac{1}{35-19}+ \\ &\frac{1}{31-18}+\frac{1}{35-20}+\frac{1}{35-18}+\frac{1}{37-18}+ \\ &\left.\frac{1}{32-18}+\frac{1}{35-18}+\frac{1}{33-18}+\frac{1}{34-20}\right) \\ &= \frac{1}{20}(1,3517) = 0,0675. \end{aligned}$$

2.3.3 Média e Variância intervalar

Como já se conhece a densidade empírica de uma variável intervalar, é possível encontrar o valor médio dessa variável. Dessa forma, a média empírica de $\overline{Z}$ em termos da função de densidade empírica é dada por:

$$\overline{Z} = \int_{-\infty}^{\infty} \xi f(\xi) d\xi.$$

Substituindo a equação (2.6), obtém-se o seguinte resultado:

$$\begin{aligned} \overline{Z} &= \frac{1}{m}\sum_{w\in E}\int_{-\infty}^{\infty}\frac{I_w(\xi)}{||Z(w)||}\xi d\xi = \frac{1}{m}\sum_{w\in E}\frac{1}{b_w-a_w}\int_{\xi\in Z(w)}\xi d\xi \\ &= \frac{1}{2m}\sum_{w\in E}\frac{b_w^2-a_w^2}{b_w-a_w} = \frac{1}{m}\sum_{w\in E}\frac{b_w+a_w}{2}. \end{aligned} \tag{2.7}$$

De maneira similar ao cálculo da média empírica é possível encontrar a variância amostral intervalar dos dados em termos da função de densidade empírica. Portanto, a variância amostral da variável é definida por:

$$S^2 = \int_{-\infty}^{\infty} (\xi - \overline{Z})^2 f(\xi) d\xi$$

ao calcular a equação acima, está é equivalente a seguinte equação

$$S^2 = \int_{-\infty}^{\infty} \xi^2 f(\xi) d\xi - (\overline{Z})^2.$$

Note que a primeira parcela da variância amostral corresponde ao segundo momento da função de interesse. Calculando o segundo momento da função, obtém-se o seguinte resultado:

$$\begin{aligned} M_2 &= \int_{-\infty}^{\infty} \xi^2 f(\xi) d\xi = \frac{1}{m} \sum_{w \in E} \int_{-\infty}^{\infty} \frac{\xi^2}{||Z(w)||} d\xi \\ &= \frac{1}{m} \sum_{w \in E} \frac{b_w^3 - a_w^3}{3||Z(w)||} \\ &= \frac{1}{3m} \sum_{w \in E} (b_w^2 + b_w a_w + a_w^2). \end{aligned} \tag{2.8}$$

Dessa forma, a variância amostral intervalar é dada pela seguinte forma:

$$\begin{aligned} S^2 &= M_2 - \overline{Z}^2 \\ &= \frac{1}{3m} \sum_{w \in E} (b_w^2 + b_w a_w + a_w^2) - \frac{1}{4m^2} \left(\sum_{w \in E} b_w + a_w \right)^2 \end{aligned} \tag{2.9}$$

em que $\overline{Z}$ e M_2 foram definidos em (2.7) e (2.8).

Exemplo: Com base nos conceitos definidos sobre média e variância intervalar, tem-se que para as variáveis Peso (Y), Altura (X_1) e Idade (X_2) são descritas por:

$$E(Y) = \frac{1}{20} \left(\frac{85 + 58}{2} + \ldots + \frac{85 + 63}{2} \right) = 74,2250,$$

$$S_Y^2 = \frac{1}{20} \left(\frac{85^2 + (85)(58) + 58^2}{3} + \ldots + \frac{85^2 + (85)(63) + 63^2}{3} \right) - 74,2250^2 = 47,9827,$$

$$E(X_1) = \frac{1}{20} \left(\frac{192 + 164}{2} + \ldots + \frac{194 + 169}{2} \right) = 180,$$

$$S^2_{X_1} = \frac{1}{20}\left(\frac{192^2 + (192)(164) + 164^2}{3} + \ldots + \frac{194^2 + (194)(169) + 169^2}{3}\right) - 180^2 = 48,4166,$$

$$E(X_2) = \frac{1}{20}\left(\frac{35+21}{2} + \ldots + \frac{34+20}{2}\right) = 26,275,$$

$$S^2_{X_2} = \frac{1}{20}\left(\frac{35^2 + (35)(21) + 21^2}{3} + \ldots + \frac{34^2 + (34)(20) + 20^2}{3}\right) - 26,275^2 = 20,491,$$

A média intervalar dos 20 times franceses para variável peso é de $74,225kg$, para variável altura é $180cm$ e para variável idade é $26,275$ anos. Já os valores das variâncias foram respectivamente de $47,9827$, $48,4166$ e $26,275$ para as variáveis peso, altura e idade.

2.3.4 Histograma intervalar

Ao construir um histograma para uma variável simbólica do tipo intervalar, é importante considerar $I = [min_{a_w|w\in E}; max_{b_w|w\in E}]$ em que todos os valores possíveis de Z estão contidos em U, e também particiona-los em r subintervalos $I_g = [\xi_{g-1}; \xi_g)$, $g = 1, \ldots, r-1$ e $I_r = [\xi_{r-1}; \xi_r]$ com g = r. Portanto, o histograma para Z será dada pela representação gráfica da distribuição de frequência $\{(I_g, p_g)g = 1, \ldots, r\}$ onde:

$$p_g = \frac{1}{m}\sum_{w\in E}\frac{||Z(w) \cap I_g||}{||Z(w)||}. \tag{2.10}$$

Nesse caso p_g, para $g = 1, \ldots, r$, representa a área da barra vertical da base de qual é o intervalo de I_g pertencente ao eixo horizontal do histograma, Billard & Diday (2006b). Dessa maneira, p_g nada mais é que a probabilidade de uma descrição individual w está no intervalo I_g.

Utilizando as variáveis intervalares do conjunto de dados Futebol, tem-se que o intervalo global I das variáveis Peso, Altura e Idade são respectivamente, $[58; 95]$, $[162; 197]$ e $[18; 37]$. Para produzir os histogramas intervalares dessas variáveis, é de suma importância definir o número classes a ser utilizada e o intervalo de cada classe. A fim de obter o número de classes intervalares, será utilizado a definição de Sturges (1926) dado por:

$$K = log_2(n) + 1$$

onde K representa o número de classes e n o número de intervalos em uma determinada variável. Então, ao fazer uma aproximação do valor de K, cada variável irá conter 6 classes de intervalos nos seus respectivos histogramas. Os intervalos de cada classe são obtidos a partir do intervalo global da variável e do tamanho da amplitude de cada classe, a amplitude de cada classe é obtida por $\frac{I_{max} - I_{min}}{K}$. As classes do histograma dessas variáveis são definidas na Tabela 2.3.

Após definir as classes de cada variável, tem-se que a frequência relativa da primeira

Tabela 2.3: Intervalo das classes das variáveis do conjunto Futebol

Peso (Y)	Altura (X_1)	Idade (X_2)
$I_1 = [58,0; 64,2)$	$I_1 = [162,00; 167,85)$	$I_1 = [18,00; 21,15)$
$I_2 = [64,2; 70,4)$	$I_2 = [167,85; 173,70)$	$I_2 = [21,15; 24,30)$
$I_3 = [70,4; 76,6)$	$I_3 = [173,70; 179,55)$	$I_3 = [24,30; 27,45)$
$I_4 = [76,6; 82,8)$	$I_4 = [179,55; 185,40)$	$I_4 = [27,45; 30,60)$
$I_5 = [82,8; 89,0)$	$I_5 = [185,40; 191,25)$	$I_5 = [30,60; 33,75)$
$I_6 = [89,0; 95,0]$	$I_6 = [191,25; 197,00]$	$I_6 = [33,75; 37,00]$

classe da variável Peso é representada por:

$$\begin{aligned} p_1 &= \frac{1}{20}\left[\left(\frac{64,2-58}{85-58}\right)_{w=1} + 0 + 0 + \left(\frac{64,2-60}{83-60}\right)_{w=4} + \left(\frac{64,2-60}{84-60}\right)_{w=5} + \right. \\ &\quad + 0 + \ldots 0 + \left(\frac{64,2-63}{84-63}\right)_{w=9} + \left(\frac{64,2-58}{88-58}\right)_{w=10} + \left(\frac{64,2-62}{86-62}\right)_{w=11} + \\ &\quad + \left(\frac{64,2-62}{80-62}\right)_{w=12} + \left(\frac{64,2-63}{85-63}\right)_{w=13} + 0 + \left(\frac{64,2-63}{83-63}\right)_{w=15} + \\ &\quad + \left(\frac{64,2-60}{87-60}\right)_{w=16} + 0 + \left(\frac{64,2-62}{83-62}\right)_{w=18} + \left(\frac{64,2-63}{84-63}\right)_{w=19} + \\ &\quad \left. + \left(\frac{64,2-63}{85-63}\right)_{w=20}\right] = \frac{1}{20}(1,5514) = 0,0775. \end{aligned}$$

A frequência relativa da primeira classe da variável Altura é dada por:

$$\begin{aligned} p_1 &= \frac{1}{20}\left[\left(\frac{164-167,85}{192-164}\right)_{w=1} + 0 + 0 + \left(\frac{162-167,85}{188-162}\right)_{w=4} + 0 + \ldots + \right. \\ &\quad + 0 + \left(\frac{167-167,85}{197-167}\right)_{w=10} + \left(\frac{167,85-164}{191-164}\right)_{w=11} + \\ &\quad \left. + 0 + \left(\frac{167,85-167}{190-167}\right)_{w=13} + 0 + \ldots\right] \\ &= \frac{1}{20}(0,5703) = 0,0285. \end{aligned}$$

Já a frequência relativa da primeira classe da variável Idade é definida por:

$$\begin{aligned} p_1 &= \frac{1}{20}\left[\left(\frac{21,15-21}{35-21}\right)_{w=1} + \left(\frac{21,15-20}{30-20}\right)_{w=2} + \left(\frac{21,15-18}{36-18}\right)_{w=3} + \right. \\ &+ \left(\frac{21,15-19}{31-19}\right)_{w=4} + \left(\frac{21,15-18}{34-18}\right)_{w=5} + \left(\frac{21,15-18}{36-18}\right)_{w=6} + \\ &+ \left(\frac{21,15-19}{34-19}\right)_{w=7} + \left(\frac{21.15-19}{31-19}\right)_{w=8} + \left(\frac{21,15-19}{34-18}\right)_{w=9} + \\ &+ \left(\frac{21,15-19}{35-19}\right)_{w=10} + \left(\frac{21,15-18}{34-18}\right)_{w=11} + \left(\frac{21,15-19}{35-19}\right)_{w=12} + \\ &+ \left(\frac{21,15-18}{31-18}\right)_{w=13} + \left(\frac{21,15-20}{35-20}\right)_{w=14} + \left(\frac{21,15-18}{35-18}\right)_{w=15} + \\ &+ \left(\frac{21,15-18}{37-18}\right)_{w=16} + \left(\frac{21,15-18}{32-18}\right)_{w=17} + \left(\frac{21,15-18}{35-18}\right)_{w=18} + \\ &+ \left.\left(\frac{21,15-18}{33-18}\right)_{w=19} + \left(\frac{21,15-20}{34-20}\right)_{w=20}\right] \\ &= \frac{1}{20}(3,2092) = 0,1604. \end{aligned}$$

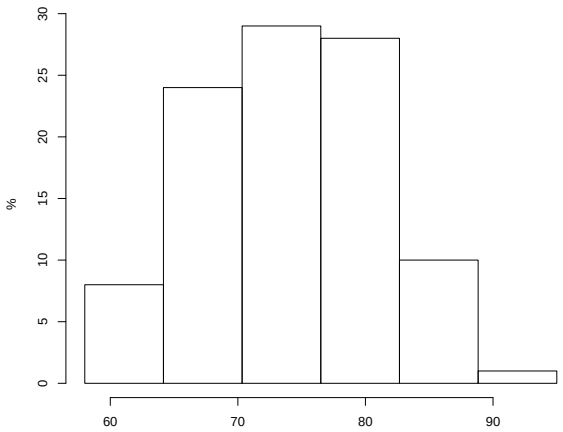
A frequência relativa das outras classes são calculadas de forma similar e podem ser verificadas na Tabela 2.4.

Tabela 2.4: Frequência relativa das 6 classes intervalares das variáveis do conjunto Futebol

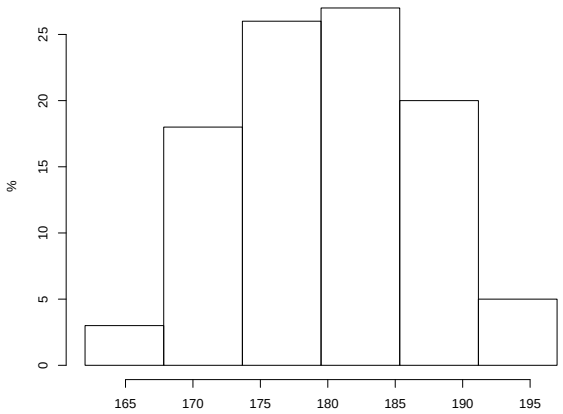
Classes	Frequência relativa		
	Peso (Y)	Altura (X_1)	Idade (X_2)
I_1	0,0775	0,0285	0,1604
I_2	0,2454	0,1860	0,2129
I_3	0,2867	0,2638	0,2129
I_4	0,2789	0,2705	0,2099
I_5	0,0989	0,2030	0,1549
I_6	0,0123	0,0480	0,0489

A Figura 2.2 descreve o comportamento dos histogramas das três variáveis Peso, Altura e Idade no conjunto Futebol. Além disso, as 6 classes dos histogramas de cada variável contém os 20 intervalos dessas respectivas variáveis.

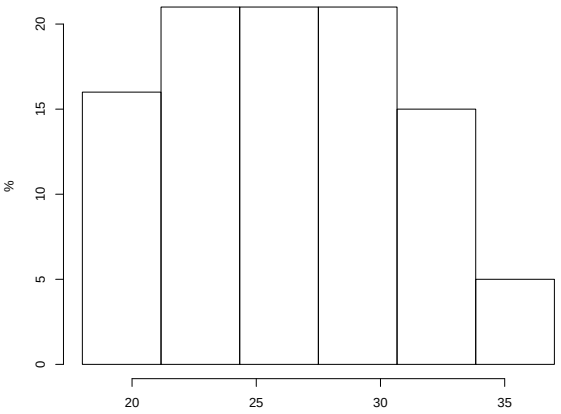
Figura 2.2: Histograma das variáveis intervalares do conjunto Futebol



(a) Peso



(b) Altura



(c) Idade

CAPÍTULO 3

Modelo de regressão elíptico bivariado intervalar

Análise de regressão é uma técnica estatística bastante utilizada quando o pesquisador tem interesse em descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas. Nessas duas últimas décadas os modelos de regressão com erros elípticos tiveram grandes avanços na literatura, Liu (2000) desenvolveu o método da influência local no modelo de regressão linear elíptico. Liu (2002) derivou o método de diagnóstico no modelo de regressão linear multivariado elíptico. Díaz García et al. (2003) propuseram novos métodos de diagnóstico para os modelos de regressão multivariado elíptico, como distância de Cook, análise de resíduos, influência local, entre outros. Savalli et al. (2006) propuseram os modelos mistos lineares elípticos, Ibacache-Pulgar et al. (2012) propuseram o modelo linear misto semiparamétrico com erros elípticos e abordaram a técnica de influência local nesse modelo, Alcantara & Cysneiros (2013) propuseram um modelo de regressão linear com erros elípticos slash, entre outros.

Os trabalhos iniciais sobre modelos de regressão intervalar não consideram uma relação entre o limite inferior e superior de uma variável simbólica intervalar, entretanto essa relação pode de fato existir e influenciar de uma certa maneira as estimativas do modelo. Com base nisso, Neto et al. (2011) propuseram um modelo linear generalizado bivariado para variáveis intervalares que considera uma possível correlação entre o limite inferior e superior. Supondo que as variáveis respostas pertencem a família exponencial bivariada.

Esse capítulo tem como objetivo propor um modelo elíptico intervalar bivariado levando em consideração a relação do limite inferior e superior das variáveis intervalares, em que as variáveis respostas pertencem a família de distribuições elípticas.

3.1 Distribuição elíptica

A classe da família de distribuições elípticas foi introduzida por Kelker (1970), estas foram conceituada como sendo uma generalização da família normal multivariada e tem como característica abranger todas as distribuições contínuas simétricas. Cambanis et al. (1981) define a classe elíptica univariada como sendo semelhante a da família de distribuições simétricas.

As distribuições elípticas representam uma família de distribuições, em que essas tem a característica de ter caudas mais pesadas ou mais leves que a distribuição normal. Dessa forma, essa família tem sido bastante utilizada na modelagem estatística quando o conjunto de dados possuem pontos aberrantes.

São exemplos de distribuições pertencente a família de distribuições elípticas a distribuição t-Student, Exponencial Potência, Logística, Cauchy, entre outras.

Conforme Cambanis et al. (1981), as distribuições elípticas são definidas da seguinte forma:

Definição: Supondo um vetor aleatório $\boldsymbol{Y}_{q\times 1}$ com um vetor de médias $\boldsymbol{\mu}_{q\times 1}$ e uma matriz de variância e covariância $\boldsymbol{\Sigma}_{q\times q} > 0$, $\mathbf{Y}$ tem distribuição elíptica q-variada se sua função característica for dada por:

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{t}) = E\left(e^{i\boldsymbol{t}^\top \boldsymbol{Y}}\right) = e^{i\boldsymbol{t}^\top \boldsymbol{\mu}_i} h(\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}) \tag{3.1}$$

em que $\boldsymbol{\mu} \in \mathbb{R}^n$ um parâmetro de locação, $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ um parâmetro de escala e $h(\cdot) : \mathbb{R} \to [0, \infty)$ é tal que $\int_0^\infty u^{\frac{np}{2}-1} h(u)du < \infty$ e $h(\cdot)$ é conhecida como função geradora de densidades, de acordo com Fang et al. (1990). Dessa maneira se a função característica de $\boldsymbol{Y}$ é representada pela equação (3.1), então $\boldsymbol{Y} \sim El(\boldsymbol{\mu}; \boldsymbol{\Sigma}, h)$.

Sabendo que a variável $\boldsymbol{Y}$ tem distribuição elíptica, quando sua função de densidade e probabilidade existe ela é dada pela seguinte forma:

$$f(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} h(u) \tag{3.2}$$

onde $u = (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$ é a distância de Mahalanobis e $h(\cdot)$ é uma função geradora de densidades. Quando essa distribuição existe tem-se que $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\mu}$ e $\mathrm{Var}(\boldsymbol{Y}) = k\Sigma$, onde k é obtido a partir da função característica na equação (3.1). Assim, $k = -2\varphi'(0)$, em que $\varphi'(\boldsymbol{t}) = \partial\varphi/\partial\boldsymbol{t}|\boldsymbol{t} = 0$.

A Tabela 3.1 é referente a algumas das distribuições elípticas citadas a cima e suas respectivas funções geradoras de densidade.

Tabela 3.1: Função geradora de densidade de distribuições elípticas

Distribuições	Função geradora de densidades
Normal	$h(u) = c_1 \exp(-u/2),\ u \geq 0$
t-Student (ν)	$h(u) = c_2(1 + \frac{u}{\nu})^{-(\nu+m)/2},\ u \geq 0$
Exponencial Potência	$h(u) = c_3 \exp(-u^{\alpha}/2),\ u \geq 0$
Logística	$h(u) = c_4 \frac{\exp(-u)}{[1+\exp(-u)]^2},\ u \geq 0$
Cauchy	$h(u) = c_5(1 + u)^{-(m+1)/2},\ u \geq 0$

Fonte: Galea et al. (2000)

em que, os c's representam constantes normalizadas, α e ν parâmetros extras, m a dimensão da distribuição e u distância de Mahalanobis.

3.1.1 Propriedades da distribuição elíptica

Muitos autores provaram que a maioria das propriedades da distribuição normal também são válidas para a família de distribuições elípticas (por exemplo, Fang et al. (1990), Arellano-Valle (1994) e Ferreira (2008)). Para obter as propriedades abaixo é importante supor que $\boldsymbol{Y} = (Y_1, \ldots, Y_q)^{\top} \in \mathbb{R}^q$, $\boldsymbol{Y} \sim El_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$, o posto$(\boldsymbol{\Sigma}) = k \leq q$ e que $\boldsymbol{\Sigma} \neq 0$.

1. Quando $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \boldsymbol{I}_q$, em que $\boldsymbol{I}_q$ é uma matriz identidade $(q \times q)$, temos que Y segue distribuição elíptica padrão;

2. $\text{Cov}(\boldsymbol{Y}) = E[(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})^{\top}] = -2h'(\mathbf{0})\boldsymbol{\Sigma}$, em que $h'(\mathbf{0})$ é a primeira derivada da função geradora de densidade h aplicada no ponto $\mathbf{0}$;

3. $\boldsymbol{Y}^*$ é um vetor encontrado a partir da combinação linear de $\boldsymbol{Y}$ dada por $\boldsymbol{Y}^* = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{b}$, onde $\boldsymbol{A}_{q\times q}$ é uma matriz não singular e $\boldsymbol{b}_{q\times 1}$ um vetor. Assim $\boldsymbol{Y}^* \sim El(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}; \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\top}; h)$;

4. Considerando as seguintes partições

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}^{(1)} \\ \boldsymbol{y}^{(2)} \end{pmatrix},\ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \text{ e } \boldsymbol{\Sigma} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

onde os sub-vetores $\boldsymbol{y}^{(1)}$ e $\boldsymbol{\mu}^{(1)}$ tem dimensões $(p \times 1)$ e $\mathbf{y}^{(2)}$ e $\boldsymbol{\mu}^{(2)}$ tem $((q-p) \times 1)$, em que $p < q$. As distribuições marginais de $\boldsymbol{y}$ são dadas por:

$$y^{(1)} \sim El_p(\boldsymbol{\mu}^{(1)}; \boldsymbol{\Sigma}_{11}; h)$$

e

$$y_2 \sim El_{q-p}(\boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma}_{22}; h).$$

Já a distribuição condicional tem a seguinte estrutura:

$$\boldsymbol{y}^{(1)}|\boldsymbol{y}_0^{(2)} \sim El_{q-p}(\boldsymbol{\mu}^{(1*)}; \boldsymbol{\Sigma}_{11}^*; h_2)$$

onde $\boldsymbol{\mu}^{(1*)} = \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_0^{(2)} - \boldsymbol{\mu}^{(2)})$, $\boldsymbol{\Sigma}_{11}^* = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ e h_2 é uma função obtida a partir de h. A distribuição condicional de $\boldsymbol{y}^{(2)}|\boldsymbol{y}_0^{(1)}$ é obtida de forma similar.

Para mais detalhes sobre as distribuições elípticas verificar Cambanis et al. (1981), Anderson et al. (1986), Anderson & Fang (1990), Fang et al. (1990), Arellano-Valle (1994), entre outros.

3.2 Modelo elíptico bivariado intervalar

3.2.1 Representação limite inferior e superior

As variáveis respostas e explicativas a serem estudadas neste trabalho são do tipo simbólicas intervalares.

Seja uma variável aleatória bidimensional $\boldsymbol{Y}$ com densidade elíptica dada pela equação (3.2), em que $\boldsymbol{Y} = (Y_I, Y_S)^\top$, $Y_I \leq Y_S$, Y_I é o limite inferior do intervalo e Y_S é o limite superior do intervalo, com matriz de locação $\boldsymbol{\mu} = (\mu_I,\ \mu_S)^\top$ e matriz escala $\boldsymbol{\Sigma}_{2\times 2}$ dada por:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix},$$

em que $\phi_{12} = \phi 21$ e $\boldsymbol{Y} \sim El(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$.

Temos o modelo de regressão elíptico bivariado intervalar definido por

$$\begin{pmatrix} Y_{Ii} \\ Y_{Si} \end{pmatrix} = \begin{pmatrix} \mu_{Ii} \\ \mu_{Si} \end{pmatrix} + \boldsymbol{\Sigma}^{1/2} \begin{pmatrix} \epsilon_{Ii} \\ \epsilon_{Si} \end{pmatrix} \quad i = 1, \ldots, n \tag{3.3}$$

em que $\boldsymbol{\Sigma}^{1/2}$ é uma matriz tal que $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} \times \boldsymbol{\Sigma}^{1/2}$, ϵ_i são vetores aleatórios bivariados identicamente distribuídos com $\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{Ii} \\ \epsilon_{Si} \end{pmatrix} \sim El_2(0, \boldsymbol{I}, h)$. A relação do vetor de médias com o preditor linear são definidas por

$$\boldsymbol{\eta}_{Ii} = g_I(\boldsymbol{\mu}_{Ii}) = x_{Ii1}\boldsymbol{\beta}_{I1} + \ldots + x_{Iip}\boldsymbol{\beta}_{Ip} \quad \text{e} \quad \boldsymbol{\eta}_{Si} = g_S(\boldsymbol{\mu}_{Si}) = x_{Si1}\boldsymbol{\beta}_{S1} + \ldots + x_{Sip}\boldsymbol{\beta}_{Sp}$$

com $\mathbf{X}_I$ e $\mathbf{X}_S$ sendo as matrizes de valores observados das variáveis X_{Ij} e X_{Sj} $(j = 1, \ldots, p)$, respectivamente, $\boldsymbol{\beta}_I = (\beta_{I1}, \ldots, \beta_{Ip})^\top$ e $\boldsymbol{\beta}_S = (\beta_{S1}, \ldots, \beta_{Sp})^\top$ são os vetores de parâmetros a serem estimados, $\boldsymbol{\eta}_I = (\eta_{I1}, \ldots, \eta_{In})^\top$ e $\boldsymbol{\eta}_S = (\eta_{S1}, \ldots, \eta_{Sn})^\top$ são os preditores lineares, $\boldsymbol{\mu}_I = (\mu_{I1}, \ldots, \mu_{In})^\top$ e $\boldsymbol{\mu}_S = (\mu_{S1}, \ldots, \mu_{Sn})^\top$ são respectivamente as médias das variáveis respostas y_{Ii} e y_{Si}, em que y_{Ii} é o limite inferior do i-ésimo intervalo e

y_{Si} é o limite superior do i-ésimo intervalo. Temos ainda que $g_I(\boldsymbol{\mu}_{Ii})$ e $g_S(\boldsymbol{\mu}_{Si})$ são funções monótonas e diferenciais, denominadas funções de ligações. Essas funções relacionam a média da variável resposta ao preditor linear , sendo na representação limite inferior e superior a função identidade.

A função de verossimilhança associado ao modelo (3.3) é dada por:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} h(u_i)$$

onde $|\boldsymbol{\Sigma}| = \phi_{11}\phi_{22} - \phi_{12}^2$. A partir dessa função tem-se que o logaritmo da função de verossimilhança do modelo (3.3) pode ser escrita pela seguinte expressão

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log\left(\prod_{i=1}^{n} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} h(u_i)\right) \\ &= \sum_{i=1}^{n}\left[-\frac{1}{2} log(\phi_{11}\phi_{22} - \phi_{12}^2) + log[h(u_i)]\right] \end{aligned} \tag{3.4}$$

onde $u_i = (\boldsymbol{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$ representa a distância de Mahalanobis, $\boldsymbol{y}_i = \begin{pmatrix} y_{Ii} \\ y_{Si} \end{pmatrix}$, $\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{Ii} \\ \mu_{Si} \end{pmatrix}$ e $\boldsymbol{\theta} = (\boldsymbol{\beta}_I^\top, \boldsymbol{\beta}_S^\top, \phi_{11}, \phi_{22}, \phi_{12})^\top$ os parâmetros do modelo a serem estimados.

3.2.2 Representação centro e amplitude

A equação (3.3) com função de ligação identidade não garante a coerência matemática em que os valores previsto do limite superior sejam maior que o inferior. Dessa forma, propomos a representação centro e amplitude definida por $\boldsymbol{Y}^* = \begin{pmatrix} Y_c \\ Y_a \end{pmatrix}$, sendo

$$Y_c = \frac{Y_I + Y_S}{2} \quad \text{e} \quad Y_a = Y_S - Y_I$$

As variáveis Y_c e Y_a representam respectivamente os valores dos centros e das amplitudes da variável intervalar $\boldsymbol{Y}^*$. Da mesma forma, temos a representação centro e amplitude para as variáveis explicativas intervalares definida por $X_c = \frac{X_I + X_S}{2}$ e $X_a = X_S - X_I$ em que X_S é o limite superior do intervalo da variável explicativa e X_I é o limite inferior do intervalo da variável explicativa. Conforme a propriedade 3 mencionada na seção 3.1.1, a variável $\boldsymbol{Y^*}$ segue uma distribuição $El(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^\top, h)$, em que $\boldsymbol{A} = \begin{bmatrix} 1/2 & 1/2 \\ -1 & 1 \end{bmatrix}$ é uma matriz não singular. Com base nisso, o vetor de médias e a matriz escala dessa nova variável são definidas por:

$$\boldsymbol{\mu}^* = \boldsymbol{A}\boldsymbol{\mu} = \begin{pmatrix} \mu_c \\ \mu_a \end{pmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ -1 & 1 \end{bmatrix} \times \begin{pmatrix} \mu_I \\ \mu_S \end{pmatrix} = \begin{pmatrix} \frac{\mu_I+\mu_S}{2} \\ \mu_S - \mu_I \end{pmatrix}$$

e

$$\begin{aligned} \boldsymbol{\Sigma}^* &= \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top = \begin{bmatrix} \phi^*_{11} & \phi^*_{12} \\ \phi^*_{21} & \phi^*_{22} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ -1 & 1 \end{bmatrix} \times \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \times \begin{bmatrix} 1/2 & -1 \\ 1/2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{4}(\phi_{11} + 2\phi_{12} + \phi_{22}) & \frac{1}{2}(\phi_{22} - \phi_{11}) \\ \frac{1}{2}(\phi_{22} - \phi_{11}) & \phi_{11} - 2\phi_{12} + \phi_{22} \end{bmatrix}. \end{aligned}$$

De forma similar temos que a matriz $\boldsymbol{\Sigma}$ também pode ser reescrita em termos dos parâmetros da matriz $\boldsymbol{\Sigma}^*$ da seguinte forma

$$\boldsymbol{\Sigma} = \begin{bmatrix} \phi^*_{11} + \frac{\phi^*_{22}}{4} - \phi^*_{12} & \phi^*_{11} - \frac{\phi^*_{22}}{4} \\ \phi^*_{11} - \frac{\phi^*_{22}}{4} & \phi^*_{11} + \frac{\phi^*_{22}}{4} + \phi^*_{12} \end{bmatrix}.$$

Baseado na relação acima podemos apresentar os seguintes resultados:

- Para representação limite inferior e superior

  $\phi_{11} = \phi^*_{11} + \frac{\phi^*_{22}}{4} - \phi^*_{12}$;

  $\phi_{22} = \phi^*_{11} + \frac{\phi^*_{22}}{4} + \phi^*_{12}$;

  $\phi_{12} = \phi_{21} = \phi^*_{11} - \frac{\phi^*_{22}}{4}$.

- Para representação centro e amplitude

  $\phi^*_{11} = \frac{1}{4}(\phi_{11} + 2\phi_{12} + \phi_{22})$;

  $\phi^*_{22} = \phi_{11} - 2\phi_{12} + \phi_{22}$;

  $\phi^*_{12} = \phi^*_{21} = \frac{1}{2}(\phi_{22} - \phi_{11})$.

De uma forma geral temos que

1. Quando $\phi_{11} = \phi_{22}$ temos que $\phi^*_{12} = 0 \ \ \forall \phi_{12} \in \mathbb{R}$;
2. Quando $\phi_{11} > \phi_{22}$ temos que $\phi^*_{12} < 0 \ \ \forall \phi_{12} \in \mathbb{R}$;
3. Quando $\phi_{11} < \phi_{22}$ temos que $\phi^*_{12} > 0 \ \ \forall \phi_{12} \in \mathbb{R}$;
4. Quando $\phi^*_{11} = \phi^*_{22}$ temos que $\phi_{12} > 0 \ \ \forall \phi^*_{12} \in \mathbb{R}$;
5. Quando $\phi^*_{11} > \phi^*_{22}$ temos que $\phi_{12} > 0 \ \ \forall \phi^*_{12} \in \mathbb{R}$;
6. Quando $\phi^*_{22} > 4\phi^*_{11}$ temos que $\phi_{12} < 0 \ \ \forall \phi^*_{12} \in \mathbb{R}$;
7. Quando $\phi^*_{22} = 4\phi^*_{11}$ temos que $\phi_{12} = 0 \ \ \forall \phi^*_{12} \in \mathbb{R}$.

em que esses resultados foram utilizados nos estudos de simulação.

Portanto, o modelo com a nova representação centro e amplitude é definido por:

$$\begin{pmatrix} Y_{ci} \\ Y_{ai} \end{pmatrix} = \begin{pmatrix} \mu_{ci} \\ \mu_{ai} \end{pmatrix} + \boldsymbol{\Sigma}^{*1/2} \begin{pmatrix} \epsilon_{ci} \\ \epsilon_{ai} \end{pmatrix} \quad i = 1, \ldots, n \tag{3.5}$$

em que $\boldsymbol{\Sigma}^{*1/2}$ é uma matriz tal que $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}^{*1/2} \times \boldsymbol{\Sigma}^{*1/2}$, ϵ_i^* são vetores aleatórios bivariados identicamente distribuídos com $\boldsymbol{\epsilon}_i^* \sim El_2(0, \boldsymbol{I}, h)$. A nova relação do vetor de médias com o preditor linear são definidas por

$$\boldsymbol{\eta}_{ci} = g_c(\boldsymbol{\mu}_{ci}) = x_{ci1}\boldsymbol{\beta}_{c1} + \ldots + x_{cip}\boldsymbol{\beta}_{cp} \quad \text{e} \quad \boldsymbol{\eta}_{ai} = g_a(\boldsymbol{\mu}_{ai}) = x_{ai1}\boldsymbol{\beta}_{a1} + \ldots + x_{aip}\boldsymbol{\beta}_{ap}$$

com $\mathbf{X}_c$ e $\mathbf{X}_a$ sendo as matrizes de valores observados das variáveis X_{cj} e X_{aj} $(j = 1, \ldots, p)$, respectivamente, $\boldsymbol{\beta}_c = (\beta_{c1}, \ldots, \beta_{cp})^\top$ e $\boldsymbol{\beta}_a = (\beta_{a1}, \ldots, \beta_{ap})^\top$ são os vetores de parâmetros a serem estimados, $\boldsymbol{\eta}_c = (\eta_{c1}, \ldots, \eta_{cn})^\top$ e $\boldsymbol{\eta}_a = (\eta_{a1}, \ldots, \eta_{an})^\top$ são os preditores lineares, $\boldsymbol{\mu}_c = (\mu_{c1}, \ldots, \mu_{cn})^\top$ e $\boldsymbol{\mu}_a = (\mu_{a1}, \ldots, \mu_{an})^\top$ são respectivamente as médias das variáveis respostas y_{ci} e y_{ai}, em que y_{ci} é o centro do i-ésimo intervalo e y_{ai} é amplitude do i-ésimo intervalo.

Neste caso ao considerar a função de ligação $g_c(\boldsymbol{\mu}_{ci})$ identidade e $g_a(\boldsymbol{\mu}_{ai})$ qualquer função positiva, garantimos a coerência matemática que os limite inferior é menor que limite superior. Nesse trabalho iremos utilizar $g_a(\boldsymbol{\mu}_{ai})$ como sendo uma função logarítma.

A função de verossimilhança associado ao novo modelo (3.5) é dada por:

$$L(\boldsymbol{\theta}^*) = \prod_{i=1}^{n} |\boldsymbol{\Sigma}^*|^{-\frac{1}{2}} h(u_i^*)$$

onde $|\boldsymbol{\Sigma}^*| = \phi_{11}^*\phi_{22}^* - \phi_{12}^{*2}$. A partir dessa função, temos que o logaritmo da função de verossimilhança do modelo (3.5) pode ser escrita por

$$\begin{aligned} l(\boldsymbol{\theta}^*) &= \log\left(\prod_{i=1}^{n} |\boldsymbol{\Sigma}^*|^{-\frac{1}{2}} h(u_i^*)\right) \\ &= \sum_{i=1}^{n}\left[-\frac{1}{2} log(\phi_{11}^*\phi_{22}^* - \phi_{12}^{*2}) + log[h(u_i^*)]\right] \end{aligned} \tag{3.6}$$

em que $u_i^* = (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^\top \boldsymbol{\Sigma}^{*-1}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)$ representa a distância de Mahalanobis, $\boldsymbol{y}_i^* = \begin{pmatrix} y_{ci} \\ y_{ai} \end{pmatrix}$, $\boldsymbol{\mu}_i^* = \begin{pmatrix} \mu_{ci} \\ \mu_{ai} \end{pmatrix}$ e $\boldsymbol{\theta}^* = (\boldsymbol{\beta}_c^\top, \boldsymbol{\beta}_a^\top, \phi_{11}^*, \phi_{22}^*, \phi_{12}^*)^\top$ os parâmetros do novo modelo a serem estimados, e y_{ci} é o valor do centro do i-ésimo intervalo e y_{ai} o valor da amplitude do i-ésimo intervalo.

3.2.3 Função *Escore* e Informação de *Fisher*

A função suporte é chamada de função (ou vetor) *escore*, essa é definida como sendo a primeira derivada do logaritmo da função de verossimilhança com relação aos parâmetros de interesse do modelo. A função *escore* é definida por:

$$\boldsymbol{U}(\boldsymbol{\theta}^*) = \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^*},$$

onde $\boldsymbol{\theta}^*$ é o vetor de parâmetros a ser estimado no modelo. Nesse trabalho o vetor $\boldsymbol{\theta}^* = (\boldsymbol{\beta}_c^\top, \boldsymbol{\beta}_a^\top, \phi_{11}^*, \phi_{22}^*, \phi_{12}^*)^\top$. Portanto, as funções *escore* dos parâmetros do modelo (3.5) são calculadas a partir da equação 3.6 dadas por

$$\mathbf{U}(\boldsymbol{\beta}_c) = (\mathbf{U}_{\boldsymbol{\beta}_{c1}}, \ldots, \mathbf{U}_{\boldsymbol{\beta}_{cp}})^\top$$

em

$$\begin{aligned}
\mathbf{U}(\boldsymbol{\beta}_{cj}) &= \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \beta_{cj}} \\
&= \sum_{i=1}^{n} \left\{ \frac{h'(u_i^*)}{h(u_i^*)} \frac{\partial u_i^*}{\partial \beta_{cj}} \right\},
\end{aligned}$$

em que $\frac{\partial u_i^*}{\partial \beta_{cj}}$ é derivada de u_i com relação a β_{cj}, $j = 1, \ldots, p$. Então,

$$\mathbf{U}(\boldsymbol{\beta}_{cj}) = -2 \sum_{i=1}^{n} \left\{ \frac{h'(u_i^*)}{h(u_i^*)} \boldsymbol{t}_{cji}^\top \boldsymbol{\Sigma}^{*-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) \right\},$$

onde o vetor $\boldsymbol{t}_{cji}^\top = \frac{\partial (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^\top}{\partial \boldsymbol{\beta}_{cj}} = (\boldsymbol{x}_{cji} g_c'^{-1}(\eta_{ci}),\ \mathbf{0})$. De forma similar, tem-se que as funções *escore*'s dos outros parâmetros ficam definidas como

$$\begin{aligned}
\boldsymbol{U}(\beta_{aj}) &= \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \beta_{aj}} \\
&= \sum_{i=1}^{n} \left\{ \frac{h'(u_i^*)}{h(u_i^*)} \frac{\partial u_i^*}{\partial \beta_{aj}} \right\} \\
&= -2 \sum_{i=1}^{n} \left\{ \frac{h'(u_i^*)}{h(u_i^*)} \mathbf{t}_{aji}^\top \boldsymbol{\Sigma}^{*-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) \right\},
\end{aligned}$$

onde o vetor $\boldsymbol{t}_{aji}^{\top} = \frac{\partial(\mathbf{y}_i^* - \boldsymbol{\mu}_i^*)^{\top}}{\partial \boldsymbol{\beta}_{aj}} = (\mathbf{0},\ \mathbf{x}_{aji} g_a'^{-1}(\eta_{ai}))$.

$$\begin{aligned}
\boldsymbol{U}(\phi_{11}^*) &= \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \phi_{11}^*} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*} \right) - \frac{h'(u_i^*)}{h(u_i^*)} \frac{\partial u_i^*}{\partial \phi_{11}^*} \right\} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*} \right) - \frac{h'(u_i^*)}{h(u_i^*)} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^{\top} \boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*} \boldsymbol{\Sigma}^{*-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) \right\},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{U}(\phi_{22}^*) &= \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \phi_{22}^*} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{22}^*} \right) - \frac{h'(u_i^*)}{h(u_i^*)} \frac{\partial u_i^*}{\partial \phi_{22}^*} \right\} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{22}^*} \right) - \frac{h'(u_i^*)}{h(u_i^*)} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^{\top} \boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{22}^*} \boldsymbol{\Sigma}^{*-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) \right\},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{U}(\phi_{12}^*) &= \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \phi_{12}^*} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{12}^*} \right) - \frac{h'(u_i^*)}{h(u_i^*)} \frac{\partial u_i^*}{\partial \phi_{12}^*} \right\} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{12}^*} \right) - \frac{h'(u_i^*)}{h(u_i^*)} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^{\top} \boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{12}^*} \boldsymbol{\Sigma}^{*-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) \right\}.
\end{aligned}$$

A matriz de Informação de *Fisher* do modelo é definida como

$$\boldsymbol{K}_{\boldsymbol{\theta}^*\boldsymbol{\theta}^*} = E[U(\boldsymbol{\theta}) U(\boldsymbol{\theta})^{\top}] = \begin{bmatrix}
\boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_c} & \boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_a} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\boldsymbol{K}_{\boldsymbol{\beta}_a\boldsymbol{\beta}_c} & \boldsymbol{K}_{\boldsymbol{\beta}_a\boldsymbol{\beta}_a} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \boldsymbol{K}_{\phi_{11}^*\phi_{11}^*} & \boldsymbol{K}_{\phi_{11}^*\phi_{22}^*} & \boldsymbol{K}_{\phi_{11}^*\phi_{12}^*} \\
\mathbf{0} & \mathbf{0} & \boldsymbol{K}_{\phi_{22}^*\phi_{11}^*} & \boldsymbol{K}_{\phi_{22}^*\phi_{22}^*} & \boldsymbol{K}_{\phi_{22}^*\phi_{12}^*} \\
\mathbf{0} & \mathbf{0} & \boldsymbol{K}_{\phi_{12}^*\phi_{11}^*} & \boldsymbol{K}_{\phi_{12}^*\phi_{22}^*} & \boldsymbol{K}_{\phi_{12}^*\phi_{12}^*}
\end{bmatrix}.$$

O elemento $(1,1)$ da matriz de informação é definido como

$$\begin{aligned}
\boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_c} &= E\left(\frac{\partial l(\boldsymbol{\theta}^*)}{\partial \beta_{cj}} \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \beta_{ck}} \right) \\
&= \sum_{i=1}^{n} \left\{ 4E[W_h (u_i^*)^2 (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^{\top} \boldsymbol{\Sigma}^{*-1} t_{cji} t_{cki}^{\top} \boldsymbol{\Sigma}^{*-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)] \right\},
\end{aligned}$$

onde $W_h(u_i^*) = \frac{h'(u_i^*)}{h(u_i^*)}$ e $Z_i = \boldsymbol{\Sigma}^{*-1/2}(y_i^* - \mu_i^*)$. Então,

$$\begin{aligned}
\boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_c} &= \sum_{i=1}^{n}\left\{4E\left[W_h^2(u_i^*)||Z_i||^2\frac{Z_i^\top}{||Z_i||}\boldsymbol{\Sigma}^{*-1/2}t_{cji}t_{cki}^\top\boldsymbol{\Sigma}^{*-1/2}\frac{Z_i}{||Z_i||}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{4E\left[E\left[W_h^2(u_i^*)||Z_i||^2\frac{Z_i^\top}{||Z_i||}\boldsymbol{\Sigma}^{*-1/2}t_{cji}t_{cki}^\top\boldsymbol{\Sigma}^{*-1/2}\frac{Z_i}{||Z_i||}\mid\ ||Z||\right]\right]\right\} \\
&= \sum_{i=1}^{n}\left\{4E\left[W_h^2(u_i^*)||Z_i||^2E\left[\frac{Z_i^\top}{||Z_i||}\boldsymbol{\Sigma}^{*-1/2}t_{cji}t_{cki}^\top\boldsymbol{\Sigma}^{*-1/2}\frac{Z_i}{||Z_i||}\mid\ ||Z||\right]\right]\right\},
\end{aligned}$$

em que $E\left(\frac{Z^\top}{||Z||}A\frac{Z}{||Z||}\mid\ ||Z||\right) = \frac{1}{k}tr(A)$ e a matriz A tem dimensão $(k\times k)$ em Lange et al. (1989). Assim,

$$\boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_c} = \sum_{i=1}^{n}\left\{4E\left[W_h^2(u_i^*)||Z_i||^2\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{*-1/2}t_{cji}t_{cki}^\top\boldsymbol{\Sigma}^{*-1/2}\right)\right]\right\},$$

onde $E\left[W_h^2(u_i^*)||Z_i||^2\right] = d_{hi}$ e $tr(AB) = tr(BA)$. Portanto,

$$\begin{aligned}
\boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_c} &= \sum_{i=1}^{n}2d_{hi}tr(t_{cji}^\top\boldsymbol{\Sigma}^{*-1}t_{cki}) \\
&= \sum_{i=1}^{n}2d_{hi}t_{cji}^\top\boldsymbol{\Sigma}^{*-1}t_{cki} \\
&= \sum_{i=1}^{n}2d_{hi}t_{cji}^\top\boldsymbol{\Sigma}^{*-1}t_{cki}.
\end{aligned}$$

De forma similar são calculados os elementos (2×2), (1×2) e (2×1), em que os elementos (1×2) e (2×1) são equivalentes.

$$\boldsymbol{K}_{\boldsymbol{\beta}_a\boldsymbol{\beta}_a} = \sum_{i=1}^{n}2d_{hi}\boldsymbol{t}_{aji}^\top\boldsymbol{\Sigma}^{*-1}\boldsymbol{t}_{aki},$$

$$\boldsymbol{K}_{\boldsymbol{\beta}_c\boldsymbol{\beta}_a} = \sum_{i=1}^{n}2d_{hi}\boldsymbol{t}_{cji}^\top\boldsymbol{\Sigma}^{*-1}\boldsymbol{t}_{aji},$$

O elemento $(3,3)$ da matriz de informação é definido como

$$\begin{aligned}
\boldsymbol{K}_{\phi_{11}^*\phi_{11}^*} &= E\left(\frac{\partial l(\boldsymbol{\theta}^*)}{\partial \phi_{11}^*}\frac{\partial l(\boldsymbol{\theta}^*)}{\partial \phi_{11}^*}\right) \\
&= \sum_{i=1}^{n}\left\{E\left[\left(-\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right) - \frac{h'(u_i^*)}{h(u_i^*)}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^\top \boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)\right)\times\right.\right. \\
&\quad \left.\left.\left(-\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right) - \frac{h'(u_i^*)}{h(u_i^*)}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)^\top \boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)\right)\right]\right\} \\
&= \sum_{i=1}^{n}\left\{E\left[\frac{1}{4}tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right) + \frac{1}{2}tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)\times\right.\right. \\
&\quad W_h(u_i^*)Z_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_i + \frac{1}{2}tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)\times \\
&\quad W_h(u_i^*)Z_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_i + W_h^2(u_i^*)Z_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_i\times \\
&\quad \left.\left.Z_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_i\right]\right\}
\end{aligned}$$

onde $c_1 = tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)$. Então,

$$\begin{aligned}
\boldsymbol{K}_{\phi_{11}^*\phi_{11}^*} &= \sum_{i=1}^{n}\left\{\frac{c_1}{4} + tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)\times E\left[W_h(u_i^*)Z_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_i\right] + \right. \\
&\quad \left. + E\left[W_h^2(u_i^*)Z_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_iZ_i^\top \boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}Z_i\right]\right\}
\end{aligned}$$

onde $E\left(\frac{Z^\top}{||Z||}A\frac{Z}{||Z||}B\frac{Z^\top}{||Z||}\mid\ ||Z||\right) = \frac{1}{k(k+2)}[2tr(AB) + tr(A)tr(B)]$, de acordo com Lange et al. (1989). Assim,

$$\begin{aligned}
\boldsymbol{K}_{\phi_{11}^*\phi_{11}^*} &= \sum_{i=1}^{n}\left\{\frac{c_1}{4} + \frac{1}{2}tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)E\left[W_h(u_i^*)||Z_i||^2\right] + \right. \\
&\quad + \frac{1}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + \right. \\
&\quad \left.\left. + tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\right)\right] \times E\left[W_h^2(u_i^*)||Z_i||^4\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{c_1}{4} + \frac{c_1}{2}E\left[W_h(u_i^*)||Z_i||^2\right] + \right. \\
&\quad + \frac{1}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial \boldsymbol{\Sigma}^*}{\partial \phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_1\right] \times \\
&\quad \left. \times E\left[W_h^2(u_i^*)||Z_i||^4\right]\right\}
\end{aligned}$$

em que $E\left(W_h^2(u_i^*)||Z_i||^4\right) = f_{hi}$ e $E\left(W_h(u_i^*)||Z_i||^2\right) = -\frac{k}{2} = -\frac{2}{2} = -1$. Portanto,

$$\boldsymbol{K}_{\phi_{11}^*\phi_{11}^*} = \sum_{i=1}^{n}\left\{-\frac{c_1}{4} + \frac{f_{hi}}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_1\right]\right\}.$$

De forma análoga, tem-se que as informações de *Fisher* restantes são definidas como

$$\boldsymbol{K}_{\phi_{22}^*\phi_{22}^*} = \sum_{i=1}^{n}\left\{-\frac{c_2}{4} + \frac{f_{hi}}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_2\right]\right\},$$

$$\boldsymbol{K}_{\phi_{12}^*\phi_{12}^*} = \sum_{i=1}^{n}\left\{-\frac{c_3}{4} + \frac{f_{hi}}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_3\right]\right\},$$

$$\boldsymbol{K}_{\phi_{11}^*\phi_{22}^*} = \sum_{i=1}^{n}\left\{-\frac{c_4}{4} + \frac{f_{hi}}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_4\right]\right\},$$

$$\boldsymbol{K}_{\phi_{11}^*\phi_{12}^*} = \sum_{i=1}^{n}\left\{-\frac{c_5}{4} + \frac{f_{hi}}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{11}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_5\right]\right\},$$

$$\boldsymbol{K}_{\phi_{22}^*\phi_{12}^*} = \sum_{i=1}^{n}\left\{-\frac{c_6}{4} + \frac{f_{hi}}{8}\left[2tr\left(\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\boldsymbol{\Sigma}^{*-1/2}\right) + c_6\right]\right\}.$$

onde c_2, c_3, c_4, c_5 e c_6 são constantes e são definidas como $c_2 = tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\right)$, $c_3 = tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\right)$, $c_4 = tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{11}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\right)$, $c_5 = tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{11}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\right)$ e $c_6 = tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{22}^*}\right)tr\left(\boldsymbol{\Sigma}^{*-1}\frac{\partial\boldsymbol{\Sigma}^*}{\partial\phi_{12}^*}\right)$. Conforme Lange et al. (1989), os elementos da matriz de informação de *Fisher* $(1,3)$, $(1,4)$, $(1,5)$, $(2,3)$, $(2,4)$, $(2,5)$, $(3,1)$, $(3,2)$, $(4,1)$, $(4,2)$, $(5,1)$ e $(5,2)$ são iguais a zero, desde que $||Z_i||$ fixado em $\frac{\partial l(\boldsymbol{\theta}^*)}{\partial\beta^*}$ seja uma função par de Z e $\frac{\partial l(\boldsymbol{\theta}^*)}{\partial\phi^*}$ seja uma função ímpar de Z.

3.2.4 Estimação

Para obter as estimativas dos parâmetros via o método de máxima verossimilhança é necessário recorrer a um processo iterativo, visto que pelo método da máxima verossimilhança não é possível obter uma forma analítica para o estimador. Atualmente existe na literatura vários métodos interativos. São exemplos, o de Newton-Rapshon, Escore de Fisher, EM, BHHH, Método Quasi-Newton (BFGS), entre outros. Neste trabalho utilizaremos o método Quasi-Newton (BFGS).

Método Quasi-Newton BFGS

Suponha que o objetivo é maximizar uma determinada função, $F : \Theta \longrightarrow \mathbb{R}$, em que Θ é um subespaço de $\mathbb{R}^p$. Sendo F uma função quadrática, podendo ser escrita como

$$F(\boldsymbol{\theta}^*) = a + b^\top \boldsymbol{\theta}^* - \frac{1}{2}\boldsymbol{\theta}^{*\top} C \boldsymbol{\theta}^*,$$

em que a é uma constante, b é um vetor de constantes de dimensão $(p \times 1)$ e C é uma matriz positiva-definida $(p \times p)$.

Na condição de primeira ordem para a maximização de F, tem-se:

$$\frac{\partial F(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^*} = b - \frac{1}{2}2C\boldsymbol{\theta}^* = 0 \Rightarrow b - C\boldsymbol{\theta}^* = 0 \Rightarrow \hat{\boldsymbol{\theta}}^* = C^{-1}b,$$

resultado obtido com a condição de que C é positiva-definida, o que garante a existência de uma matriz inversa C. Nesse caso, a solução possui forma fechada. No entanto, a maioria dos problemas encontrados são aqueles em que a condição de primeira ordem forma um sistema de equações não-lineares que não apresenta solução em forma fechada. Para solucionar problemas como esse é proposto um esquema iterativo dado por

$$\boldsymbol{\theta}^*{}_{t+1} = \boldsymbol{\theta}^*{}_t + \lambda_t \boldsymbol{\Delta}_t.$$

Partindo de $\boldsymbol{\theta}^*{}_0$, se na iteração t o valor máximo de $\boldsymbol{\theta}^*$ não tiver sido alcançado, calcula-se $\boldsymbol{\Delta}_t$ o vetor direcional de dimensão $(p \times 1)$ e λ_t o "tamanho do passo". Assim, um problema auxiliar de otimização é encontrado, pois em cada passo deve-se encontrar o valor máximo do tamanho do passo mais adequado. Esse processo é conhecido como busca em linha. A ideia consiste em encontrar o valor de λ_t que resolva a equação abaixo:

$$\frac{\partial F(\boldsymbol{\theta}^*{}_t + \lambda_t \boldsymbol{\Delta}_t)}{\partial \lambda_t} = d(\boldsymbol{\theta}^*{}_t + \lambda_t \boldsymbol{\Delta}_t)^\top \boldsymbol{\Delta}_t = 0,$$

em que d é um vetor de derivadas parciais de F.

Do ponto de vista computacional, a utilização de buscas em linha em algoritmos de otimização não-linear é custosa. Dessa forma, uma proposta menos trabalhosa é a substituição desse processo por um conjunto de regras *ad hoc*.

A classe mais utilizada de algoritmos iterativos é a classe de métodos gradiente. Sendo $\boldsymbol{\Delta}_t = Q_t d_t$, o esquema iterativo tem a seguinte representação:

$$\boldsymbol{\theta}^*{}_{t+1} = \boldsymbol{\theta}^*{}_t + \lambda_t Q_t d_t,$$

em que Q_t é uma matriz positiva-definida e $d_t = d_t(\boldsymbol{\theta}^*) = \dfrac{\partial F(\boldsymbol{\theta}^*{}_t)}{\partial \boldsymbol{\theta}^*{}_t}$ é o gradiente da função F.

A classe Quasi-Newton propõe a construção da seguinte sequência de matrizes:

$$Q_{t+1} = Q_t + E_t,$$

tal que E_t é uma matriz positiva-definida. Assim, todos os elementos da sequência são matrizes positivas-definidas.

O nome da classe Quasi-Newton se deve ao fato de que a matriz hessiana não é utilizada, e sim, uma aproximação para ela feita por iterações. Nessa classe, o método mais utilizado é o BFGS, proposto por Shanno (1985).

Seja $\kappa_t = \gamma_t^\top Q_t \gamma_t$, $\boldsymbol{\delta}_t = \boldsymbol{\theta^*}_{t+1} - \boldsymbol{\theta^*}_t$ e $\gamma_t = d(\boldsymbol{\theta^*}_{t+1}) - d(\boldsymbol{\theta^*}_t)$, o método BFGS é dado por

$$Q_{t+1} = Q_t + \frac{\boldsymbol{\delta}_t \boldsymbol{\delta}_t^\top}{\boldsymbol{\delta}_t^\top \gamma_t} + \frac{Q_t \gamma_t \gamma_t^\top Q_t}{\gamma_t^\top Q_t \boldsymbol{\delta_t}} - \kappa_t \left(\frac{\boldsymbol{\delta}_t}{\boldsymbol{\delta}_t^\top \gamma_t} - \frac{Q_t \gamma_t}{\gamma_t^\top Q_t \gamma_t} \right) \left(\frac{\boldsymbol{\delta}_t}{\boldsymbol{\delta}_t^\top \gamma_t} - \frac{Q_t \gamma_t}{\gamma_t^\top Q_t \gamma_t} \right)^\top.$$

Portanto, Q_t é sempre positiva-definida, desde que a sequência se inicie com uma matriz de mesma propriedade. Assim, além de não envolver o cálculo de segundas derivadas e a necessidade de inverter a matriz hessiana, soluciona-se o problema apresentado no método de Newton-Raphson, onde Q_t poderia não ser positiva-definida quando distante do ponto máximo da função.

Então, com o interesse de estimar $\boldsymbol{\theta}^* = (\boldsymbol{\beta}_c^\top, \boldsymbol{\beta}_a^\top, \phi_{11}^*, \phi_{22}^*, \phi_{12}^*)^\top$ temos

$$\hat{\theta}^* = \text{argmax}_{\theta^*} L(\theta^*).$$

CAPÍTULO 4

Estudo de simulação

Um estudo de simulação, a fim de verificar o comportamento do erro previsão do modelo de regressão elíptico bivariado intervalar (MREBI) na presença de *outliers* nos dados foi proposto. O estudo foi realizado considerando o uso das distribuições Normal e t-Student ($\nu = 4$) para o componente aleatório do modelo (4.1).

$$\begin{pmatrix} Y_{ci} \\ Y_{ai} \end{pmatrix} = \begin{pmatrix} \mu_{ci} \\ \mu_{ai} \end{pmatrix} + \mathbf{\Sigma}^{*1/2} \begin{pmatrix} \epsilon_{ci} \\ \epsilon_{ai} \end{pmatrix} \quad i = 1, \ldots, n \tag{4.1}$$

Em cada cenário foram considerados quatro diferentes tamanhos amostrais $n = 30$, 40, 60 e 100 e cinco diferentes correlações entre as variáveis respostas do modelo (ρ ou ρ^*) $= 0; 0,3; 0,5; 0,7$ e $0,9$. Foram considerados 1000 réplicas de Monte Carlo para cada cenário

4.1 Erro de previsão do modelo

Um dos objetivos do modelos de regressão é avaliar a performace do modelo quando a previsão de novas observações. Nesse trabalho utilizaremos o *Mean Absolute Error* (MAD) que pode ser encontrando em diversos trabalhos, tal como Momeni et al. (2010) sendo descrito pela equação (4.2).

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4.2}$$

em que y_i corresponde ao verdadeiro valor da observação i e $\hat{y}_i$ corresponde ao valor estimando da observação i no modelo.

Para o modelo de regressão elíptico bivariado intervalar dado por (4.1), definimos o *Mean Absolute Error Interval* (MADI) como

$$MADI = \frac{1}{n}\sum_{i=1}^{n} erro_i$$

em que $erro_i = |(y_{ci} - \hat{y}_{ci})| + |(y_{ai} - \hat{y}_{ai})|$, y_{ci} é o verdadeiro valor do centro do i-ésimo intervalo e $\hat{y}_{ci} = g_c^{-1}(\hat{\eta}_{ci})$ o valor estimado do centro do i-ésimo intervalo, y_{ai} é o verdadeiro valor da amplitude do i-ésimo intervalo e $\hat{y}_{ai} = g_a^{-1}(\hat{\eta}_{ai})$ o valor estimado da amplitude i-ésimo intervalo.

4.2 Simulações

Os experimentos consistem em gerar n observações intervalares para as variáveis respostas e explicativas, a partir dessa geração o conjunto de dados foi dividido em dois subconjuntos. O primeiro subconjunto foi definido como os dados de treinamento e o segundo como os dados de teste. Os dados de treinamento correspondem aos 75% das observações e foram utilizadas para compor a base de ajustamento do modelo. Já os dados de teste corresponderam a 25% das observações que compuseram a base que foi utilizada para fazer previsões do modelo.

As variáveis explicativas $X_{c1} = 1$, $X_{a1} = 1$, $X_{c2} \sim U[10, 20]$, $X_{a2} \sim U[1, 5]$, os β_{cj} e β_{aj} foram gerados de $U[0.9, 1.1]$, em que $j = 1$ e 2 , sendo mantidos fixados nas 1000 réplicas de Monte Carlo. As variáveis respostas foram geradas a partir do modelo (4.3)

$$\begin{pmatrix} Y_{ci} \\ Y_{ai} \end{pmatrix} = \begin{pmatrix} \mu_{ci} \\ \mu_{ai} \end{pmatrix} + \Sigma^{*1/2} \begin{pmatrix} \epsilon_{ci} \\ \epsilon_{ai} \end{pmatrix} \quad i = 1, \ldots, n \tag{4.3}$$

em que o vetor $\begin{pmatrix} \epsilon_{ci} \\ \epsilon_{ai} \end{pmatrix} \sim N_2(0, \boldsymbol{I})$, $\mu_{ci} = g_c^{-1}(\eta_{ci})$, $\mu_{ai} = g_a^{-1}(\eta_{ai})$, $\eta_{ci} = x_{ci}^{\top}\beta_c$, $\eta_{ai} = x_{ai}^{\top}\beta_a$, $x_{ci}^{\top} = (x_{ci1}, x_{ci2})$, $x_{ai}^{\top} = (x_{ai1}, x_{ai2})$, $\beta_c = (\beta_{c1}, \beta_{c2})^{\top}$, $\beta_a = (\beta_{a1}, \beta_{a2})^{\top}$, a matriz escala $\boldsymbol{\Sigma^*}$ foram definidas em cada cenário de forma diferente.

Foram definidos sete cenários, em que cada um desses corresponde a uma das relações apresentadas na seção 3.2.2. Nos sete cenários descritos posteriormente consideramos 0%, 5%, 10% e 15% de *outliers* nas 75% das observações da variável resposta $\boldsymbol{Y}_c$. O número de observações *outliers* t_o são calculados e selecionados a partir do tamanho da amostra n e dos percentuais de *outliers* considerados. Após obter os conjuntos de dados $(y_{ci},\ x_{ci})$ e $(y_{ai},\ x_{ai})$ $(i = 1, \ldots, n)$ em cada um dos cenários, selecionar o $(y_{ci},\ x_{ci})$ em ordem crescente e obtenha o último elemento que corresponde ao primeiro elemento t_o ordenado de y_{ci}. Os *outliers* de centro nesses cenários são obtidos por $y_{ci} = y_{ci} + 6 * \sqrt{Var(y_c)}$, em que $i = 1, \ldots, t_o$. A partir disso, ajustou o MEBI e observou o comportamento do erro de previsão a medida que o percentual de *outliers* aumentava 1000 vezes em cada

um dos sete cenários. Em cada réplica de monte carlo o modelo (4.1) foi ajustado e o MADI foi calculado. Para cada cenário foi calculado o $\overline{MADI} = \sum_{B=1}^{1000} \frac{MADI_B}{1000}$ e $s_{MADI} = \sqrt{\frac{1}{999}\sum_{B=1}^{1000}(MADI_B - \overline{MADI})^2}$.

As correlações entre as variáveis respostas na representação limite inferior e superior, centro e amplitude foram calculadas em cada cenário pelas expressões

$$\rho = \frac{\phi_{12}}{\sqrt{\phi_{11}\phi_{22}}}$$

e

$$\rho^* = \frac{\phi_{12}^*}{\sqrt{\phi_{11}^*\phi_{22}^*}}.$$

4.2.1 Cenário 1

Nesse cenário iremos utilizar a relação de quando $\phi_{11} = \phi_{22}$ temos que $\phi_{12}^* = 0 \ \ \forall \phi_{12} \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11} = \phi_{22} = 10$ e $\rho = 0; 0,3; 0.5; 0,7$ e $0,9$ entre as variáveis y_I e y_S. Dado que temos os valores de ϕ_{11}, ϕ_{22} e ϕ_{12} e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}^*, ϕ_{22}^*, ϕ_{12}^* e ρ^* foram obtidos e podem ser vistos na Tabela 4.1. Note que nesse cenário as variáveis y_c e y_a não são correlacionadas, os valores de ϕ_{11}^* aumentam e ϕ_{22}^* diminuem a medida que a correlação entre y_I e y_S cresce.

Tabela 4.1: Valores assumidos pelos parâmetros ϕ_{11}^*, ϕ_{22}^*, ϕ_{12}^* e ρ^* no Cenário 1

Parâmetros	ρ				
	0	0,3	0,5	0,7	0,9
ϕ_{11}^*	5,0	6,5	7,5	8,5	9,5
ϕ_{22}^*	20,0	14,0	10,0	6,0	2,0
ϕ_{12}^*	0	0	0	0	0
ρ^*	0	0	0	0	0

A Tabela 4.2 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 1 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers* de 0%, o erro de previsão não sofreu variações a medida que a correlação entre y_I e y_S aumentava. Quando esse percentual aumenta, os erros de previsão também aumentam. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 0% e um tamanho de amostra 60, notou-se que com $\rho = 0$ o erro de previsão foi $5,1481$, com $\rho = 0,3$ o

erro foi $5,1523$, com $\rho = 0,5$ foi $5,1720$, com $\rho = 0,7$ foi $5,1563$ e com $\rho = 0,9$ foi $5,1404$, indicando pequenas variações a medida que ρ crescia. Já com um percentual de *outliers* de 5%, notou-se que com $\rho = 0$ o erro de previsão foi $6,4666$, com $\rho = 0,3$ o erro foi $6,5509$, com $\rho = 0,5$ foi $6,5956$, com $\rho = 0,7$ foi $6,6466$ e com $\rho = 0,9$ foi $6,6990$, indicando um aumento dos erros a medida que ρ cresce. Quando fixamos um percentual de *outliers* e correlação em 0% e $\rho = 0,3$ respectivamente, o erro de previsão foi $5,3413$ com $n = 30$, $5,2023$ com $n = 40$, $5,1523$ com $n = 60$ e $5,0941$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão não sofreu variações a medida que a correlação entre y_I e y_S aumentava. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 5% e um tamanho de amostra 40, notou-se que com $\rho = 0$ o erro de previsão foi $5,2835$, com $\rho = 0,3$ o erro foi $5,2421$, com $\rho = 0,5$ foi $5,2327$, com $\rho = 0,7$ foi $5,2626$ e com $\rho = 0,9$ foi $5,2449$, indicando pequenas variações no erro a medida que ρ aumenta. Já para um percentual de *outliers* e correlação fixados em 5% e $\rho = 0$ respectivamente, o erro de previsão foi $5,3552$ com $n = 30$, $5,2835$ com $n = 40$, $5,2118$ com $n = 60$ e $5,1556$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 30 e $\rho = 0.9$, o erro de previsão do modelo normal foi $5,2438$ com 0% de *outliers*, $7,7058$ com 5%, $9,5320$ com 10% e $11,7486$ com 15%. Já para modelo t-Student(4) o erro foi $5,3016$ com 0% de *outliers*, $5,4202$ com 5%, $5,3427$ com 10% e $6,6279$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações a medida que a correlação entre y_I e y_S aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.2: Cenário 1: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	5,2277 (1,1135)	7,2439 (1,8971)	8,8281 (2,4978)	10,5685 (3,2213)	5,2510 (1,1206)	5,3552 (1,0605)	5,3806 (1,1163)	6,6132 (2,3588)
	0,3	5,3413 (1,1454)	7,3501 (2,1827)	9,1122 (2,9430)	10,8769 (3,6426)	5,3707 (1,1529)	5,3225 (1,1255)	5,3221 (1,2188)	6,6525 (2,4776)
	0,5	5,2737 (1,1986)	7,4093 (2,2000)	9,2145 (3,0390)	11,1546 (3,7757)	5,2970 (1,2103)	5,3550 (1,2561)	5,4061 (1,2603)	6,6270 (2,7907)
	0,7	5,2729 (1,3017)	7,6289 (2,3662)	9,5951 (3,3851)	11,5985 (4,1384)	5,2657 (1,2865)	5,2752 (1,2557)	5,3530 (1,3651)	6,7146 (2,8036)
	0,9	5,2438 (1,4051)	7,7058 (2,6032)	9,5320 (3,2350)	11,7486 (4,1039)	5,3016 (1,4375)	5,4202 (1,4538)	5,3427 (1,4533)	6,6279 (2,9094)
40	0	5,2336 (0,9200)	7,7000 (1,9025)	8,9200 (2,3315)	10,2414 (2,7192)	5,2542 (0,9335)	5,2835 (0,9003)	5,3247 (0,9299)	6,3606 (2,0854)
	0,3	5,2023 (0,9290)	7,8470 (2,0672)	9,2097 (2,5820)	10,4113 (3,0580)	5,2415 (0,9194)	5,2421 (0,9451)	5,3283 (0,9807)	6,3494 (2,1963)
	0,5	5,2073 (0,9710)	7,8939 (2,1359)	9,1946 (2,5958)	10,6523 (3,0456)	5,2798 (1,0010)	5,2327 (0,9750)	5,3737 (1,0564)	6,3617 (2,1619)
	0,7	5,1844 (1,0456)	8,0613 (2,2959)	9,6231 (2,9754)	11,0047 (3,2285)	5,2517 (1,0912)	5,2626 (1,1025)	5,3787 (1,1035)	6,1842 (2,0674)
	0,9	5,2549 (1,2144)	8,0819 (2,3910)	9,6155 (2,9543)	10,9849 (3,3121)	5,2316 (1,2260)	5,2449 (1,2427)	5,3460 (1,2485)	6,2674 (2,3533)
60	0	5,1481 (0,6899)	6,4666 (1,1448)	8,7587 (1,8926)	10,4603 (2,3168)	5,1738 (0,7068)	5,2118 (0,7279)	5,2699 (0,7660)	7,2240 (2,3751)
	0,3	5,1523 (0,7444)	6,5509 (1,2544)	8,9626 (2,0379)	10,7031 (2,5858)	5,1834 (0,7573)	5,2094 (0,7147)	5,2618 (0,7676)	7,1599 (2,2296)
	0,5	5,1720 (0,7954)	6,5956 (1,3409)	9,1478 (2,1293)	10,9341 (2,6201)	5,1753 (0,8016)	5,1594 (0,7937)	5,2752 (0,8563)	6,9934 (2,3916)
	0,7	5,1563 (0,8350)	6,6466 (1,4265)	9,2878 (2,3067)	11,2387 (2,8890)	5,2016 (0,8450)	5,2301 (0,8447)	5,2797 (0,9639)	6,7680 (2,3120)
	0,9	5,1404 (0,9417)	6,6990 (1,5554)	9,2344 (2,3577)	11,3957 (2,9638)	5,1765 (0,9361)	5,1500 (0,9570)	5,2861 (0,9856)	6,8035 (2,3074)
100	0	5,1287 (0,5633)	6,4494 (0,9255)	7,9226 (1,2625)	9,9057 (1,6579)	5,1467 (0,5514)	5,1556 (0,5459)	5,1232 (0,5312)	5,9881 (1,0967)
	0,3	5,0941 (0,5643)	6,5269 (0,9475)	8,0503 (1,4314)	10,0945 (1,8734)	5,1213 (0,5846)	5,1215 (0,5601)	5,1426 (0,5587)	5,9056 (1,1048)
	0,5	5,1041 (0,6178)	6,5889 (1,0327)	8,0789 (1,4005)	10,3749 (1,9879)	5,1217 (0,6424)	5,1309 (0,6154)	5,1552 (0,6087)	5,8820 (1,1501)
	0,7	5,1197 (0,7039)	6,6601 (1,1502)	8,1997 (1,5733)	10,3335 (2,0020)	5,1185 (0,6638)	5,1090 (0,6517)	5,1274 (0,6586)	5,7407 (1,1238)
	0,9	5,1382 (0,6921)	6,6556 (1,2205)	8,2060 (1,5886)	10,5693 (2,2165)	5,1305 (0,7201)	5,1569 (0,7140)	5,1318 (0,7391)	5,6565 (1,1163)

4.2.2 Cenário 2

Nesse cenário iremos utilizar a relação de quando $\phi_{11} > \phi_{22}$ temos que $\phi_{12}^* < 0 \ \forall \phi_{12} \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11} = 15$, $\phi_{22} = 10$ e $\rho = 0; 0,3; 0.5; 0,7$ e $0,9$ entre as variáveis y_I e y_S. Dado que temos os valores de ϕ_{11}, ϕ_{22} e ϕ_{12} e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}^*, ϕ_{22}^*, ϕ_{12}^* e ρ^* foram obtidos e podem ser vistos na Tabela 4.3. Note que nesse cenário as variáveis y_c e y_a são correlacionadas e negativamente, os valores de ϕ_{11}^* aumentam e ϕ_{22}^* diminuem a medida que a correlação entre y_I e y_S cresce.

Tabela 4.3: Valores assumidos pelos parâmetros ϕ_{11}^*, ϕ_{22}^*, ϕ_{12}^* e ρ^* no Cenário 2

Parâmetros	ρ				
	0	0,3	0,5	0,7	0,9
ϕ_{11}^*	6,25	8,08	9,31	10,53	11,76
ϕ_{22}^*	25,00	17,65	12,75	7,85	2,95
ϕ_{12}^*	-2,50	-2,50	-2,50	-2,50	-2,50
ρ^*	-0,20	-0,21	-0,23	-0,27	-0,42

A Tabela 4.4 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 2 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_I e y_S aumenta. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 0% e um tamanho de amostra 100, notou-se que com $\rho = 0$ o erro de previsão foi $5,6922$, com $\rho = 0,3$ o erro foi $5,6957$, com $\rho = 0,5$ foi $5,6770$, com $\rho = 0,7$ foi $5,6793$ e com $\rho = 0,9$ foi $5,6594$, indicando pequenas variações a medida que ρ crescia. Quando fixamos um percentual de *outliers* e correlação em 10% e $\rho = 0,5$ respectivamente, o erro de previsão foi $9,8765$ com $n = 30$, $9,7665$ com $n = 40$, $9,6761$ com $n = 60$ e $8,5769$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_I e y_S aumenta. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 15% e um tamanho de amostra 30, notou-se que com $\rho = 0$ o erro de previsão foi $7,3371$, com $\rho = 0,3$ o erro foi $7,3395$, com $\rho = 0,5$ foi $7,3171$, com $\rho = 0,7$ foi $7,3813$ e com $\rho = 0,9$ foi $7,3508$, indicando pequenas variações a medida que ρ aumenta. Já para um percentual de *outliers* e correlação fixados em 5%

e $\rho = 0$ respectivamente, o erro de previsão foi $5,9620$ com $n = 30$, $5,8883$ com $n = 40$, $5,7964$ com $n = 60$ e $5,7329$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 40 e $\rho = 0$, o erro de previsão do modelo normal foi $5,8080$ com 0% de *outliers*, $8,2313$ com 5%, $9,4472$ com 10% e $10,7021$ com 15%. Já para modelo t-Student(4) o erro foi $5,8484$ com 0% de *outliers*, $5,8883$ com 5%, $5,9476$ com 10% e $6,9994$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações a medida que a correlação entre y_I e y_S aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.4: Cenário 2: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	5,8152 (1,2384)	7,8266 (2,0719)	9,3870 (2,6898)	11,0939 (3,3687)	5,8456 (1,2573)	5,9620 (1,1936)	6,0117 (1,2493)	7,3371 (2,5354)
	0,3	5,9417 (1,2739)	7,9569 (2,3275)	9,7330 (3,1461)	11,4836 (3,7289)	5,9685 (1,2883)	5,9320 (1,2659)	5,9352 (1,3616)	7,3395 (2,7258)
	0,5	5,8737 (1,3361)	8,0238 (2,3192)	9,8765 (3,2888)	11,9057 (4,0275)	5,8997 (1,3458)	5,9606 (1,3917)	6,0457 (1,4692)	7,3171 (2,6718)
	0,7	5,8525 (1,4345)	8,2487 (2,5213)	10,2150 (3,5295)	12,2043 (4,2994)	5,8757 (1,4415)	5,8741 (1,3976)	5,9690 (1,5241)	7,3813 (3,0590)
	0,9	5,8412 (1,5795)	8,2515 (2,7259)	10,1535 (3,3537)	12,3584 (4,2410)	5,8832 (1,5837)	6,0202 (1,5984)	5,9597 (1,6385)	7,3508 (3,1699)
40	0	5,8080 (1,0129)	8,2313 (2,0457)	9,4472 (2,4202)	10,7021 (2,7457)	5,8484 (1,0068)	5,8883 (1,0126)	5,9476 (1,0703)	6,9994 (2,3017)
	0,3	5,7441 (1,0365)	8,4373 (2,1920)	9,7912 (2,5915)	11,0335 (3,0208)	5,8366 (1,0760)	5,8377 (1,0531)	5,9552 (1,1217)	7,0031 (2,3278)
	0,5	5,8055 (1,0741)	8,4898 (2,2657)	9,7663 (2,6092)	11,3972 (3,3018)	5,8333 (1,1216)	5,8407 (1,0913)	5,9558 (1,1550)	6,8114 (2,1150)
	0,7	5,8171 (1,1952)	8,6691 (2,5792)	10,1412 (3,0781)	11,4413 (3,5785)	5,8382 (1,2158)	5,8446 (1,2354)	5,9499 (1,3123)	6,8515 (2,1913)
	0,9	5,7768 (1,3104)	8,7080 (2,5756)	10,1976 (3,0438)	11,6295 (3,5096)	5,8092 (1,3333)	5,8332 (1,3900)	5,9157 (1,3786)	6,9035 (2,4532)
60	0	5,7153 (0,7865)	6,9862 (1,1783)	9,3778 (1,9433)	10,9447 (2,4913)	5,7629 (0,8224)	5,7964 (0,8161)	5,8436 (0,8618)	7,7445 (2,3150)
	0,3	5,6869 (0,7865)	7,0385 (1,2828)	9,5027 (2,1200)	11,2384 (2,5627)	5,7904 (0,8614)	5,7938 (0,8003)	5,8700 (0,9118)	7,5767 (2,3755)
	0,5	5,7163 (0,8763)	7,1221 (1,4246)	9,6761 (2,3260)	11,3766 (2,5242)	5,7478 (0,8993)	5,7376 (0,8867)	5,8655 (0,9824)	7,4487 (2,3932)
	0,7	5,7616 (0,9626)	7,2208 (1,5208)	9,8645 (2,3684)	11,7267 (2,9169)	5,7678 (0,9491)	5,8166 (0,9399)	5,8589 (1,0462)	7,5141 (2,4041)
	0,9	5,7583 (1,0887)	7,2414 (1,6047)	10,1036 (2,4738)	11,7636 (2,9649)	5,7558 (1,0777)	5,7156 (1,0679)	5,8571 (1,0942)	7,3624 (2,4507)
100	0	5,6922 (0,5913)	6,9389 (0,9032)	8,3687 (1,3256)	10,3996 (1,8342)	5,7100 (0,6185)	5,7329 (0,6089)	5,7027 (0,6342)	6,4362 (1,1252)
	0,3	5,6957 (0,6501)	7,0130 (1,0175)	8,4958 (1,3866)	10,5870 (1,8896)	5,6838 (0,6608)	5,6934 (0,6184)	5,7230 (0,6432)	6,3674 (1,0402)
	0,5	5,6770 (0,7204)	7,0974 (1,0879)	8,5769 (1,6189)	10,7291 (1,9347)	5,7146 (0,6636)	5,7028 (0,6873)	5,6568 (0,6832)	6,3422 (1,0999)
	0,7	5,6793 (0,7332)	7,0617 (1,1368)	8,6440 (1,6473)	10,9742 (2,1220)	5,7136 (0,7528)	5,6760 (0,7231)	5,6975 (0,7300)	6,3121 (1,0733)
	0,9	5,6594 (0,7948)	7,1617 (1,2571)	8,8348 (1,6593)	11,1047 (2,1518)	5,6940 (0,8080)	5,7260 (0,7967)	5,6913 (0,8240)	6,2944 (1,2465)

4.2.3 Cenário 3

Nesse cenário iremos utilizar a relação de quando $\phi_{11} < \phi_{22}$ temos que $\phi_{12}^* > 0 \ \forall \phi_{12} \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11} = 10$, $\phi_{22} = 15$ e $\rho = 0; 0,3; 0.5; 0,7$ e $0,9$ entre as variáveis y_I e y_S. Dado que temos os valores de ϕ_{11}, ϕ_{22} e ϕ_{12} e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}^*, ϕ_{22}^*, ϕ_{12}^* e ρ^* foram obtidos e podem ser vistos na Tabela 4.5. Note que nesse cenário as variáveis y_c e y_a são correlacionadas e positivamente, os valores de ϕ_{11}^* aumentam e ϕ_{22}^* diminuem a medida que a correlação entre y_I e y_S cresce.

Tabela 4.5: Valores assumidos pelos parâmetros ϕ_{11}^*, ϕ_{22}^*, ϕ_{12}^* e ρ^* no Cenário 3

Parâmetros	ρ				
	0	0,3	0,5	0,7	0,9
ϕ_{11}^*	6,25	8,08	9,31	10,53	11,76
ϕ_{22}^*	25,00	17,65	12,75	7,85	2,95
ϕ_{12}^*	2,50	2,50	2,50	2,50	2,50
ρ^*	0,20	0,21	0,23	0,27	0,42

A Tabela 4.6 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 3 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_I e y_S aumenta. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 0% e um tamanho de amostra 40, notou-se que com $\rho = 0$ o erro de previsão foi $5,8639$, com $\rho = 0,3$ o erro foi $5,8174$, com $\rho = 0,5$ foi $5,8436$, com $\rho = 0,7$ foi $5,8924$ e com $\rho = 0,9$ foi $5,8178$, indicando pequenas variações a medida que ρ crescia. Quando fixamos um percentual de *outliers* e correlação em 10% e $\rho = 0$ respectivamente, o erro de previsão foi $9,3945$ com $n = 30$, $9,5260$ com $n = 40$, $9,3299$ com $n = 60$ e $8,2740$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_I e y_S aumenta. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 5% e um tamanho de amostra 60, notou-se que com $\rho = 0$ o erro de previsão foi $5,7988$, com $\rho = 0,3$ o erro foi $5,7933$, com $\rho = 0,5$ foi $5,7355$, com $\rho = 0,7$ foi $5,8135$ e com $\rho = 0,9$ foi $5,7229$, indicando pequenas variações no erro a medida que ρ aumenta. Já para um percentual de *outliers* e correlação fixados

em 10% e $\rho = 0,3$ respectivamente, o erro de previsão foi $5,9523$ com $n = 30$, $5,9346$ com $n = 40$, $5,8391$ com $n = 60$ e $5,7233$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 100 e $\rho = 0,9$, o erro de previsão do modelo normal foi $5,7026$ com 0% de *outliers*, $7,1736$ com 5%, $8,6840$ com 10% e $10,9737$ com 15%. Já para modelo t-Student(4) o erro foi $5,7109$ com 0% de *outliers*, $5,7294$ com 5%, $5,6944$ com 10% e $6,3558$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações a medida que a correlação entre y_I e y_S aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.6: Cenário 3: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	5,8121 (1,2346)	7,8351 (2,0671)	9,3945 (2,6685)	11,1777 (3,4282)	5,8497 (1,2503)	5,9592 (1,2017)	6,0160 (1,2856)	7,4495 (2,6831)
	0,3	5,9481 (1,2866)	7,9456 (2,3413)	9,7297 (3,0837)	11,4826 (3,6662)	5,9791 (1,2968)	5,9254 (1,2596)	5,9523 (1,3669)	7,3558 (2,6474)
	0,5	5,8779 (1,3377)	8,0046 (2,3359)	9,9133 (3,3177)	11,8994 (4,0393)	5,9050 (1,3435)	5,9677 (1,4030)	6,0585 (1,4531)	7,3672 (2,8307)
	0,7	5,8564 (1,4182)	8,1898 (2,5710)	10,2355 (3,5705)	12,2850 (4,3171)	5,8753 (1,4308)	5,8648 (1,4166)	5,9764 (1,5410)	7,4150 (3,0032)
	0,9	5,8644 (1,5900)	8,2970 (2,7185)	10,0424 (3,4303)	12,4071 (4,2613)	5,8876 (1,5975)	6,0139 (1,6134)	5,9535 (1,6445)	7,4681 (3,1546)
40	0	5,8639 (1,0321)	8,2475 (1,9538)	9,5260 (2,4141)	10,7850 (2,8489)	5,8667 (1,0082)	5,8684 (1,0016)	5,9338 (1,0689)	7,0267 (2,0827)
	0,3	5,8174 (1,0329)	8,4210 (2,1623)	9,5918 (2,6481)	11,1214 (3,1675)	5,8115 (1,0493)	5,8242 (1,0621)	5,9346 (1,0980)	6,9805 (2,5577)
	0,5	5,8436 (1,1219)	8,3656 (2,2331)	9,8621 (2,8556)	11,2179 (3,2749)	5,7764 (1,1226)	5,8012 (1,0837)	5,9856 (1,1955)	6,9500 (2,3060)
	0,7	5,8924 (1,2358)	8,5928 (2,4393)	10,0460 (2,8127)	11,5457 (3,4770)	5,8635 (1,2450)	5,8674 (1,2165)	5,9880 (1,2450)	6,9936 (2,4125)
	0,9	5,8178 (1,3330)	8,5750 (2,5199)	10,0711 (3,0835)	11,7299 (3,7830)	5,8461 (1,3185)	5,8284 (1,3678)	5,9280 (1,3724)	6,8836 (2,3711)
60	0	5,7128 (0,7970)	6,9066 (1,1920)	9,3299 (2,0070)	10,9843 (2,4451)	5,7445 (0,7750)	5,7988 (0,8107)	5,8630 (0,8432)	7,6354 (2,1693)
	0,3	5,7902 (0,8354)	7,0495 (1,2863)	9,4668 (2,1541)	11,2644 (2,6077)	5,7708 (0,8408)	5,7933 (0,7988)	5,8391 (0,8524)	7,5558 (2,2892)
	0,5	5,7875 (0,9006)	7,1278 (1,4552)	9,7402 (2,2167)	11,5868 (2,8106)	5,7920 (0,9002)	5,7355 (0,8820)	5,8649 (0,9745)	7,5224 (2,3228)
	0,7	5,7255 (0,9521)	7,2766 (1,5313)	9,9284 (2,3824)	11,7503 (3,0240)	5,7442 (0,9794)	5,8135 (0,9413)	5,8716 (1,0551)	7,4132 (2,4464)
	0,9	5,6941 (1,0878)	7,1770 (1,6393)	9,9524 (2,4326)	12,1051 (3,1979)	5,6914 (1,0441)	5,7229 (1,0560)	5,8729 (1,0870)	7,5400 (2,4569)
100	0	5,6934 (0,6185)	6,9420 (0,9515)	8,2740 (1,2139)	10,3387 (1,7555)	5,6876 (0,5992)	5,7351 (0,6110)	5,7007 (0,5975)	6,4196 (1,0392)
	0,3	5,6886 (0,6162)	6,9759 (0,9910)	8,5202 (1,3546)	10,6146 (1,9397)	5,6763 (0,6302)	5,6997 (0,6317)	5,7233 (0,6276)	6,3277 (1,0727)
	0,5	5,7049 (0,6761)	7,0068 (1,0596)	8,6025 (1,5640)	10,7918 (1,9927)	5,6836 (0,6893)	5,7066 (0,6858)	5,7373 (0,6848)	6,3572 (1,1113)
	0,7	5,7103 (0,7476)	7,0979 (1,1479)	8,6576 (1,5444)	11,0581 (2,0859)	5,6953 (0,6791)	5,6819 (0,7306)	5,7014 (0,7336)	6,3252 (1,1269)
	0,9	5,7026 (0,8325)	7,1736 (1,2370)	8,6840 (1,6321)	10,9737 (2,0960)	5,7109 (0,8201)	5,7294 (0,7909)	5,6944 (0,8096)	6,3558 (1,2314)

4.2.4 Cenário 4

Nesse cenário iremos utilizar a relação de quando $\phi_{11}^* = \phi_{22}^*$ temos que $\phi_{12} > 0 \ \forall \phi_{12}^* \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11}^* = \phi_{22}^* = 10$ e $\rho^* = (0; 0,3; 0,5; 0,7$ e $0,9)$ entre as variáveis y_c e y_a. Dado que temos os valores de ϕ_{11}^*, ϕ_{22}^* e ϕ_{12}^* e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ foram obtidos e podem ser vistos na Tabela 4.7. Note que nesse cenário as variáveis y_I e y_S são correlacionadas e positivamente, os valores de ϕ_{11} diminuem e ϕ_{22} aumentam a medida que a correlação entre y_c e y_a cresce.

Tabela 4.7: Valores assumidos pelos parâmetros ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ no Cenário 3

Parâmetros	ρ^*				
	0	0,3	0,5	0,7	0,9
ϕ_{11}	12,50	9,50	7,50	5,50	3,50
ϕ_{22}	12,50	15,50	17,50	19,50	21,50
ϕ_{12}	7,50	7,50	7,50	7,50	7,50
ρ	0,60	0,62	0,65	0,72	0,86

A Tabela 4.8 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 4 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 5% e um tamanho de amostra 30, notou-se que com $\rho^* = 0$ o erro de previsão foi $8,1140$, com $\rho^* = 0,3$ o erro foi $8,0580$, com $\rho^* = 0,5$ foi $8,0294$, com $\rho^* = 0,7$ foi $8,0387$ e com $\rho^* = 0,9$ foi $7,8852$, indicando variações a medida que ρ^* crescia. Quando fixamos um percentual de *outliers* e correlação em 0% e $\rho^* = 0$ respectivamente, o erro de previsão foi $5,8146$ com $n = 30$, $5,8570$ com $n = 40$, $5,7629$ com $n = 60$ e $5,7575$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 5% e um tamanho de amostra 30, notou-se que com $\rho^* = 0$ o erro de previsão foi $5,9774$, com $\rho^* = 0,3$ o erro foi $5,8812$, com $\rho^* = 0,5$ foi $5,8554$, com $\rho^* = 0,7$ foi $5,5983$ e com $\rho^* = 0,9$ foi $5,4206$, indicando uma diminuição do erro a medida que ρ^* aumenta. Já para um percentual de *outliers* e correlação fixados em 5% e $\rho^* = 0,3$ respectivamente, o erro de previsão foi $5,8812$ com $n = 30$, $5,8037$ com $n = 40$, $5,7790$

com $n = 60$ e $5,6827$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 40 e $\rho^* = 0,7$, o erro de previsão do modelo normal foi $5,5641$ com 0% de *outliers*, $8,4875$ com 5%, $9,9350$ com 10% e $11,3100$ com 15%. Já para modelo t-Student(4) o erro foi $5,6421$ com 0% de *outliers*, $5,6397$ com 5%, $5,7027$ com 10% e $6,7584$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.8: Cenário 4: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	5,8146 (1,3803)	8,1140 (2,4643)	9,8447 (3,1368)	11,7970 (4,0146)	5,8473 (1,3785)	5,9774 (1,3503)	5,9865 (1,4199)	7,0333 (2,6640)
	0,3	5,9039 (1,4247)	8,0580 (2,5789)	9,9837 (3,3756)	11,7746 (3,8961)	5,9421 (1,4352)	5,8812 (1,3887)	5,9601 (1,5239)	7,2441 (2,7238)
	0,5	5,7755 (1,4108)	8,0294 (2,3934)	10,0162 (3,3339)	11,9963 (4,0058)	5,7915 (1,4083)	5,8554 (1,4743)	5,9467 (1,4771)	7,3055 (2,8987)
	0,7	5,5665 (1,3407)	8,0387 (2,4809)	10,1124 (3,3690)	12,1308 (3,9751)	5,5941 (1,3598)	5,5983 (1,3967)	5,7447 (1,5679)	7,2681 (2,8976)
	0,9	5,3408 (1,4160)	7,8852 (2,2626)	9,7561 (3,0146)	11,9228 (3,7392)	5,3545 (1,4128)	5,4206 (1,4547)	5,4220 (1,4919)	6,9211 (2,8347)
40	0	5,8570 (1,1430)	8,6860 (2,5463)	10,0262 (2,8942)	11,3499 (3,2720)	5,9091 (1,1668)	5,9070 (1,1400)	5,9299 (1,2046)	6,7779 (2,0056)
	0,3	5,7303 (1,1317)	8,7023 (2,3837)	9,9551 (2,8988)	11,6330 (3,4609)	5,8069 (1,1503)	5,8037 (1,1814)	5,9351 (1,2439)	6,9885 (2,2067)
	0,5	5,6510 (1,1313)	8,4820 (2,4207)	9,8197 (2,7159)	11,3129 (3,1976)	5,7514 (1,1604)	5,6634 (1,1446)	5,8764 (1,2685)	6,9343 (2,3836)
	0,7	5,5641 (1,1430)	8,4875 (2,3539)	9,9350 (2,6417)	11,3100 (3,2477)	5,6421 (1,2043)	5,6397 (1,1726)	5,7027 (1,2087)	6,7584 (2,3563)
	0,9	5,2914 (1,1733)	8,3809 (2,2397)	9,8134 (2,5550)	11,1688 (2,9953)	5,3380 (1,2372)	5,3359 (1,1986)	5,3452 (1,1880)	6,3322 (2,0356)
60	0	5,7629 (0,9196)	7,1870 (1,4340)	9,8052 (2,2837)	11,5330 (2,7489)	5,7745 (0,9072)	5,8416 (0,9253)	5,8500 (0,9657)	7,2047 (2,0789)
	0,3	5,6863 (0,9367)	7,2222 (1,5546)	9,7128 (2,2485)	11,6054 (2,8230)	5,7959 (0,9268)	5,7790 (0,8827)	5,8302 (0,9466)	7,4699 (2,3472)
	0,5	5,6553 (0,9664)	7,1077 (1,4859)	9,7219 (2,2488)	11,6079 (2,8519)	5,7020 (0,9533)	5,6227 (0,9277)	5,7700 (0,9930)	7,3821 (2,3798)
	0,7	5,5197 (0,9873)	7,0076 (1,4263)	9,7829 (2,3039)	11,5884 (2,6158)	5,4868 (0,9387)	5,5623 (0,9292)	5,6352 (1,0108)	7,3074 (2,2582)
	0,9	5,2372 (0,9500)	6,9102 (1,3228)	9,5372 (2,0380)	11,5015 (2,5560)	5,2307 (0,9899)	5,2549 (0,9370)	5,3289 (0,9951)	7,1408 (2,1828)
100	0	5,7575 (0,7242)	7,1736 (1,2191)	8,7753 (1,6513)	10,9610 (2,1010)	5,7742 (0,7157)	5,7634 (0,7222)	5,6898 (0,6982)	6,2386 (1,0900)
	0,3	5,6781 (0,7099)	7,1316 (1,1259)	8,6698 (1,6097)	10,9106 (2,0417)	5,6921 (0,7141)	5,6827 (0,7042)	5,7084 (0,6880)	6,3253 (1,1524)
	0,5	5,5792 (0,7203)	6,9587 (1,0938)	8,5671 (1,4623)	10,8638 (2,0858)	5,6083 (0,7147)	5,6065 (0,7184)	5,6421 (0,7259)	6,2664 (1,1694)
	0,7	5,4370 (0,7351)	7,0449 (1,1394)	8,4923 (1,4961)	10,8299 (1,9330)	5,3964 (0,7223)	5,4456 (0,7232)	5,4745 (0,7229)	6,1662 (1,1206)
	0,9	5,2127 (0,7456)	6,8890 (1,0346)	8,5491 (1,4647)	10,7781 (1,8329)	5,2313 (0,7447)	5,2560 (0,7086)	5,2226 (0,7289)	5,8411 (1,1089)

4.2.5 Cenário 5

Nesse cenário iremos utilizar a relação de quando $\phi_{11}^* > \phi_{22}^*$ temos que $\phi_{12} > 0 \ \forall \phi_{12}^* \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11}^* = 15$, $\phi_{22}^* = 10$ e $\rho^* = (0; 0,3; 0,5; 0,7 \text{ e } 0,9)$ entre as variáveis y_c e y_a. Dado que temos os valores de ϕ_{11}^*, ϕ_{22}^* e ϕ_{12}^* e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ foram obtidos e podem ser vistos na Tabela 4.9. Note que nesse cenário as variáveis y_I e y_S são correlacionadas e positivamente, os valores de ϕ_{11} diminuem e ϕ_{22} aumentam a medida que a correlação entre y_c e y_a cresce.

Tabela 4.9: Valores assumidos pelos parâmetros ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ no Cenário 5

Parâmetros	ρ^*				
	0	0,3	0,5	0,7	0,9
ϕ_{11}	17,50	13,82	11,37	8,92	6,47
ϕ_{22}	17,50	21,17	23,62	26,07	28,52
ϕ_{12}	12,50	12,50	12,50	12,50	12,50
ρ	0,71	0,73	0,76	0,82	0,92

A Tabela 4.10 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 5 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 15% e um tamanho de amostra 60, notou-se que com $\rho^* = 0$ o erro de previsão foi $13,0538$, com $\rho^* = 0,3$ o erro foi $13,0559$, com $\rho^* = 0,5$ foi $13,0082$, com $\rho^* = 0,7$ foi $12,8512$ e com $\rho^* = 0,9$ foi $12,6902$, indicando uma diminuição do erro a medida que ρ^* crescia. Quando fixamos um percentual de *outliers* e correlação em 0% e $\rho^* = 0,5$ respectivamente, o erro de previsão foi $5,8507$ com $n = 30$, $6,8129$ com $n = 40$, $6,7464$ com $n = 60$ e $6,6688$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 0% e um tamanho de amostra 100, notou-se que com $\rho^* = 0$ o erro de previsão foi $6,8281$, com $\rho^* = 0,3$ o erro foi $6,7407$, com $\rho^* = 0,5$ foi $6,6410$, com $\rho^* = 0,7$ foi $6,5246$ e com $\rho^* = 0,9$ foi $6,3671$, indicando uma diminuição do erro a medida que ρ^* aumenta. Já para um percentual de *outliers* e correlação fixados em 10% e $\rho^* = 0$

respectivamente, o erro de previsão foi $7,0759$ com $n = 30$, $6,9741$ com $n = 40$, $6,8927$ com $n = 60$ e $6,7258$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 30 e $\rho^* = 0,9$, o erro de previsão do modelo normal foi $6,4751$ com 0% de *outliers*, $9,0204$ com 5%, $10,9295$ com 10% e $13,1182$ com 15%. Já para modelo t-Student(4) o erro foi $6,4820$ com 0% de *outliers*, $6,5634$ com 5%, $6,5271$ com 10% e $7,6999$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.10: Cenário 5: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	6,8851 (1,7142)	9,4066 (2,9323)	11,2099 (3,6928)	13,1875 (4,4674)	6,9114 (1,7138)	7,0640 (1,6899)	7,0759 (1,7651)	7,9502 (2,7865)
	0,3	6,9947 (1,7685)	9,2988 (3,0473)	11,2909 (3,8947)	13,2578 (4,3966)	7,0330 (1,7836)	6,9530 (1,7299)	7,0459 (1,8728)	8,0851 (2,8334)
	0,5	6,8507 (1,7435)	9,1917 (2,7464)	11,2295 (3,7493)	13,3214 (4,5256)	6,8796 (1,7505)	6,9642 (1,8324)	7,0304 (1,7938)	8,1451 (3,0357)
	0,7	6,6718 (1,6827)	9,1674 (2,8187)	11,3162 (3,8704)	13,4259 (4,4060)	6,6894 (1,6994)	6,6871 (1,7386)	6,8235 (1,8591)	7,9594 (2,7727)
	0,9	6,4751 (1,7537)	9,0204 (2,5991)	10,9295 (3,4264)	13,1182 (4,1581)	6,4820 (1,7534)	6,5634 (1,7906)	6,5271 (1,8325)	7,6999 (2,8200)
40	0	6,9538 (1,4438)	9,9377 (2,7402)	11,2110 (3,2334)	12,7893 (3,9105)	6,9860 (1,4698)	6,9860 (1,4208)	6,9741 (1,4681)	7,5046 (2,0224)
	0,3	6,8327 (1,4074)	9,5507 (2,6038)	11,2734 (3,3143)	12,6213 (3,6937)	6,8614 (1,4746)	6,8705 (1,4707)	6,9843 (1,4767)	7,6691 (2,2431)
	0,5	6,8129 (1,4683)	9,7704 (2,7455)	11,2219 (3,3613)	12,6755 (3,7942)	6,7504 (1,4253)	6,7308 (1,4312)	6,9532 (1,5398)	7,6544 (2,1718)
	0,7	6,6407 (1,4895)	9,4661 (2,6434)	10,9866 (3,0129)	12,7126 (3,6193)	6,6962 (1,4499)	6,7306 (1,4537)	6,7849 (1,4934)	7,4235 (2,1113)
	0,9	6,4289 (1,4678)	9,4898 (2,5917)	10,9651 (2,9288)	12,6069 (3,5666)	6,3944 (1,4915)	6,4515 (1,4908)	6,4606 (1,4542)	7,2164 (2,1179)
60	0	6,7847 (1,1212)	8,5114 (1,9496)	11,0912 (2,7598)	13,0538 (3,3084)	6,7875 (1,1604)	6,9152 (1,1514)	6,8927 (1,1831)	7,9123 (2,0915)
	0,3	6,7841 (1,1039)	8,1674 (1,7186)	10,9245 (2,6274)	13,0559 (3,1768)	6,7944 (1,1328)	6,8495 (1,0975)	6,8767 (1,1618)	8,2366 (2,4247)
	0,5	6,7464 (1,1504)	8,2040 (1,7199)	11,0723 (2,6199)	13,0082 (3,0815)	6,7532 (1,1511)	6,6775 (1,1609)	6,8361 (1,2159)	8,0634 (2,1929)
	0,7	6,5799 (1,1450)	8,1808 (1,6499)	10,7645 (2,4407)	12,8512 (2,9406)	6,5091 (1,2027)	6,6518 (1,1558)	6,7048 (1,2314)	7,9149 (2,0144)
	0,9	6,3293 (1,1818)	8,0558 (1,6015)	10,7416 (2,4514)	12,6902 (2,8791)	6,3633 (1,1590)	6,3592 (1,1585)	6,4446 (1,2246)	7,6853 (2,0599)
100	0	6,7353 (0,8894)	8,3457 (1,3855)	10,1809 (1,9531)	12,2951 (2,3904)	6,8281 (0,8250)	6,8205 (0,9041)	6,7258 (0,8734)	7,2238 (1,1728)
	0,3	6,7161 (0,8290)	8,1744 (1,2601)	9,6033 (1,7380)	11,9757 (2,2099)	6,7407 (0,8996)	6,7331 (0,8762)	6,7608 (0,8498)	7,2335 (1,2115)
	0,5	6,6688 (0,8882)	8,0320 (1,3052)	9,6930 (1,7150)	12,0737 (2,2685)	6,6410 (0,8987)	6,6644 (0,8945)	6,7014 (0,9044)	7,0961 (1,1251)
	0,7	6,5277 (0,8960)	7,9035 (1,2587)	9,7119 (1,7221)	11,9279 (2,1372)	6,5246 (0,9044)	6,5167 (0,8928)	6,5462 (0,8977)	7,0572 (1,1318)
	0,9	6,3174 (0,8905)	7,9001 (1,2466)	9,5280 (1,5749)	12,0043 (2,1080)	6,3671 (0,9245)	6,3723 (0,8791)	6,3279 (0,9029)	6,7832 (1,0524)

4.2.6 Cenário 6

Nesse cenário iremos utilizar a relação de quando $\phi_{22}^* > 4\phi_{11}^*$ temos que $\phi_{12} < 0 \ \forall \phi_{12}^* \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11}^* = 3$, $\phi_{22}^* = 15$ e $\rho^* = (0; 0,3; 0,5; 0,7 \text{ e } 0,9)$ entre as variáveis y_c e y_a. Dado que temos os valores de ϕ_{11}^*, ϕ_{22}^* e ϕ_{12}^* e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ foram obtidos e podem ser vistos na Tabela 4.11. Note que nesse cenário as variáveis y_I e y_S são correlacionadas e negativamente, os valores de ϕ_{11} diminuem e ϕ_{22} aumentam a medida que a correlação entre y_c e y_a cresce.

Tabela 4.11: Valores assumidos pelos parâmetros ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ no Cenário 6

Parâmetros	ρ^*				
	0	0,3	0,5	0,7	0,9
ϕ_{11}	6,75	4,73	3,39	2,05	0,71
ϕ_{22}	6,75	8,76	10,10	11,44	12,78
ϕ_{12}	-0,75	-0,75	-0,75	-0,75	-0,75
ρ	-0,11	-0,11	-0,12	-0,15	-0,24

A Tabela 4.12 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 6 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 15% e um tamanho de amostra 60, notou-se que com $\rho^* = 0$ o erro de previsão foi $9,5631$, com $\rho^* = 0,3$ o erro foi $9,6751$, com $\rho^* = 0,5$ foi $9,6898$, com $\rho^* = 0,7$ foi $9,6398$ e com $\rho^* = 0,9$ foi $9,5779$, indicando variações do erro a medida que ρ^* crescia. Quando fixamos um percentual de *outliers* e correlação em 0% e $\rho^* = 0,3$ respectivamente, o erro de previsão foi $4,3537$ com $n = 30$, $4,2262$ com $n = 40$, $4,1989$ com $n = 60$ e $4,1358$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 10% e um tamanho de amostra 30, notou-se que com $\rho^* = 0$ o erro de previsão foi $4,4278$, com $\rho^* = 0,3$ o erro foi $4,3554$, com $\rho^* = 0,5$ foi $4,3237$, com $\rho^* = 0,7$ foi $4,1329$ e com $\rho^* = 0,9$ foi $3,7365$, indicando uma diminuição do erro a medida que ρ^* aumenta. Já para um percentual de *outliers* e correlação fixados em 5% e $\rho^* = 0,9$ respectivamente, o erro de previsão foi $3,7051$ com $n = 30$, $3,6548$ com $n = 40$, $3,5965$

com $n = 60$ e $3,5583$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 100 e $\rho^* = 0,7$, o erro de previsão do modelo normal foi $3,8832$ com 0% de *outliers*, $5,4770$ com 5%, $6,9393$ com 10% e $9,0529$ com 15%. Já para modelo t-Student(4) o erro foi $3,8811$ com 0% de *outliers*, $3,8790$ com 5%, $3,9094$ com 10% e $5,1426$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.12: Cenário 6: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	4,2969 (0,9271)	6,3569 (1,6991)	7,8892 (2,2153)	9,6719 (2,9158)	4,3182 (0,9339)	4,4034 (0,8866)	4,4278 (0,9196)	5,8551 (2,3398)
	0,3	4,3537 (0,9291)	6,3084 (1,7778)	8,0513 (2,3766)	9,7505 (2,9755)	4,3744 (0,9311)	4,3538 (0,8940)	4,3554 (0,9633)	6,2236 (2,4873)
	0,5	4,1792 (0,8893)	6,2631 (1,6438)	8,0298 (2,2697)	9,8773 (2,9426)	4,2018 (0,9022)	4,2316 (0,9384)	4,3237 (0,9926)	6,2631 (2,7361)
	0,7	3,9912 (0,8663)	6,2522 (1,6670)	8,1509 (2,3572)	9,9999 (3,0275)	4,0062 (0,8730)	4,0463 (0,9047)	4,1329 (1,0427)	6,3041 (2,7742)
	0,9	3,6559 (0,8681)	6,2111 (1,5478)	8,0272 (2,1642)	9,9867 (2,8595)	3,6647 (0,8744)	3,7051 (0,9128)	3,7365 (0,9355)	6,1782 (2,6825)
40	0	4,3225 (0,7682)	6,7555 (1,6303)	8,1298 (1,9779)	9,2336 (2,4099)	4,3295 (0,7466)	4,3442 (0,7450)	4,3842 (0,7820)	5,5842 (2,0013)
	0,3	4,2262 (0,7300)	6,7385 (1,5840)	8,0883 (2,0659)	9,4390 (2,5231)	4,2316 (0,7458)	4,2673 (0,7508)	4,3828 (0,8182)	5,8963 (2,2186)
	0,5	4,1635 (0,7148)	6,7807 (1,6328)	8,0787 (2,0275)	9,2329 (2,4417)	4,1638 (0,7422)	4,1460 (0,7330)	4,2924 (0,8298)	5,8251 (2,1517)
	0,7	3,9600 (0,7233)	6,7711 (1,5726)	8,0457 (1,9239)	9,4042 (2,4634)	3,9958 (0,7206)	4,0484 (0,7369)	4,1115 (0,7718)	5,6480 (1,9380)
	0,9	3,6052 (0,7153)	6,6616 (1,5452)	8,1326 (1,9191)	9,3597 (2,2975)	3,6486 (0,7460)	3,6548 (0,7348)	3,6389 (0,7477)	5,4867 (2,2437)
60	0	4,2822 (0,5870)	5,5829 (0,9651)	7,8867 (1,6650)	9,5631 (2,0942)	4,2356 (0,5939)	4,2815 (0,6014)	4,3511 (0,6379)	6,8503 (2,3950)
	0,3	4,1989 (0,5843)	5,5555 (0,9648)	7,9919 (1,6056)	9,6751 (2,1033)	4,2047 (0,5978)	4,2321 (0,5823)	4,3042 (0,6384)	7,0293 (2,2357)
	0,5	4,0616 (0,5928)	5,4174 (0,9484)	7,9185 (1,6694)	9,6898 (2,1371)	4,1134 (0,6281)	4,1084 (0,5952)	4,2093 (0,6536)	6,9170 (2,2269)
	0,7	3,9329 (0,6029)	5,4517 (0,9242)	7,9904 (1,6310)	9,6398 (2,0525)	3,9378 (0,6025)	3,9799 (0,5933)	4,0218 (0,6275)	6,7567 (2,1156)
	0,9	3,5953 (0,5873)	5,3252 (0,8859)	7,9231 (1,5141)	9,5779 (2,0199)	3,5859 (0,5858)	3,5965 (0,5793)	3,6431 (0,6330)	6,8650 (2,2433)
100	0	4,1953 (0,4350)	5,5947 (0,7674)	7,0469 (1,1290)	9,0635 (1,5008)	4,2108 (0,4592)	4,2357 (0,4503)	4,2165 (0,4382)	5,3643 (1,2185)
	0,3	4,1358 (0,4510)	5,6094 (0,7879)	7,0207 (1,0829)	9,1429 (1,5534)	4,1461 (0,4606)	4,1692 (0,4530)	4,1856 (0,4515)	5,3564 (1,2210)
	0,5	4,0393 (0,4367)	5,5063 (0,7638)	7,0399 (1,0912)	9,0780 (1,4654)	4,0749 (0,4575)	4,0745 (0,4543)	4,1092 (0,4583)	5,3283 (1,2110)
	0,7	3,8832 (0,4557)	5,4770 (0,7729)	6,9393 (1,0928)	9,0529 (1,4804)	3,8811 (0,4508)	3,8790 (0,4657)	3,9094 (0,4630)	5,1426 (1,1855)
	0,9	3,5415 (0,4584)	5,3651 (0,6949)	6,8847 (1,0388)	9,0725 (1,4764)	3,5409 (0,4466)	3,5583 (0,4388)	3,5548 (0,4632)	4,6495 (1,1707)

4.2.7 Cenário 7

Nesse cenário iremos utilizar a relação de quando $\phi_{22}^* = 4\phi_{11}^*$ temos que $\phi_{12} = 0 \ \forall \phi_{12}^* \in \mathbb{R}$. Então, iremos supor que os parâmetros $\phi_{11}^* = 5$, $\phi_{22}^* = 20$ e $\rho^* = (0; 0,3; 0,5; 0,7$ e $0,9)$ entre as variáveis y_c e y_a. Dado que temos os valores de ϕ_{11}^*, ϕ_{22}^* e ϕ_{12}^* e a partir das relações definidas na seção 3.2.2, os valores de ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ foram obtidos e podem ser vistos na Tabela 4.11. Note que nesse cenário as variáveis y_I e y_S não são correlacionadas, os valores de ϕ_{11} diminuem e ϕ_{22} aumentam a medida que a correlação entre y_c e y_a cresce.

Tabela 4.13: Valores assumidos pelos parâmetros ϕ_{11}, ϕ_{22}, ϕ_{12} e ρ no Cenário 7

Parâmetros	ρ^*				
	0	0,3	0,5	0,7	0,9
ϕ_{11}	10	7	5	3	1
ϕ_{22}	10	13	15	17	19
ϕ_{12}	0	0	0	0	0
ρ	0	0	0	0	0

A Tabela 4.14 mostra o comportamento dos valores médios e desvio padrão do MADI de previsão no cenário 7 quando consideramos normalidade e t-Student(4) para a componente aleatória. Ao considerar o modelo sob normalidade podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 0% e um tamanho de amostra 40, notou-se que com $\rho^* = 0$ o erro de previsão foi $5,2606$, com $\rho^* = 0,3$ o erro foi $5,1716$, com $\rho^* = 0,5$ foi $5,0501$, com $\rho^* = 0,7$ foi $4,8607$ e com $\rho^* = 0,9$ foi $4,4011$, indicando uma diminuição do erro a medida que ρ^* crescia. Quando fixamos um percentual de *outliers* e correlação em 0% e $\rho^* = 0,9$ respectivamente, o erro de previsão foi $4,4365$ com $n = 30$, $4,4011$ com $n = 40$, $4,3459$ com $n = 60$ e $4,3168$ com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra crescia.

Nessa Tabela ao consideramos t-Student(4) para a componente aleatória, podemos observar que ao fixar um tamanho de amostra e um percentual de *outliers*, o erro de previsão sofreu variações a medida que a correlação entre y_c e y_a aumenta. Quando o percentual de *outliers* e correlação foram fixados, o erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Como por exemplo, considerando o percentual de *outliers* de 15% e um tamanho de amostra 40, notou-se que com $\rho^* = 0$ o erro de previsão foi $6,2736$, com $\rho^* = 0,3$ o erro foi $6,6337$, com $\rho^* = 0,5$ foi $6,3458$, com $\rho^* = 0,7$ foi $6,2544$ e com $\rho^* = 0,9$ foi $6,1529$, indicando variações do erro a medida que ρ^* aumenta. Já para um percentual de *outliers* e correlação fixados em 5% e $\rho^* = 0$ respectivamente, o erro de previsão foi $5,3552$ com $n = 30$, $5,2835$ com $n = 40$, $5,2118$ com $n = 60$ e $5,1556$

com $n = 100$, indicando uma diminuição dos valores dos erros a medida que o tamanho da amostra cresce.

Ao verificar o erro de previsão MADI nos dois modelos, notou que com o aumento do percentual de *outliers* o modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia. Esses comportamentos podem ser verificados ao considerar um tamanho de amostra 30 e $\rho^* = 0,5$, o erro de previsão do modelo normal foi $5,0889$ com 0% de *outliers*, $7,1094$ com 5%, $8,8835$ com 10% e $10,7219$ com 15%. Já para modelo t-Student(4) o erro foi $5,1137$ com 0% de *outliers*, $5,1534$ com 5%, $5,2711$ com 10% e $6,9262$ com 15%, mostrando que o modelo normal sofreu uma influência maior com o aumento do percentual de *outliers*.

Ao analisar o desvio padrão do erro MADI de previsão desses modelos, verificou um comportamento semelhante ao erro médio de previsão. Ao fixar um tamanho de amostra e um percentual de *outliers*, o desvio do erro de previsão sofreu variações ou em alguns caso nenhuma a medida que a correlação entre y_c e y_a aumenta. Já com um percentual de *outliers* e correlação fixados, o desvio do erro de previsão diminuiu a medida que o tamanho da amostra aumenta. Com relação ao aumento do percentual de *outliers*, o desvio padrão do modelo normal sofreu uma influência maior que o modelo t-Student(4) a medida que esse percentual crescia.

Tabela 4.14: Cenário 7: Comportamento da Média e Desvio Padrão (entre parênteses) do MADI de previsão no modelo Normal e t-Student(4) com o aumento do percentual de *outlier*

		Normal				t-Student(4)			
n	ρ	0%	5%	10%	15%	0%	5%	10%	15%
30	0	5,2277 (1,1135)	7,2439 (1,8971)	8,8281 (2,4978)	10,5685 (3,2213)	5,2510 (1,1206)	5,3552 (1,0605)	5,3806 (1,1163)	6,6132 (2,3588)
	0,3	5,2942 (1,1223)	7,1796 (1,9892)	8,8807 (2,6424)	10,5640 (3,2124)	5,3219 (1,1263)	5,2811 (1,0814)	5,3016 (1,1738)	6,8849 (2,4086)
	0,5	5,0889 (1,0742)	7,1094 (1,8562)	8,8835 (2,5789)	10,7219 (3,2489)	5,1137 (1,0889)	5,1534 (1,1317)	5,2711 (1,2029)	6,9262 (2,7949)
	0,7	4,8514 (1,0393)	7,0750 (1,8583)	8,9768 (2,6425)	10,8081 (3,2616)	4,8691 (1,0481)	4,9122 (1,0861)	5,0354 (1,2598)	6,9166 (2,6908)
	0,9	4,4365 (1,0435)	6,9762 (1,7775)	8,7979 (2,4153)	10,7472 (3,1170)	4,4467 (1,0505)	4,4940 (1,0958)	4,5354 (1,1195)	6,6715 (2,7256)
40	0	5,2606 (0,9238)	7,6373 (1,8313)	8,8750 (2,3097)	10,1224 (2,7457)	5,2700 (0,9134)	5,2835 (0,9003)	5,3247 (0,9299)	6,2736 (1,8819)
	0,3	5,1716 (0,8844)	7,5587 (1,8555)	8,7929 (2,2505)	10,3001 (2,8573)	5,1492 (0,9013)	5,1845 (0,9120)	5,3027 (0,9843)	6,6337 (2,3211)
	0,5	5,0501 (0,8830)	7,5256 (1,8041)	8,8298 (2,2055)	10,1353 (2,6369)	5,0671 (0,8976)	5,0351 (0,8831)	5,2126 (0,9962)	6,3458 (2,0475)
	0,7	4,8607 (0,8918)	7,5096 (1,6819)	8,8429 (2,1632)	10,1072 (2,6187)	4,8656 (0,8662)	4,9181 (0,8898)	5,0030 (0,9478)	6,2544 (2,0715)
	0,9	4,4011 (0,8819)	7,4742 (1,7736)	8,8074 (2,2507)	10,1928 (2,4259)	4,4180 (0,8953)	4,4324 (0,8859)	4,4279 (0,9009)	6,1529 (2,2669)
60	0	5,1747 (0,7101)	6,4699 (1,1173)	8,7740 (1,8719)	10,5646 (2,3335)	5,1550 (0,7240)	5,2118 (0,7279)	5,2699 (0,7660)	7,0885 (2,0448)
	0,3	5,1010 (0,7144)	6,3684 (1,0646)	8,8112 (1,8732)	10,5771 (2,4198)	5,1191 (0,7237)	5,1467 (0,6985)	5,2196 (0,7516)	7,1543 (2,0905)
	0,5	4,9734 (0,7259)	6,3596 (1,1442)	8,8017 (1,7633)	10,4087 (2,2742)	5,0072 (0,7611)	4,9939 (0,7159)	5,1085 (0,7757)	7,4003 (2,3482)
	0,7	4,7473 (0,7165)	6,2590 (1,0663)	8,7336 (1,8033)	10,4587 (2,2429)	4,7869 (0,7268)	4,8376 (0,7152)	4,8942 (0,7577)	7,1582 (2,1694)
	0,9	4,3459 (0,7249)	6,0522 (1,0404)	8,5622 (1,6604)	10,4154 (2,1839)	4,3467 (0,7003)	4,3591 (0,6957)	4,4284 (0,7581)	7,0964 (2,0635)
100	0	5,1397 (0,5359)	6,4768 (0,9055)	7,8391 (1,2200)	9,8531 (1,6898)	5,1270 (0,5543)	5,1556 (0,5459)	5,1232 (0,5312)	5,9350 (1,1127)
	0,3	5,0457 (0,5416)	6,3657 (0,8928)	7,8368 (1,2453)	9,7982 (1,6575)	5,0458 (0,5598)	5,0716 (0,5487)	5,0910 (0,5458)	5,9591 (1,1748)
	0,5	4,9303 (0,5383)	6,3354 (0,8731)	7,7583 (1,2220)	9,9131 (1,6765)	4,9569 (0,5510)	4,9567 (0,5478)	4,9975 (0,5548)	5,8981 (1,0848)
	0,7	4,6922 (0,5533)	6,1661 (0,8948)	7,6887 (1,2091)	9,7660 (1,5538)	4,7157 (0,5414)	4,7165 (0,5608)	4,7528 (0,5592)	5,7559 (1,1158)
	0,9	4,3168 (0,5485)	6,0950 (0,7998)	7,6304 (1,1142)	9,5829 (1,4682)	4,2937 (0,5407)	4,3184 (0,5253)	4,3108 (0,5537)	5,2641 (1,0949)

CAPÍTULO 5

Análise de dados reais

Esse capítulo tem como objetivo fazer uma análise detalhada de um conjunto de dados reais ilustrando a metodologia apresentada. O conjunto de dados Futebol introduzido no Capítulo 2 não será analisado, pois esse não mostrou conter intervalos *outliers* e também as variáveis respostas na representação centro e amplitude mostraram não serem correlacionadas.

5.1 Dados de Cardiologia

Medidas clínicas de 59 pacientes foram coletadas pelo Departamento de Nefrologia do hospital del Valle Naln, na cidade de Langreo, Espanha. Em cada paciente foram feitas três medições da pressão arterial sistólica, pressão arterial diastólica e taxa de pulso. Estes dados já foram analisados por Neto & de Carvalho (2008), Fagundes (2013), entre outros pesquisadores.

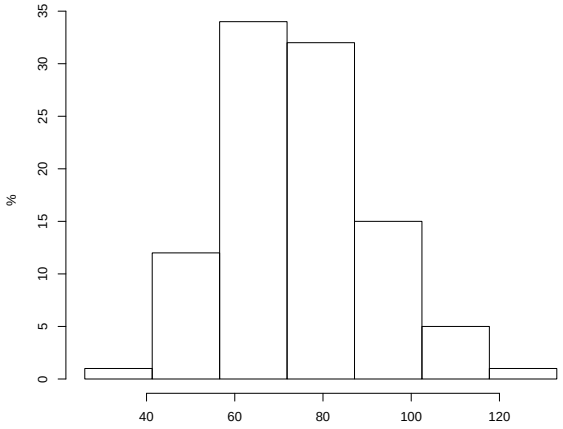
As variáveis taxa de pulso, pressão arterial sistólica e pressão arterial diastólica do conjunto cardiologia são classificadas como variáveis simbólicas intervalares, e a Tabela 5.1 descreve os resultados de duas medidas intervalares para essas variáveis. A média intervalar das variáveis taxa de pulso, pressão arterial sistólica e pressão arterial diastólica dos 59 pacientes foram respectivamente $74,5169$ batimentos/minuto, $146,7034$ mmHg e $83,4491$ mmHg. Sendo as variâncias intervalares nessas variáveis foram $274,8288$, $761,0363$ e $371,1994$ respectivamente. A Figura 5.1 mostra os histogramas intervalares de cada variável desse conjunto de dados, esses histogramas intervalares indicaram um comportamento simétrico em todas as variáveis.

A Figura 5.2 representa em três dimensões os intervalos das três variáveis do conjunto de dados Cardiologia. Nesta figura a um retângulo destoante dos demais, indicando um possível *outlier*.

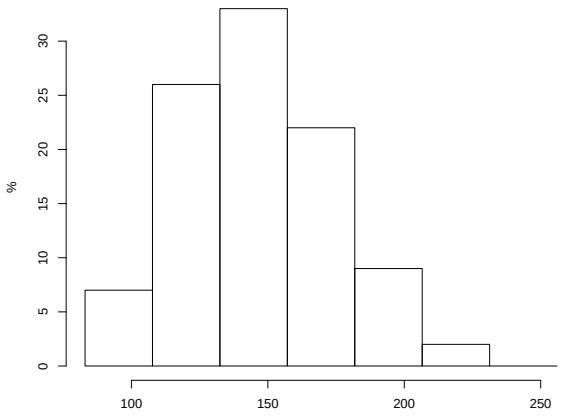
Tabela 5.1: Análise descritiva intervalar das variáveis do conjunto Cardiologia

Variáveis	Medidas	
	$\bar{x}$	s^2
Taxa de pulso	74,5169	274,8288
Pressão arterial sistólica	146,0363	761,0363
Pressão arterial diastólica	83,4491	371,1994

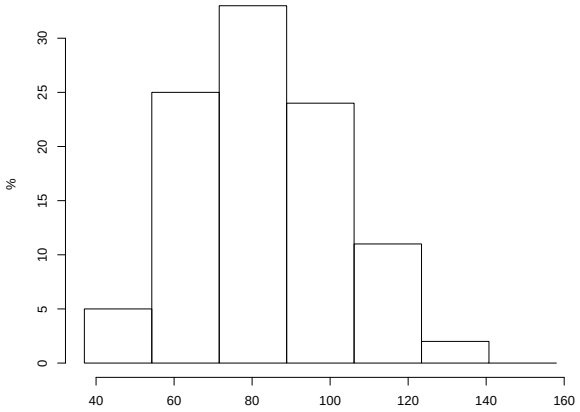
Figura 5.1: Histograma das variáveis intervalares do conjunto Cardiologia



(a) Taxa de Pulso

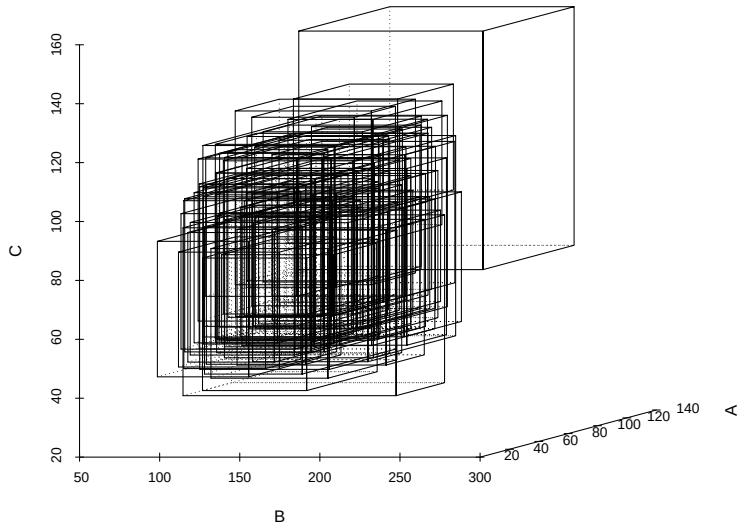


(b) Pressão Arterial Sistólica



(c) Pressão Arterial Diastólica

Figura 5.2: Gráfico 3D entre as variáveis Taxa de Pulso (A), Pressão Arterial Sistólica (B) e Pressão Arterial Diastólica (C)



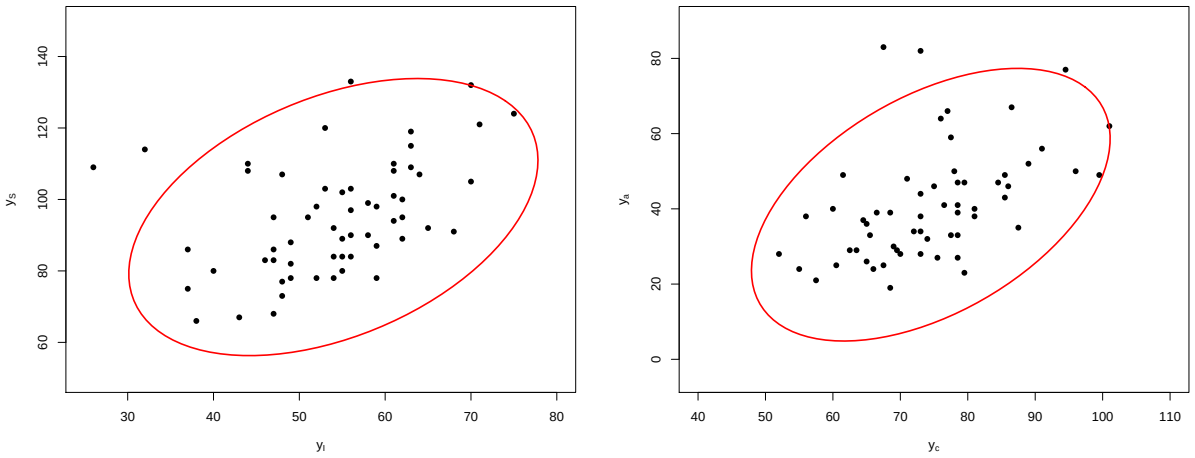
Um dos objetivos desse conjunto de dados foi verificar a influência das variáveis pressão arterial sistólica e pressão arterial diastólica sobre taxa de pulso. Verificamos também que o coeficiente de correlação linear de *pearson* entre o limite inferior e superior da variável taxa de pulso é de $0,41$. Na representação centro e amplitude temos $\tilde{\rho}^* = 0,485$. Esse comportamento também podem ser observado na Figura 5.3, que representam elipsóides com 95% de confiança obtidas a partir do vetor de médias e matriz escala da variável taxa de pulso. A Figura 5.3 também indicou três pontos fora da elipsóide de contorno, indicando possíveis *outliers*.

Nesse sentido, propomos ajustar um modelo de regressão elíptico bivariado intervalar com representação centro e amplitude para descrever a relação entre a variável resposta taxa de pulso e as variáveis explicativas: pressão diastólica e pressão sistólica. O MREBI é definido por:

$$\begin{pmatrix} tp_{ci} \\ tp_{ai} \end{pmatrix} = \begin{pmatrix} \mu_{ci} \\ \mu_{ai} \end{pmatrix} + \Sigma^{1/2} \begin{pmatrix} \epsilon_{ci} \\ \epsilon_{ai} \end{pmatrix} \quad i = 1, \ldots, 59 \tag{5.1}$$

em que tp_{ci} é a taxa de pulso centro do i-ésimo intervalo, tp_{ai} é taxa de pulso amplitude do i-ésimo intervalo, $\mu_{ci} = \beta_{c1} + \beta_{c2}x_{c2i} + \beta_{c3}x_{c3i}$, x_{c2} representa a variável pressão arterial sistólica centro e x_{c3} pressão arterial diastólica centro, e $\mu_{ai} = \exp(\beta_{a1} + \beta_{a2}x_{a2i} + \beta_{a3}x_{a3i})$, x_{a2} representa a variável pressão arterial sistólica amplitude e x_{a3} pressão arterial diastólica amplitude.

Figura 5.3: Elipsóide com 95% de confiança da variável taxa de pulso



(a) Limite inferior versus limite superior

(b) Centro versus amplitude

Inicialmente, foi considerado uma distribuição normal bivariada para o modelo (5.1). Esse modelo resultou um AIC de $922,2273$, as estimativas e desvio padrão dos parâmetros desse modelo são vistos na Tabela 5.2, em que $\hat{\rho}^* = \frac{\hat{\phi}^*_{12}}{\sqrt{\hat{\phi}^*_{11}\hat{\phi}^*_{22}}}$.

Tabela 5.2: Estimativas e desvio padrão do MREBI considerando distribuição Normal para os dados de Cardiologia

Parâmetros	Estimativas	Erro padrão
β_{c1}	$70,5223$	$0,4063$
β_{c2}	$5,8150$	$1,3614$
β_{c3}	$4,2324$	$1,0633$
β_{a1}	$3,6214$	$0,0093$
β_{a2}	$-0,2671$	$0,0392$
β_{a3}	$0,4393$	$0,0284$
ϕ^*_{11}	$116,1814$	$2,7668$
ϕ^*_{22}	$207,0231$	$4,9563$
ϕ^*_{12}	$79,0590$	$2,5099$
ρ^*	$0,5099$	$0,0151$

O MREBI sob distribuição normal obteve um MADI intervalar de previsão de 21.2645 com desvio padrão de 12.7548.

O resíduo ordinário de um modelo mensura a diferença entre o valor observado e o ajustado e é dado por

$$r_i = y_i - \hat{y}_i \quad i = 1, \ldots, 59$$

em que r_i representa o resíduo da i-ésima observação.

Para resíduos intervalares temos as seguintes expressões

$$r_{Ii} = y_{Ii} - \hat{y}_{Ii}$$

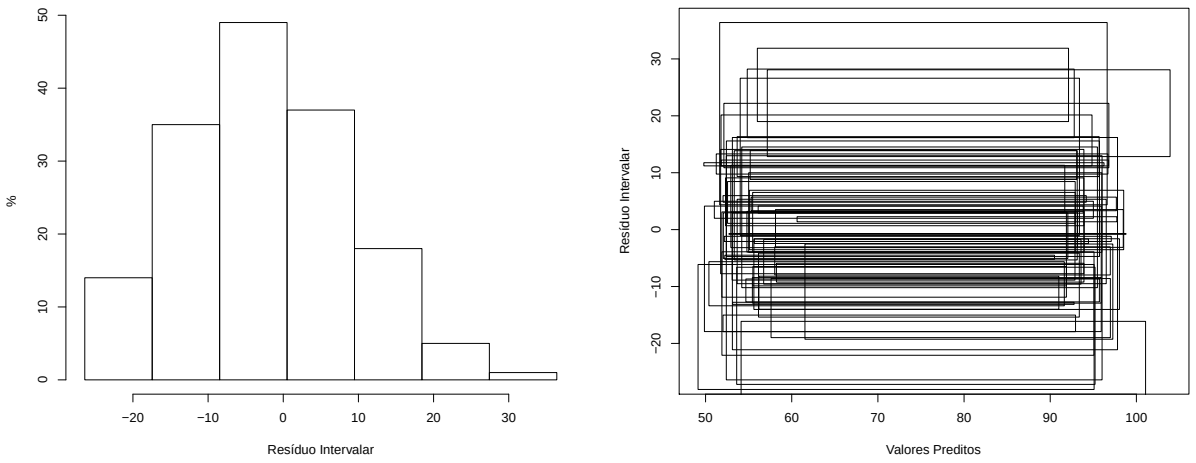
e

$$r_{Si} = y_{Si} - \hat{y}_{Si}$$

em que r_{Ii} representa o resíduo da i-ésima observação do limite inferior e r_{Si} representa o resíduo da i-ésima observação do limite superior.

Pelo histograma intervalar dos resíduos (Figura 5.5(a)) indicou que os resíduos tem média próximo de zero e com comportamento aproximadamente simétrico. Em relação ao gráfico dos retângulos (Figura 5.5(b)) indicaram não haver correlação dos retângulos e um comportamento homogêneo.

Figura 5.4: Resíduos intervalares do MREBI sob distribuição normal



(a) Histograma

(b) Gráfico dos retângulos

Como os dados apresentaram indícios de um possível retângulo *outlier*, um modelo MREBI sob distribuição t-Student é proposto com o objetivo de obter um modelo menos sensível a presença de retângulos *outliers*. Os graus de liberdade ν considerado no MREBI com distribuição t-Student foi escolhido baseado no menor valor do critério de *Akaike*(AIC) dentre uma grade de valores. Pela Tabela 5.3, o MREBI t-Student com 6 g.l apresentou o menor valor de AIC.

Na Tabela 5.4 é apresentado as estimativas de máxima verossimilhança dos parâmetros $\hat{\beta}^*$'s, $\hat{\phi}^*$'s e erros padrões assintóticos desse modelo. Ao analisar esses valores, temos que as estimativas dos β^*'s foram bem próximas aos do modelo normal. Contudo, as estimativas dos ϕ^*'s tiveram valores menores.

O MADI de previsão do MREBI sob distribuição t-Student(6) foi de $21,2366$ com desvio padrão de $12,8403$.

Tabela 5.3: Valores do AIC supondo distribuição t-Student nos modelos para os dados de Cardiologia

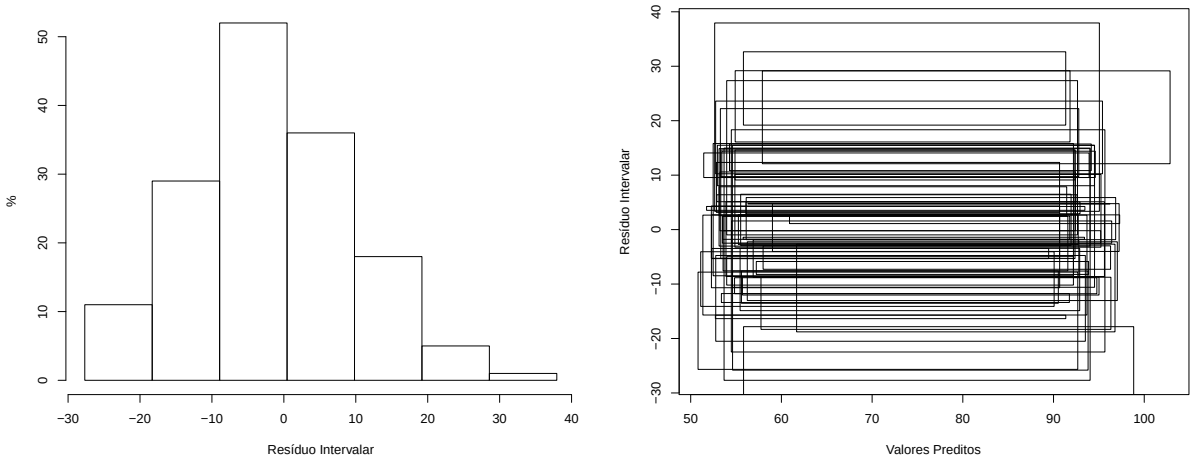
Modelo	AIC
t-Student (3)	$922,0649$
t-Student (4)	$920,3099$
t-Student (5)	$919,5853$
t-Student (6)	**$919,2936$**
t-Student (10)	$919,4087$

Tabela 5.4: Estimativas e desvio padrão do MREBI considerando distribuição t-Student(6) para os dados de Cardiologia

Parâmetros	Estimativas	Erro padrão
β_{c1}	$70,0278$	$0,4137$
β_{c2}	$5,1339$	$1,5944$
β_{c3}	$5,2851$	$1,2135$
β_{a1}	$3,5768$	$0,0099$
β_{a2}	$-0,1050$	$0,0348$
β_{a3}	$0,3125$	$0,0255$
ϕ^*_{11}	$95,6972$	$5,0123$
ϕ^*_{22}	$140,5114$	$10,0813$
ϕ^*_{12}	$67,0476$	$5,5149$
ρ^*	$0,5778$	$0,0188$

Analisando o comportamento dos resíduos intervalares desse modelo nas Figuras 5.6(a) e 5.6(b), observamos comportamento semelhante ao do modelo normal. Visto que, o gráfico dos retângulos mostram não haver indícios de resíduos correlacionados e também um comportamento homogêneo.

Figura 5.5: Resíduos intervalares do MREBI sob distribuição t-Student(6)



(a) Histograma

(b) Gráfico dos retângulos

CAPÍTULO 6

Considerações finais

No desenvolvimento desta dissertação, foi feito uma revisão sobre dados simbólicos, principais metodologias abordadas para analisar dados dessa natureza e análise descritiva de dados simbólicos intervalares. Introduzimos o modelo de regressão elíptico bivariado intervalar que considera a dependência entre o limite inferior e superior do intervalo, e que assegura a coerência matemática ao considerar a representação centro e amplitude com uma função de ligação *log* no componente sistemático da amplitude.

Foram propostos estudos de simulações de Monte Carlo sob distribuição normal e t-Student(4) para o componente aleatório. Consideramos diferentes cenários e adicionamos diversos percentuais de *outliers* no conjunto de ajustamento para verificar o comportamento do erro de previsão médio e desvio padrão MADI. Nessas simulações, vimos que, em geral, o erro de previsão médio e desvio padrão MADI possuem comportamento semelhante nos dois modelos, ambos diminuem com o aumento do tamanho amostral quando a correlação e percentual de *outliers* foram fixados. Ao fixar um tamanho de amostra e percentual de *outliers*, os erros médios e desvio padrão sofreram variações com o aumento da correlação em todos os cenários. Contudo, em algumas ocasiões os erros médios não sofreram variações como nos cenários 1, 2 e 3. Já o desvio padrão não apresentou essas variações nos cenários 4, 5, 6 e 7. Ao analisar o aumento do percentual de *outliers* com correlação e tamanho fixados, observou que o erro médio e desvio padrão MADI do MREBI sob distribuição normal é mais sensível que o MREBI sob t-Student(4).

Além das simulações, aplicação a um conjunto de dados reais foi feita para o MREBI sob distribuição normal e t-Student. Ao ajustar os modelos, pudemos observar os valores estimados e erro padrão dos parâmetros, os erros médios e desvio padrão de previsão MADI e análise de resíduo intervalar em cada um dos modelos. Ao comparar o erro médio de predição do MREBI sob distribuição normal e t-Student(6), observamos que os erros foram bem próximos.

Referências

Alcantara, I. C. & F. J. A. Cysneiros (2013). Linear regression models with slash-elliptical errors. *Computational Statistics & Data Analysis 64*, 153–164.

Anderson, T. & K.-T. Fang (1990). Inference in multivariate elliptically contoured distributions based on maximum likelihood. *Statistical inference in elliptically contoured and related distributions*, 201–216.

Anderson, T., H. Hsu, & K.-T. Fang (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. *Canadian Journal of Statistics 14*(1), 55–59.

Arellano-Valle, R. (1994). *Distribuições elıpticas: Propriedades, inferência e aplicações a modelos de regressão*. Ph. D. thesis, Tese de Doutorado, Programa de pós-graduação em Estatıstica, Universidade de Sao Paulo.

Bertrand, P. & F. Goupil (2000). Descriptive statistics for symbolic data. in: Analysis of symbolic data: Explanatory methods for extracting statistics information from complex data. *Springer-Verlag*, 103–124.

Billard, L. & E. Diday (2000). Regression analysis for interval-valued data. In *Data Analysis, Classification, and Related Methods*, pp. 369–374. Springer.

Billard, L. & E. Diday (2002). Symbolic regression analysis. In *Classification, Clustering, and Data Analysis*, pp. 281–288. Springer.

Billard, L. & E. Diday (2003). From the statistics of data to the statistics of knowledge: symbolic data analysis. *Journal of the American Statistical Association 98*(462), 470–487.

Billard, L. & E. Diday (2006a). Symbolic data analysis: Conceptual statistics and data mining.

Billard, L. & E. Diday (2006b). Symbolic data analysis: Conceptual statistics and data mining.

Bock, H. H. & E. Diday (Heidelberg, 2000). *Analysis of Symbolic Data. Studies in Classification*. Data Analysis and Knowledge Organization.

Cambanis, S., S. Huang, & G. Simons (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis 11*(3), 368–385.

Campos, V. M. F. (2008). Análise simbólica de dados e a sua aplicação na extração de informação de estatísticas oficiais: análise do inquérito à ocupação do tempo.

Carvalho, F. (1995). Histograms in symbolic data analysis. *Annals of Operations Research 55*(2), 299–322.

Carvalho, F., J. T. Pimentel, L. X. Bezerra, & R. M. de Souza (2007). Clustering symbolic interval data based on a single adaptive hausdorff distance. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pp. 451–455. IEEE.

Carvalho, F. d. A. (1994). Proximity coefficients between boolean symbolic objects. In *New approaches in classification and data analysis*, pp. 387–394. Springer.

Carvalho, F. d. A. d. & R. M. C. R. de Souza (2010). Unsupervised pattern recognition models for mixed feature-type symbolic data. *Pattern Recognition Letters 31*(5), 430–443.

Chouakria, A., E. Diday, & P. Cazes (1998). An improved factorial representation of symbolic objects. *Knowledge Extraction from Statistical Data*, 301–305.

Díaz García, J., M. Galea, & V. Leiva (2003). Influence diagnostics for elliptical multivariate linear regression models. *Communication in Statistics–Theory and Methods. 32*, 625–641.

Diday, E. (1988). The symbolic approach in clustering and related methods of data analysis, classification and related methods of data analysis. In *Proceedings of the first Conference of the Federation of the classification societies. North Holland.*

Diday, E. (1989). Introduction à l'analyse des données symboliques.

Diday, E. & M. P. Brito (1989). Symbolic cluster analysis. In *Conceptual and Numerical Analysis of Data*, pp. 45–84. Springer.

Diday, E. & M. Noirhomme-Fraiture (2008). *Symbolic data analysis and the SODAS software*. Wiley Online Library.

Domingues, M. A., R. M. de Souza, & F. J. A. Cysneiros (2010). A robust method for linear regression of symbolic interval data. *Pattern Recognition Letters 31*(13), 1991–1996.

Fagundes, R. A., R. M. De Souza, & F. J. A. Cysneiros (2013). Robust regression with application to symbolic interval data. *Engineering Applications of Artificial Intelligence 26*(1), 564–573.

Fagundes, R. A., R. M. De Souza, & F. J. A. Cysneiros (2014). Interval kernel regression. *Neurocomputing 128*, 371–388.

Fagundes, R. A. d. A. (2013). *Métodos de regressão robusta e kernel para dados intervalares*. Ph. D. thesis, Tese de Doutorado, Programa de pós-graduaçao em Ciência da Computação, Universidade Federal de Pernambuco.

Fang, K.-T., S. Kotz, & K. W. Ng (1990). *Symmetric multivariate and related distributions*. Chapman and Hall.

Ferreira, D. F. (2008). *Estatística multivariada*. Editora UFLA.

Galea, M., M. Riquelme, & G. A. Paula (2000). Diagnostics methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics 14*, 167–184.

Gordon, A. D. (2000). An iterative relocation algorithm for classifying symbolic data. In *Data Analysis*, pp. 17–23. Springer.

Ibacache-Pulgar, G., G. A. Paula, & M. Galea (2012). Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling 12*(2), 165–193.

Kelker, D. (1970). Distribution theory of spherical distributions and a location–scale parameter generalization. *Sankhyã: The Indian Journal of Statistics, Series A 32*, 419–430.

Lange, K. L., R. J. A. Litte, & J. M. G. Taylor (1989). Robust statistical modeling using the t distribution. *Jounal of the American Statistical Association 84*, 881–896.

Liu, S. (2000). On local influence for elliptical linear models. *Statistical Papers 41*, 211–224.

Liu, S. (2002). Local influence in multivariate elliptical linear regression models. *Linear Algebra Application. 354*, 159–174.

Maia, A. L. S., F. d. A. de Carvalho, & T. B. Ludermir (2008). Forecasting models for interval-valued time series. *Neurocomputing 71*(16), 3344–3352.

Momeni, M., M. D. Nayeri, A. F. Ghayoumi, & H. Ghorbani (2010). Robust regression and its application in financial data analysis. *World Academy of Science, Engineering and Technology 47*, 521–526.

Neto, E. d. A. L., G. M. Cordeiro, & F. d. A. de Carvalho (2011). Bivariate symbolic regression models for interval-valued variables. *Journal of Statistical Computation and Simulation 81*(11), 1727–1744.

Neto, E. d. A. L. & F. d. A. de Carvalho (2008). Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis 52*(3), 1500–1515.

Neto, E. d. A. L. & F. d. A. de Carvalho (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics & Data Analysis 54*(2), 333–347.

Salazar, D. R. d. S. (2008). Classificadores para dados simbólicos do tipo intervalo baseados em modelos geométricos.

Savalli, C., G. A. Paula, & F. J. Cysneiros (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling 6*(1), 59–76.

Shanno, D. F. (1985). On broyden-fletcher-goldfarb-shanno method. *Journal of Optimization Theory and Applications 46*(1), 87–94.

Souza, R. M. d., D. C. Queiroz, & F. J. A. Cysneiros (2011). Logistic regression-based pattern classifiers for symbolic interval data. *Pattern Analysis and Applications 14*(3), 273–282.

Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association 21*(153), 65–66.