

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CAMPUS DE FOZ DO IGUAÇU
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA E COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

**MODELO MATEMÁTICO MULTIVARIADO PARA ANÁLISE DE
RISCO EM SISTEMAS DINÂMICOS**

ALINE CHAVES

FOZ DO IGUAÇU
2018

Aline Chaves

**Modelo Matemático Multivariado para Análise de Risco em Sistemas
Dinâmicos**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Sistemas Dinâmicos e Energéticos.

Orientador: Prof. Dr. Carlos Roberto Mendonça da Rocha

Coorientador: Prof. Dr. Orlando Catarino da Silva

Foz do Iguaçu
2018

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Chaves, Aline

Modelo Matemático Multivariado para Análise de Risco em Sistemas Dinâmicos : Modelo Matemático Multivariado para Análise de Risco em Sistemas Dinâmicos / Aline Chaves; orientador(a), Dr. Carlos Roberto Mendonça da Rocha; coorientador(a), Orlando Catarino da Silva, 2018. 108 f.

Dissertação (mestrado), Universidade Estadual do Oeste do Paraná, Centro de Engenharias e Ciências Exatas, Programa de Pós-Graduação em Engenharia Elétrica e Computação, 2018.

1. Sistemas Dinâmicos. 2. Análise Multivariada. 3. Análise Fatorial. 4. Método Estatístico Multivariado. I. Rocha, Dr. Carlos Roberto Mendonça da . II. Silva, Orlando Catarino da. III. Título.

**MODELO MATEMÁTICO MULTIVARIADO PARA ANÁLISE DE RISCO EM
SISTEMAS DINÂMICOS**

Aline Chaves

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Engenharia
Elétrica e Computação e aprovada pela Banca Examinadora:

Data da defesa 10/09/2018



Prof. Dr. Carlos Roberto Mendonça da Rocha – (Orientador)
Universidade Estadual do Oeste do Paraná - UNIOESTE



Prof. Dr. Orlando Catarino da Silva – (Coorientador)
Universidade Estadual do Oeste do Paraná - UNIOESTE



Prof. Dr^a. Sheila Regina Oro
Universidade Tecnológica Federal do Paraná - UTFPR



Prof. Dr. Ricardo Luiz Barros de Freitas
Universidade Estadual do Oeste do Paraná - UNIOESTE

Resumo

A área de Sistemas Dinâmicos estuda processos em que a evolução é estabelecida por uma lei matemática, cujo objetivo é desenvolver uma teoria matemática destes processos que permita compreender e prever sua evolução. Um sistema estrutural complexo, geralmente possui um vasto banco de dados contendo as variáveis aleatórias relacionadas com o sistema, que podem fornecer informações inestimáveis sobre as condições de segurança da estrutura. Neste trabalho foi desenvolvido, com o auxílio das técnicas da Análise Multivariada e da Confiabilidade Estrutural, um modelo matemático que possui a capacidade de auxiliar na atividade de monitoramento e na prevenção de anomalias em bloco de concreto altamente instrumentado, contribuindo para a realização de ações antes da ocorrência de problemas. O modelo foi estabelecido com nove componentes, com cada um representando uma anomalia, e foi implementado um sistema de alerta para a representação da situação do Sistema Dinâmico. Foi realizado um estudo de caso considerando um bloco de concreto altamente instrumentado da Usina Hidrelétrica de Itaipu. A aplicação do modelo apresentou resultados satisfatórios.

Palavras-chave: Sistema Dinâmico, Análise Multivariada, Análise Fatorial, Método Estatístico Multivariado, Modos de Falha..

Abstract

The area of Dynamic Systems studies processes in which the evolution is established by a mathematical law, whose objective is to develop a mathematical theory of these processes that allows to understand and to predict its evolution. A complex structural system usually has a vast database containing random variables related to the system, which can provide invaluable information on the safety conditions of the structure. In this work, a mathematical model was developed, with the aid of the Multivariate Analysis and Structural Reliability techniques, which has the capacity to assist in the monitoring activity and in the prevention of anomalies in a highly instrumented concrete block, contributing to the accomplishment of actions before the occurrence of problems. The model was established with nine components, with each one representing an anomaly, and an alert system was implemented to represent the situation of the Dynamic System. A case study was carried out considering a highly instrumented concrete block of the Itaipu Hydroelectric Power Plant. The application of the model presented satisfactory results.

Keywords: Dynamic System, Multivariate Analysis, Factor Analysis, Multivariate Statistical Method, Failure Modes.

Dedico esse trabalho aos meus pais e ao meu irmão.

Agradecimentos

Agradeço primeiramente à Deus, por me dar forças para chegar até o fim de mais uma importante jornada em minha vida acadêmica, só ele sabe das dificuldades e obstáculos que enfrentei durante esse período, mas com sua ajuda consegui superar todas as adversidades.

Agradeço imensamente aos meus pais e ao meu irmão que, mesmo de longe, sempre me incentivaram e sempre me apoiaram em todos os momentos do curso, me dando conselhos, me acalmando nas horas de "desespero e dificuldades", mas que nunca duvidaram da minha capacidade de chegar até o fim dessa etapa, obrigada por serem meu alicerce. Agradeço também aos demais familiares, que sempre me apoiaram e sei que torcem muito pelo meu sucesso.

Agradeço à todos meus professores da Graduação pelo conhecimento transmitido, mas um agradecimento especial, ao professor Emídio Santos Portilho Júnior e à professora Kelly Roberta Mazzutti Lübeck, por acreditarem no meu potencial e na minha dedicação e, aceitaram meu pedido de encaminhar a carta de recomendação ao programa de Pós-Graduação em Engenharia Elétrica e Computação, no qual o mesmo foi aceito e, estou chegando ao fim de mais uma conquista.

Agradeço ao professor Carlos Henrique Farias dos Santos por ter aceito o pedido para ser meu orientador, porém por falta de conhecimento e incompetência da minha parte, não consegui desenvolver a pesquisa que me foi proposta, mas, mesmo assim, teve uma importante contribuição para a conclusão desse curso.

Agradeço ao professor Carlos Roberto Mendonça da Rocha por ter aceito o desafio de se tornar meu novo orientador, o qual sempre me ajudou com seu conhecimento, fazendo-o com muita dedicação, paciência e entusiasmo, sua contribuição foi de extrema importância para a conclusão desse trabalho. Agradeço também ao professor Orlando Catarino da Silva por aceitar o convite de ser o coorientador dessa pesquisa, sempre me auxiliando, apoiando e aconselhando, sua ajuda foi fundamental para a realização de mais esse sonho.

Agradeço aos demais professores do Programa de Pós-Graduação em Engenharia Elétrica e Computação (PGEEC) por todo conhecimento compartilhado.

Agradeço de coração ao professor Itamar Pena Nieradka e aos colegas de projeto, Christian Carvalho e Lucas Reinaldet, pelas inúmeras reuniões e encontros, que sem medir esforços, sempre estavam dispostos à ajudar, e que tiveram uma imensa importância no desenvolvimento do trabalho.

Agradeço aos profissionais da Usina Hidrelétrica (UHE) de Itaipu, em especial ao Dimilson Pinto Coelho e à Silvia Frazão, que sempre estavam dispostos à ajudar, fornecendo todas as informações necessárias para a realização dessa pesquisa.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro e ao Centro de Estudos Avançados em Segurança de Barragens (CEASB) pelo espaço disponível para realizar o estudo, agradeço imensamente aos profissionais que ali trabalham por toda dedicação e comprometimento que possuem com relação aos estudantes e pesquisadores que ali frequentam e, também agradeço pelo apoio financeiro durante o curso.

Por fim, mas não menos importante, agradeço à todos meus amigos, o qual a maioria estão à quilômetros de distância mas, mesmo assim, sempre me apoiaram, incentivaram, torceram e vibraram com cada conquista em minha vida.

Agradeço de coração à cada um de vocês!!!

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xv
Lista de Siglas	xvii
Capítulo 1 Introdução	1
1.1 Objetivos	2
1.1.1 Objetivo Geral	2
1.1.2 Objetivo específico	2
1.2 Motivação	3
1.3 Justificativa	3
1.4 Estrutura do Trabalho	3
Capítulo 2 Variável Aleatória, Análise Fatorial e Confiabilidade Estrutural	5
2.1 Variáveis Aleatórias Unidimensionais	5
2.1.1 Variáveis Aleatórias Unidimensionais Discretas	6
2.1.2 Variáveis Aleatórias Unidimensionais Contínuas	6
2.2 Estatística Univariada	7
2.2.1 Teste de Normalidade Lilliefors	9
2.3 Análise Multivariada	10
2.3.1 Estatística Multivariada	10
2.3.2 Teorema Espectral	13
2.3.3 Componentes Principais	15
2.4 Análise Fatorial	22
2.4.1 Suposições da Análise Fatorial	23
2.4.2 O Modelo Fatorial Ortogonal	25
2.5 Confiabilidade Estrutural	32
2.6 Estudo da Arte	33
Capítulo 3 Modelo Matemático Multivariado	37

3.1	Caracterização do Modelo Matemático para o Caso Geral	38
3.2	Modelo Matemático para o Caso Geral.....	42
Capítulo 4 Testes e Resultados		45
4.1	Aplicação do Método Estatístico Multivariado ao Bloco-Chave A-7 da Barragem de Itaipu	46
4.1.1	Testes Estatísticos	46
4.2	Determinação dos Modos de Falha	49
4.2.1	Escorregamento na Brecha D	50
4.2.2	Subpressão	51
4.2.3	Escorregamento na Junta D	52
4.2.4	Fluência	53
4.2.5	Tombamento	54
4.2.6	Sinais de Alerta	55
4.3	Semi-Componentes da Função de Desempenho	56
4.3.1	Recalque ou Distensão Pé Montante na Junta D	56
4.3.2	Recalque ou Distensão Pé Montante na Brecha D	57
4.4	Função de Desempenho Multivariado para o Caso Aplicado	58
4.5	Validação do Modelo Matemático	60
Capítulo 5 Conclusão		67
Referências Bibliográficas		69
Apêndice A		73

Lista de Figuras

3.1: Fluxograma da Metodologia	38
3.2: Fluxograma do Algoritmo para Determinar os Modos de Falha.....	42
4.1: Bloco-Chave A-7.....	46
4.2: Região Crítica do Escorregamento na Brecha D.....	51
4.3: Região Crítica da Subpressão na Brecha D.....	52
4.4: Região Crítica do Escorregamento na Junta D.....	53
4.5: Região Crítica da Fluência	54
4.6: Região Crítica do Tombamento.....	55
4.7: Região Crítica do Recalque ou Distensão Pé Montante na Junta D.....	57
4.8: Região Crítica do Recalque ou Distensão Pé Montante na Brecha D.....	58

Lista de Tabelas

2.1: Valores de Interpretação do Critério de Kaiser-Meyer-Olkin – KMO.....	24
4.1: Grupos de Instrumentos Altamente Correlacionados.....	485
4.2: Grupos de Instrumentos Altamente Correlacionados.....	48
4.3: Resultados Obtidos após o Teste Lilliefors	49
4.4: Associação dos Fatores com as Anomalias	59
4.5: Resultados da Validação com Dados Sintéticos.....	61
4.6: Resultados da Validação com Dados Reais.....	63
4.7: Resultados da Validação com Dados Reais Modificados	64

Lista de Siglas

PGEEC	Programa de Pós-Graduação em Engenharia Elétrica e Computação
UHE	Usina Hidrelétrica
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CEASB	Centro de Estudos Avançados em Segurança de Barragens
LF	Teste de Lilliefors
KS	Teste de Kolmogorov-Smirnov
TDE	Teorema da Decomposição Espectral
AF	Análise Fatorial
KMO	Critério de Kaiser-Meyer-Olkin
FORM	Métodos de Confiabilidade de Primeira Ordem
SORM	Método de Confiabilidade de Segunda Ordem
MC	Mudança de Comportamento
BC	Bom Comportamento
RC	Região Crítica
SOM	Sistema de Operação e Manutenção de Itaipu

Capítulo 1

Introdução

A construção da Usina Hidrelétrica (UHE) de Itaipu foi de grande importância para o desenvolvimento do país e demonstrou grande capacidade da engenharia brasileira. Uma obra muito avançada para época que envolveu em sua construção, profissionais de alta capacidade. No entanto, uma obra tão grande requer muita vigilância. No mundo todo, a segurança de barragens é um assunto muito discutido, e o Brasil mostra-se preocupado com a "saúde" de suas barragens (SILVA, 2016).

A teoria dos sistemas dinâmicos teve sua origem no estudo de modelos matemáticos que descrevem a evolução de um sistema físico. Atualmente seu escopo é muito mais amplo, uma vez que conexões com outras áreas da Matemática foram surgindo ao longo de sua história. Tanto seus métodos quanto resultados se relacionam de maneira profunda com áreas como Topologia, Probabilidade, Ações de Grupos, Geometria Hiperbólica e Análise Numérica (MATEMÁTICA, 2017).

Contribuíram para o desenvolvimento da teoria dos sistemas dinâmicos, os trabalhos desenvolvidos por Johannes Kepler, sobre a mecânica dos corpos celestes, e Isaac Newton, com seus trabalhos sobre modelagem matemática através da formalização da mecânica clássica. No entanto, o francês Henri Poincaré é considerado um dos criadores da teoria moderna dos sistemas dinâmicos, introduzindo muitos dos aspectos do estudo qualitativo das equações diferenciais que permitiram estudar propriedades assintóticas das soluções (ou da maior parte das soluções) de uma equação diferencial, como estabilidade e periodicidade, sem ser necessário resolver explicitamente a equação diferencial. Tal abordagem pode ser encontrada na sua obra-prima *Les Méthodes Nouvelles de La Mécanique Céleste*, publicada em três volumes entre 1892 e 1899.

O objetivo do estudo teórico de sistemas dinâmicos é prever o futuro (ou explicar o passado) de modo científico. Para fazer isso, é necessário conhecer como os processos funcionam e compreender as regras que governam as mudanças que ocorrerão.

Segundo Monteiro (2006), um sistema pode ser definido como um conjunto de objetos agrupados por alguma interação ou interdependência, de modo que existam relações de causa e efeito nos fenômenos que ocorrem com os elementos desse conjunto. Neste contexto, nesse trabalho pretendeu-se mostrar que um bloco de concreto de uma barragem altamente

instrumentada pode ser tratado como um sistema dinâmico, revelando através de métodos matemáticos e estatísticos, as relações entre os instrumentos instalados nesse bloco. Tais instrumentos têm a função de verificar comportamentos na estrutura que podem ser admissíveis ou não.

Um sistema estrutural complexo geralmente possui um vasto banco de dados contendo as variáveis aleatórias relacionadas com o sistema, que podem fornecer inúmeras informações sobre as condições de segurança da estrutura. Buscou-se, com o auxílio das técnicas da Análise Multivariada e da Confiabilidade Estrutural, desenvolver um modelo matemático capaz de prever anomalias na estrutura, permitindo uma ação prévia do problema antes que ele ocorra. Para isso, esses modelos deveriam apresentar um sistema de alertas que fossem condizentes com a atual situação do Sistema Dinâmico.

Nesse trabalho foi desenvolvido um modelo com a aplicação de um método estatístico multivariado, para definir uma função de desempenho multivariada capaz de diagnosticar as condições de segurança do bloco. Para tanto, utiliza-se as variáveis aleatórias que são geradas pela instrumentação instalada no bloco, os conceitos teóricos apresentados na literatura sobre a instrumentação e suas funções, além do conhecimento prático dos profissionais em segurança de barragens da UHE de Itaipu.

1.1 Objetivos

A seguir são apresentados o objetivo geral e os objetivos específicos que foram perseguidos com a realização desta pesquisa.

1.1.1 Objetivo Geral

O objetivo geral deste trabalho foi desenvolver um modelo matemático multivariado, que utilizasse as variáveis aleatórias que estão associadas à um sistema dinâmico, de modo que permitisse avaliar as condições de segurança da estrutura em termos probabilísticos, além de permitir diagnosticar com antecedência se uma mudança brusca de comportamento estaria ocorrendo.

1.1.2 Objetivo específico

Para cumprir o objetivo geral, foram considerados os seguintes objetivos específicos:

- Analisar as variáveis aleatórias relacionadas à um Sistema Dinâmico;
- Definir a importância de cada variável no desempenho do sistema;
- Aplicar a análise fatorial na matriz multivariada gerada pelas variáveis aleatórias que possuíssem relevância acerca do sistema;
- Avaliar a associação dos fatores com as possíveis anomalias que esse sistema estrutural poderia sofrer;
- Aplicar os conceitos da Confiabilidade Estrutural aos resultados encontrados com a utilização da Análise Fatorial;

- Construir um modelo matemático multivariado como função de desempenho que fosse capaz de monitorar o Sistema Dinâmico proposto;
- Aplicar os conceitos da confiabilidade estrutural aos dados, considerando o modelo matemático, para calcular estimativas da probabilidade de cada evento de falha.

1.2 Motivação

Quando se trabalha com sistemas dinâmicos, é preciso tomar decisões e fazer inferências com base em variáveis aleatórias que estão presentes no sistema. Situações mais elaboradas, oriundas de aplicações, dão origem a problemas mais complexos, onde se necessita de mais de uma variável para tornar possível a modelagem e a busca por soluções. Para esses casos, o conhecimento de técnicas da Análise Multivariada e da Confiabilidade Estrutural pode, muitas vezes, dar o suporte necessário para obtenção do conhecimento com base num conjunto de dados armazenados.

1.3 Justificativa

No mundo todo, a segurança de barragens é um assunto muito discutido, e o Brasil mostra-se preocupado com a “saúde” de suas barragens. Essa segurança representa um motivo de preocupação para a sociedade, já que existem riscos potenciais a que ficam sujeitas às pessoas e bens instalados nas cidades.

As barragens são obras que necessitam de uma certa atenção, a fim de evitar algum tipo de problema na estrutura ou até mesmo ao meio ambiente ou perdas de vidas humanas. O reconhecimento de problemas e uma possível tomada de decisão para reparos, pode por sua vez, diminuir o fator risco. Porém, esses estudos requerem muita atenção para a determinação de soluções adequadas.

1.4 Estrutura do Trabalho

Neste capítulo foi apresentada uma breve introdução ao assunto da pesquisa, com a descrição dos objetivos (geral e específicos), a motivação e a justificativa para a sua realização.

No capítulo 2, são apresentados os conceitos teóricos que foram necessários para a realização do trabalho. Apresenta-se uma breve revisão sobre variáveis aleatórias discretas e contínuas, aprofundando nos resultados referentes à análise multivariada, com os principais resultados da estatística multivariada, análise de componentes principais e Análise Fatorial, que foi a principal ferramenta utilizada para o desenvolvimento do trabalho. É mencionado também o Teorema Espectral, que é uma ferramenta da álgebra linear, necessária para comprovação de

muitos resultados demonstrados nesse capítulo.

No capítulo 3, é apresentado o Modelo Matemático Multivariado desenvolvido por Silva (2016). Este modelo visa a determinação de uma função de desempenho multivariada, que seja capaz de monitorar o comportamento estrutural do bloco-chave em estudo, fazendo com que ele possa ser tratado como um sistema dinâmico. Neste capítulo também, alguns conceitos e algoritmos que são necessários para a aplicação do modelo são definidos e comprovados, tais como, intervalos de mudança de comportamento e determinação dos modos de falha.

No capítulo 4 é especificado os Testes e Resultados obtidos com a realização desse trabalho, no qual é apresentado o bloco-chave A-7 da UHE de Itaipu que pretende-se tratar como um sistema dinâmico, também são mostrados as regiões críticas de cada anomalia encontradas nesse bloco e é apresentada a função de desempenho multivariada para o caso aplicado e, por fim, é realizado a validação do modelo matemático, incluindo os resultados de simulações com dados simulados para calibração do código, dados reais e reais modificados.

No Capítulo 5, é apresentada a Conclusão sobre a realização desse estudo, bem como, algumas sugestões de trabalhos futuros e as bibliografias utilizadas para o desenvolvimento do trabalho.

Capítulo 2

Variável Aleatória, Análise Fatorial e Confiabilidade Estrutural

Este capítulo tem como objetivo expor os conceitos e definições sobre variáveis aleatórias que são necessários para a compreensão do trabalho. Para o desenvolvimento desta pesquisa, são destacadas as duas técnicas da Análise Multivariada (Análise Fatorial e Confiabilidade Estrutural) amplamente difundidas na literatura e com aplicações em várias áreas do conhecimento. Em seguida, é apresentado um estudo da arte sobre trabalhos relacionados com o tema proposto.

2.1 Variáveis Aleatórias Unidimensionais

Nessa seção é realizado um breve resumo sobre os conceitos estatísticos univariados, tendo como objetivo definir os principais parâmetros estatísticos de posição e dispersão de um conjunto de dados. Posteriormente, ver que tais parâmetros podem ser estendidos para o caso multivariado. Os resultados aqui apresentados estão disponíveis em (MENDES, 2010).

Sejam E um experimento aleatório e S o espaço amostral associado a esse experimento. Uma função X , que associa um número real x a cada resultado s do espaço amostral S , isto é, $X(s) = x$, é denominada variável aleatória. Em algumas situações, pode ocorrer mais de um resultado s do espaço amostral S levar à um mesmo valor x da variável aleatória X . Portanto, usamos a letra maiúscula X para representar uma variável aleatória unidimensional, e a correspondente letra minúscula x pode representar os valores que esta variável aleatória pode tomar.

Se a variável aleatória unidimensional X puder tomar um número finito de valores, ou um número infinito numerável de valores, diz-se que ela é variável aleatória discreta. Se a variável aleatória unidimensional X tomar um número infinito e não enumerável de valores, ela é dita uma variável aleatória contínua.

2.1.1 Variáveis Aleatórias Unidimensionais Discretas

Seja X uma variável aleatória unidimensional discreta. Isso significa que X assume um número finito de valores, ou um número infinito e enumerável de valores.

Para cada valor x que a variável aleatória X assume, será distribuído um número real, representando a probabilidade da variável aleatória X ser igual ao valor x . Esta probabilidade será denotada por $P(X = x) = p(x)$. Esses números reais $p(x)$, como representam probabilidades, devem satisfazer às seguintes condições:

1. $0 \leq p(x) \leq 1$;
2. $\sum_x p(x) = 1$.

As probabilidades $p(x)$, para todos os valores x que a variável aleatória X assume, formam o que se denomina função de probabilidade da variável aleatória X .

Para quaisquer valores a e b teremos $P(a \leq X \leq b) = \sum p(x)$, onde o somatório deve incluir todos os valores possíveis de X entre a e b inclusive.

Define-se função $f(x)$ como função de distribuição acumulada da variável aleatória discreta X , para todos os valores possíveis de X que a variável aleatória X assume, com $F(x) = P(X \leq x)$.

2.1.2 Variáveis Aleatórias Unidimensionais Contínuas

Seja X uma variável aleatória unidimensional contínua. Isto significa que X assume um número infinito não numerável de valores.

Define-se uma função, denotada por $f(x)$, e denominada função densidade de probabilidade da variável aleatória X , representando fisicamente, a curva das probabilidades de todos os infinitos valores da variável aleatória X . Esta função $f(x)$ deve satisfazer às seguintes condições:

1. $f(x) \geq 0$, para todo x , isto é, $f(x)$ é válida para todos os valores que a variável aleatória X assume.
2. $\int_x f(x)dx = 1$, ou seja, a área abaixo de $f(x)$, em todo o intervalo de validação da variável aleatória X , deve ser igual a 1.

Para quaisquer valores a e b pertencentes ao intervalo de validação da variável aleatória contínua X , teremos, $P(a \leq X \leq b) = \int_a^b f(x)dx$ representando a área abaixo de $f(x)$ no intervalo $[a, b]$.

Assim, teremos $P(X = x) = 0$, uma vez que $\int_a^b f(x)dx = 0$, o que nos leva a ter $P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b)$.

A função de distribuição acumulada da variável aleatória contínua X , denotada por $F(x)$, é dada por $F(x)$ no intervalo que compreende desde o limite inferior de validação da variável aleatória X até o valor de interesse x , conseqüentemente, $f(x) = \frac{d}{dx}F(x) = F'(x)$, para todo x no qual $F(x)$ seja derivável.

2.2 Estatística Univariada

Definição 2.1 - Valor Médio ou Esperança: Seja uma variável aleatória X , discreta que assume valores no conjunto x_1, x_2, \dots . Chama-se valor médio ou esperança matemática de X o valor

$$\mu = E = \sum_{i=1}^n x_i P(X = x_i) = \sum_{i=1}^n \frac{x_i}{n}$$

Definição 2.2 - Variância: Chama-se variância da variável aleatória X ao valor

$$\sigma^2 = V(X) = E[X - E(X)]^2 = \sum_{i=1}^n (x_i - \mu)^2 P(X = x_i) = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n}$$

Definição 2.3 - Desvio Padrão: A raiz quadrada da variância da variável aleatória X é denominada desvio padrão e é definido por

$$\sigma = \sqrt{V(X)}$$

Observação 2.4: Uma relação muito importante é, $V(X) = E(X^2) - [E(X)]^2$, onde $E(X^2) = \sum_{i=1}^n x_i^2 P(X = x_i)$.

Da mesma forma, se a variável aleatória contínua tem-se a esperança de X dada por

$$E(X) = \mu = \int_{-\infty}^{+\infty} xf(x)dx$$

então, a variância é dada por

$$V(X) = (X - \mu)^2 = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx.$$

A variância mede a dispersão (espalhamento) dos dados em torno da média $\mu = E(X)$ e o desvio padrão faz isso também, mas na mesma unidade de medida de dados.

Definição 2.5 - Covariância: Se as variáveis aleatórias X e Y não são independentes. Existe uma diferença entre $E(XY)$ e $E(X)E(Y)$, esta diferença é chamada de covariância e definida por

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

e se $Cov(X, Y) = 0$, as variáveis aleatórias são chamadas de não-correlacionadas.

Definição 2.6 - Coeficiente de Correlação: A covariância entre as variáveis aleatórias X e Y padronizadas é chamada de coeficiente de correlação

$$\rho = E \left[\left(\frac{X - E(X)}{\theta_x} \right) \left(\frac{Y - E(Y)}{\theta_y} \right) \right].$$

Proposição 2.7: $Cov(X, Y) = E(XY) - E(X)E(Y)$.

De fato!

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] = E[XY - XE(Y) - YE(X) + E(X)E(Y)] = \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) = E(XY) - E(X)E(Y). \end{aligned}$$

Os próximos resultados apresentam propriedades sobre a Esperança e a Variância de uma variável aleatória unidimensional.

Propriedades da Esperança: Sejam X uma variável aleatória e c uma constante, então:

1. $E(C) = C$;
2. $E(X_1 + X_2) = E(X_1) + E(X_2)$;
3. $E(cX_1) = cE(X_1)$.

Demonstração 2.8:

$$\text{Sejam } X_1 = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, X_2 = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, c \in R, \underline{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}.$$

Como visto anteriormente, $E(X_i) = \sum_{i=1}^m \frac{x_i}{m}$, no caso de variáveis discretas.

1. Fazendo-se $X = C$, temos

$$E(X) = \int_x xf(x)dx = \int_x Cf(x)dx = C \int_x f(x)dx = C,$$

uma vez que $\int_x f(x)dx = 1$.

Provaremos que $E(X_1 + X_2) = E(X_1) + E(X_2)$.

Seja (X_1, X_2) uma variável aleatória bidimensional com função densidade de probabilidade conjunta dada por $f(x_1, x_2)$; fazendo-se $W = H(X_1 + X_2) = X$ e $K = L(X_1 + X_2) = Y$, então

$$E(X_1 + X_2) = E(W + K) = \int_x \int_y [H(X_1 + X_2) + L(X_1 + X_2)] f(x_1, x_2) dx dy =$$

$$\int_x \int_y H(X_1 + X_2) f(x_1, x_2) dx dy + \int_x \int_y L(X_1 + X_2) f(x_1, x_2) dx dy =$$

$$E(H(X_1 + X_2)) + E(L(X_1 + X_2)) = E(X_1) + E(X_2).$$

2. $E(cX) = cE(X)$.

$$E(CX) = \int_x Cxf(x)dx = C \int_x xf(x)dx = CE(X).$$

■

Propriedades da Variância: Sejam X e Y variáveis aleatórias e c uma constante, então:

1. $V(cX) = c^2V(X)$;
2. $V(X) = E(X^2) - [E(X)]^2$;
3. $V(X + Y) = V(X) + V(Y)$.

Demonstração:

1. $V(cX) = c^2V(X)$ onde $V(X) = E[(X - \mu)(X - \mu)']$.

$$V(cX) = E[(cX - c\mu)(cX - c\mu)'] = E[c^2(X - \mu)(X - \mu)'] = c^2E[(X - \mu)(X - \mu)'] = c^2V(X).$$

2. $V(X) = E(X^2) - [E(X)]^2$, quando X é uma variável aleatória contínua.

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx = \int_{-\infty}^{+\infty} (x^2 - 2\mu_x x + \mu_x^2) f(x) dx = \\ &= \int_{-\infty}^{+\infty} x^2 f(x) dx - 2\mu_x \int_{-\infty}^{+\infty} x f(x) dx + \mu_x^2 \int_{-\infty}^{+\infty} f(x) dx = \\ &= E(X^2) - 2\mu_x E(X) + \mu_x^2 = E(X^2) - 2\mu_x \mu_x + \mu_x^2 = E(X^2) - 2\mu_x^2 + \mu_x^2 = \\ &= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - [E(X)]^2. \end{aligned}$$

3. $V(X + Y) = V(X) + V(Y)$.

$$\begin{aligned} V(X + Y) &= E[(X + Y)^2] - E[(X + Y)]^2 = E[X^2 + 2XY + Y^2] - [E(X) + E(Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 = \\ &= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 = V(X) + V(Y). \end{aligned}$$

■

2.2.1 Teste de Normalidade Lilliefors

Uma das consequências da aplicação da análise fatorial é a redução da dimensão do problema, substituindo as variáveis originais por novas variáveis não correlacionadas chamadas de escores fatoriais. Se caso, essas novas variáveis possuírem uma distribuição normal de probabilidade, é possível realizar o cálculo da estimativa das probabilidades de falha de cada evento de falha possível de ser monitorado. Por esse motivo, é apresentado um método para determinar se a distribuição de probabilidade é normal ou não, quando trabalhamos com dados amostrais.

O Teste de Lilliefors (LF) é uma modificação do teste de Kolmogorov-Smirnov (KS). O teste (KS) é adequado numa situação em que os parâmetros da distribuição hipotética são completamente conhecidos. No entanto, por vezes, é difícil inicialmente ou completamente especificar os parâmetros quando a distribuição é desconhecida. Neste caso, os parâmetros precisam ser estimados com base nos dados da amostra. Quando a estatística original (KS) é utilizada em tal situação, os resultados podem ser enganosos e a probabilidade do erro tipo I tende a ser menor do que as dadas na tabela-padrão do ensaio (KS) (Lilliefors, 1967).

Em contraste com o teste (KS), os parâmetros para o teste Lilliefors são estimados com base na amostra. Portanto, nesta situação, o teste (LF) é preferido sobre o teste (KS) (Razali, 2011). Dada uma amostra de observações, a estatística (LF) é definida como (Lilliefors, 1967),

$$D = \max_x |F^*(x) - S_n(X)|$$

onde $S_n(X)$ é a função de distribuição acumulada da amostra e $F^*(x)$ é a função de distribuição normal acumulada com média $\mu = \bar{x}$ e variância s^2 .

O teste estatístico (KS) testa as seguintes hipóteses:

$\{H_0: F(x) = F^*(x) \quad \forall x \in (-\infty, \infty)$ os dados seguem uma distribuição especificada;
 $H_1: F(x) \neq F^*(x)$ para algum x os dados não seguem uma distribuição especificada;

Mesmo que a estatística (LF) seja a mesma que a estatística (KS), a tabela para os valores críticos é diferente, o que leva à uma conclusão diferente sobre a normalidade dos dados (Mendes, 2003). A tabela de valores críticos para este teste pode ser encontrada na Tabela A15 de (Conover, 1999). Se D excede o valor crítico correspondente na tabela, então a hipótese nula é rejeitada.

2.3 Análise Multivariada

Os resultados apresentados nesta seção podem ser encontrados em Corrar, Paulo e Dias (2009), Johnson e Wichern (1987) e Mingoti (2007).

A análise multivariada é um conjunto de técnicas estatísticas que tratam dos dados correspondentes às medidas de muitas variáveis simultaneamente. Basicamente, a análise multivariada consiste no estudo estatístico dos problemas relacionados com:

- Inferências sobre médias multivariadas;
- Análise da estrutura de covariância de uma matriz de dados;
- Técnicas de reconhecimento de padrão, classificação e agrupamento.

No estudo de $p \geq 1$ variáveis, geralmente, toma-se n observações de cada variável para obter informações sobre parâmetros, relacionamentos entre as variáveis, comparações, etc. Assim, as medidas registradas são x_{ij} com $i = \{1, 2, \dots, n\}$ (observações) e $j = 1, 2, \dots, p$ (variáveis) que podem ser agrupadas na matriz de dados $X_{n \times p}$, com n linhas e p colunas.

$$X_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

A matriz de dados $X_{n \times p}$ contém n observações do vetor aleatório p -dimensional $\underline{X}' = [X_1, X_2, \dots, X_p]$.

2.3.1 Estatística Multivariada

Muitas das informações contidas na matriz de dados podem ser dadas pelo cálculo de números sumários conhecidos como estatísticas descritivas.

Definição 2.10 - Vetor Média Amostral:

$$\underline{\bar{X}}' = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p],$$

com $\bar{X}_j = \frac{\sum_{i=1}^n x_{ij}}{n}, j = 1, 2, \dots, p.$

Definição 2.11 - Matriz de Covariância Amostral:

$$S = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{pmatrix}$$

onde,

$S_{ij} = \frac{\sum_{i=1}^n (x_{ij} - \bar{X}_j)^2}{n-1}$ é a variância da variável aleatória x_j .

$S_{ij} = \frac{\sum_{i=1}^n (x_{ij} - \bar{X}_j)(x_{ik} - \bar{X}_k)}{n-1}; j, k = 1, 2, \dots, p$ é a covariância entre x_j e x_k .

Observação 2.12: A matriz de covariância amostral de uma matriz multivariada também pode ser definida como

$$V(\underline{X}) = E \left[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' \right].$$

Essa observação apresentada acima, contribui para demonstração de vários resultados que serão mostrados adiante.

Definição 2.13 - Matriz de Correlação Amostral:

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}$$

onde, $r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}}$

Proposição 2.14: Sejam $\underline{X} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}$ e $\underline{\mu} = \begin{pmatrix} \mu_1 & \mu_2 \\ \mu_1 & \mu_2 \end{pmatrix}$. Defina $\underline{X} - \underline{\mu} = \begin{pmatrix} x_1 - \mu_1 & y_1 - \mu_2 \\ x_2 - \mu_1 & y_2 - \mu_2 \end{pmatrix}$ e $\underline{c}' = (c_1 \ c_2)$. Então

$$\underline{c}' (\underline{X} - \underline{\mu}) = \underline{c}' \underline{X} - \underline{c}' \underline{\mu}.$$

Demonstração 2.15:

$$\begin{aligned} \underline{c}' (\underline{X} - \underline{\mu}) &= \underline{c}' = (c_1 \ c_2) \begin{pmatrix} x_1 - \mu_1 & y_1 - \mu_2 \\ x_2 - \mu_1 & y_2 - \mu_2 \end{pmatrix} = \\ & [c_1(x_1 - \mu_1) + c_2(x_2 - \mu_1) \quad c_1(y_1 - \mu_2) + c_2(y_2 - \mu_2)] = \\ & [(c_1x_1 + c_2x_2) + (-c_1\mu_1 - c_2\mu_1) \quad (c_1y_1 + c_2y_2) + (-c_1\mu_2 - c_2\mu_2)] = \\ & [c_1x_1 + c_2x_2 \quad c_1y_1 + c_2y_2] + [-c_1\mu_1 - c_2\mu_1 \quad -c_1\mu_2 - c_2\mu_2] = \\ & (c_1 \ c_2) \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} - [(c_1 + c_2)\mu_1 \quad (c_1 + c_2)\mu_2] = \end{aligned}$$

$$\underline{c}'\underline{X} - [c_1\mu_1 + c_2\mu_1 \quad c_1\mu_2 + c_2\mu_2] = \underline{c}'\underline{X} - (c_1 \quad c_2) \begin{pmatrix} \mu_1 & \mu_2 \\ \mu_1 & \mu_2 \end{pmatrix} = \underline{c}'\underline{X} - \underline{c}'\underline{\mu}.$$

De forma análoga, mostra-se que $(\underline{X} - \underline{\mu})'\underline{c} = (\underline{c}'\underline{X} - \underline{c}'\underline{\mu})'$.

■

Proposição 2.16 – Propriedades de uma Matriz Multivariada: Sejam \underline{X} uma matriz multivariada, c uma constante e \underline{c}' um vetor de constantes, então:

1. $E(c\underline{X}) = cE(\underline{X})$;
2. $E(\underline{c}'\underline{X}) = \underline{c}'E(\underline{X})$;
3. $V(c\underline{X}) = c^2V(\underline{X})$;
4. $V(\underline{c}'\underline{X}) = \underline{c}'V(\underline{X})\underline{c}$.

Demonstração 2.17: Sejam $c \in R$ uma constante, $\underline{c}' = [c_{11}, c_{12}, \dots, c_{1p}]$ onde $c_{ij} \in R$;

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \text{ e } E(\underline{X}) = [\mu_1, \mu_2, \dots, \mu_p] \text{ onde } \mu_i = \sum_{i=1}^m \frac{x_i}{m}. \text{ Então:}$$

1. $E(c\underline{X}) = cE(\underline{X})$.

$$\begin{aligned} E(c\underline{X}) &= E \left[c \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \right] = E \left[\begin{pmatrix} cX_1 \\ cX_2 \\ \vdots \\ cX_p \end{pmatrix} \right] = \begin{bmatrix} E[cX_1] \\ E[cX_2] \\ \vdots \\ E[cX_p] \end{bmatrix} = \begin{bmatrix} cE[X_1] \\ cE[X_2] \\ \vdots \\ cE[X_p] \end{bmatrix} = \\ &= c \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \underline{c}'E(\underline{X}). \end{aligned}$$

2. Queremos demonstrar que $E(\underline{c}'\underline{X}) = \underline{c}'E(\underline{X})$;

$$E(\underline{c}'\underline{X}) = E \left[(c_{11} \quad c_{12} \quad \dots \quad c_{1p}) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \right] = E[c_{11}X_1 + c_{12}X_2 + \dots + c_{1p}X_p] =$$

$$E[c_{11}X_1] + E[c_{12}X_2] + \dots + E[c_{1p}X_p] = c_{11}E(X_1) + c_{12}E(X_2) + \dots + c_{1p}E(X_p) =$$

$$[c_{11} \quad c_{12} \quad \dots \quad c_{1p}] \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \underline{c}'E(\underline{X}).$$

3. $V(c\underline{X}) = c^2V(\underline{X})$ onde $V(\underline{X}) = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})']$.

$$\begin{aligned} V(c\underline{X}) &= E[(c\underline{X} - c\underline{\mu})(c\underline{X} - c\underline{\mu})'] = E[c(\underline{X} - \underline{\mu})c(\underline{X} - \underline{\mu})'] = \\ &= E[c^2(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'] = c^2E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'] = c^2V(\underline{X}). \end{aligned}$$

$$4. \quad V(\underline{c}'\underline{X}) = \underline{c}'V(\underline{X})\underline{c}.$$

$$\begin{aligned} V(\underline{c}'\underline{X}) &= E \left[(\underline{c}'\underline{X} - \underline{c}'\underline{\mu}) (\underline{c}'\underline{X} - \underline{c}'\underline{\mu})' \right] = E \left[\underline{c}' (\underline{X} - \underline{\mu}) [\underline{c}' (\underline{X} - \underline{\mu})]' \right] = \\ &= E \left[\underline{c}' (\underline{X} - \underline{\mu}) (\underline{X} - \underline{\mu})' \underline{c} \right] = \underline{c}' E \left[(\underline{X} - \underline{\mu}) (\underline{X} - \underline{\mu})' \right] \underline{c} = \underline{c}' V(\underline{X}) \underline{c}. \end{aligned}$$

■

2.3.2 Teorema Espectral

Os métodos da análise multivariada que são apresentados no decorrer do trabalho, exigem o conhecimento de propriedades das matrizes reais simétricas. Essa necessidade vem do fato de se trabalhar com as matrizes multivariadas geradas a partir das variáveis aleatórias envolvidas na análise e, conseqüentemente, com suas matrizes de covariância e de correlação, que são obrigatoriamente simétricas. Os conceitos algébricos apresentados neste momento, assim como seus resultados e demonstrações, podem ser encontrados em (STEVEN, 2011).

Matriz Hermitiana

Seja $M = (m_{ij})$ uma matriz $n \times n$ com $m_{ij} = a_{ij} + ib_{ij}$ para todo i e j . Podemos escrever M sob a forma

$$M = A + iB$$

onde $A = (a_{ij})$ e $B = (b_{ij})$ tem elementos reais. Definimos a conjugada de M como

$$\overline{M} = A - iB$$

Logo, \overline{M} é a matriz formada fazendo o conjugado de cada elemento de M . A transposta de \overline{M} será escrita M^H . O espaço vetorial de todas as matrizes $m \times n$ com elementos complexos é escrito $C^{m \times n}$. Se A e B são elementos de $C^{m \times n}$ e $C \in C^{m \times n}$, então as seguintes regras são verificadas:

1. $(A^H)^H = A$;
2. $(\alpha A + \beta B)^H = \overline{\alpha} A^H + \overline{\beta} B^H$;
3. $(AC)^H = C^H A^H$.

Definição 2.18 - Matriz Hermitiana: Uma matriz M é dita hermitiana se $M = M^H$.

Observação 2.19: Se M é uma matriz com elementos reais, então $M^H = M^T$. Em particular, se M é uma matriz real simétrica, então M é hermitiana. Então, podemos encarar as matrizes hermitianas como o análogo complexo das matrizes reais simétricas. As matrizes hermitianas tem muitas propriedades interessantes, como veremos no próximo teorema.

Teorema 2.20: Os autovalores de uma matriz hermitiana são todos reais. Além disso, autovetores associados a diferentes autovalores são ortogonais.

Definição 2.21 - Matriz Unitária: Uma matriz $n \times n$, U , é dita unitária se seus vetores coluna formam um conjunto ortonormal em C^n .

Corolário 2.22: Se os autovalores de uma matriz hermitiana A são distintos, então existe uma matriz unitária U que diagonaliza A .

Teorema 2.23 - Teorema de Schur: Para cada matriz $n \times n$, A , existe uma matriz unitária U tal que $U^H A U$ é triangular superior.

Teorema 2.24 - A decomposição Real de Schur: Se A é uma matriz $n \times n$ com elementos reais, então A pode ser fatorada em um produto $Q T Q^T$, no qual Q é uma matriz ortogonal e T está na forma de Schur.

Demonstração 2.25: No caso de $n = 2$ se os autovalores de A são reais, nós podemos fazer \underline{q}_1 um autovetor unitário associado ao primeiro autovalor λ_1 e fazer \underline{q}_2 qualquer vetor unitário ortogonal a \underline{q}_1 . Se fizermos $Q = (\underline{q}_1 \underline{q}_2)$ então Q é uma matriz ortogonal. Se fizermos $T = Q^T A Q$, então a primeira coluna de T é

$$Q^T A \underline{q}_1 = \lambda_1 Q^T \underline{q}_1 = \lambda_1 \underline{e}_1$$

Assim, T é triangular superior e $A = Q T Q^T$. Se os autovalores de A são complexos, então simplesmente definimos $T = A$ e $Q = I$. Então, toda matriz 2×2 real tem uma decomposição de Schur real. Agora seja A uma matriz $k \times k$ em que $k \geq 3$ e suponha-se que, para $2 \leq m < k$, cada matriz $m \times m$ real tem uma decomposição de Schur. Seja λ_1 um autovalor de A . Se λ_1 é real, seja \underline{q}_1 um autovetor unitário associado a λ_1 e escolham-se $\underline{q}_2, \underline{q}_3, \dots, \underline{q}_n$ de modo que $Q_1 = (\underline{q}_1, \underline{q}_2, \dots, \underline{q}_n)$ seja uma matriz ortogonal. Tal como na prova do Teorema de Schur, segue-se da primeira coluna de $Q_1^T A Q_1$ será $\lambda_1 \underline{e}_1$. No caso em que λ_1 é complexo, seja $z = x + iy$ (em que x e y são reais) um autovetor associado a λ_1 e seja $S = \text{Cob}(x, y)$, então $\dim S = 2$ e S é invariante sobre A . Seja $\underline{q}_1, \underline{q}_2$ uma base ortonormal para S . Escolha $\underline{q}_1, \underline{q}_2, \dots, \underline{q}_n$ para que $Q_1 = (\underline{q}_1, \underline{q}_2, \dots, \underline{q}_n)$ seja uma matriz ortogonal. Desde que S é invariante sob A , segue-se que

$$A \underline{q}_1 = b_{11} \underline{q}_1 + b_{21} \underline{q}_2$$

e

$$A \underline{q}_2 = b_{12} \underline{q}_1 + b_{22} \underline{q}_2$$

para alguns escalares $b_{11}, b_{21}, b_{12}, b_{22}$ e, portanto, as duas primeiras colunas de $Q_1^T A Q_1$ serão

$$\left(Q_1^T A \underline{q}_1, Q_1^T A \underline{q}_2 \right) = (b_{11} \underline{e}_1 + b_{21} \underline{e}_2, b_{12} \underline{e}_1 + b_{22} \underline{e}_2)$$

Assim, em geral, $Q_1^T A Q_1$ será uma matriz em blocos

$$Q_1^T A Q_1 = \begin{pmatrix} B_1 & X \\ 0 & A_1 \end{pmatrix}$$

onde, $B_1 = \lambda_1$ e A_1 é $(k-1) \times (k-1)$ se λ_1 é real. Caso λ_1 seja complexo, B_1 é 2×2 e A_1 é $(k-2) \times (k-2)$

Em ambos os casos, pode-se aplicar nossa hipótese de indução a A_1 e obter uma decomposição de Schur $A_1 = UT_1U^T$. Supondo que a forma de Schur T_1 tem $j - 1$ blocos diagonais, B_2, B_3, \dots, B_j . Se fizermos

$$Q_2 = \begin{pmatrix} I & O \\ O & Q_1 \end{pmatrix} \text{ e } Q = Q_1Q_2$$

Então, tanto Q_1 quanto Q_2 são matrizes ortogonais $k \times k$. Se, em seguida, definirmos $T = Q^T A Q$, obteremos uma matriz na fórmula de Schur, e segue-se que A terá decomposição de Schur QTQ^T .

■

No caso em que todos os autovalores de A são reais, a forma real de Schur T será triangular superior. No caso em que A é real e simétrica, então, uma vez que todos os autovalores de A são reais, T deve ser triangular superior, porém, neste caso T também deve ser simétrica. Então, vamos acabar com a diagonalização de A . Assim, para matrizes simétricas reais, temos a seguinte versão do Teorema Espectral.

Teorema 2.26 - Teorema da Decomposição Espectral (TDE) - Matrizes Simétricas Reais: Se A é uma matriz real simétrica, então existe uma matriz ortogonal Q que diagonaliza A , isto é, $Q^T A Q = D$, na qual D é diagonal.

Observação 2.27: Podemos denotar como P uma matriz gerada pelos autovetores normalizados da matriz de covariância $V(\underline{X})$ e de Λ uma matriz diagonal formada pelos autovalores de $V(\underline{X})$.

A análise de componentes principais que é apresentada a seguir exige o cálculo dos autovalores e autovetores a partir da matriz de covariância ou de correlação. O próximo teorema afirma que, se os autovalores forem todos positivos, pode-se concluir que a matriz de covariância ou de correlação é positiva definida.

Teorema 2.28: Seja A uma matriz real simétrica $n \times n$. Então, A é positiva definida se e somente se todos os seus autovalores são positivos.

2.3.3 Componentes Principais

Conforme Lattin (2011), a análise dos componentes principais é um método que pode ser usado para reduzir a dimensionalidade dos dados multivariados. Ela permite que o pesquisador reexpresse os dados (fazendo combinações lineares das variáveis originais) para que as primeiras poucas variáveis novas resultantes (chamadas componentes) respondam por tantas informações disponíveis quanto possível. Se uma redundância substancial estiver presente no conjunto de dados, então é possível explicar a maioria das informações do conjunto original de dados com um número relativamente pequeno de componentes. Essa redução da dimensão torna mais direta a visualização dos dados e sua subsequente análise, mais administrável.

Ainda, segundo Lattin (2011), o pesquisador deve decidir quantos componentes principais reter para a análise subsequente, fazendo uma difícil escolha entre simplicidade (um

pequeno número de dimensões é mais fácil de administrar) e completude (um grande número de dimensões capta uma quantidade maior de informações disponíveis).

A análise de componentes principais é um método utilizado para reorganizar um conjunto de variáveis em um novo conjunto. Ela procura explicar a estrutura de variância-covariância da matriz de dados através de combinações lineares não correlacionadas das p variáveis originais. As variáveis do novo conjunto são chamadas de componentes e são geradas de forma que o primeiro componente contenha a maior parcela possível da informação originalmente existente. O segundo componente contém menos informação que o primeiro e mais que os subsequentes, assim, sucessivamente. O número de componentes principais é sempre igual ao número de variáveis originais. Portanto, os objetivos da Análise de Componentes Principais são: redução de dados; obtenção de variáveis aleatórias não correlacionadas e interpretação.

Componentes Principais da População

Algebricamente, componentes principais são combinações lineares particulares das p variáveis aleatórias X_1, X_2, \dots, X_p . Geometricamente, estas combinações lineares representam a seleção de um novo sistema de coordenadas obtido por rotação do sistema original com X_1, X_2, \dots, X_p como eixos. Os novos eixos Y_1, Y_2, \dots, Y_p representam as direções com variabilidade máxima e fornecem uma descrição mais simples da estrutura de covariância (Johnson, 1987).

As componentes principais dependem da matriz de covariâncias Σ (ou da matriz de correlação ρ) das variáveis aleatórias X_1, X_2, \dots, X_p . O seu desenvolvimento não necessita da suposição de Gaussianidade.

Seja o vetor aleatório $\underline{X}' = [X_1, X_2, \dots, X_p]$, com vetor de médias $\underline{\mu} = E(\underline{X})$ e matriz de covariância $\Sigma = V(\underline{X})$, com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Considere as combinações lineares

$$Y_1 = \underline{c}'_1 \underline{X} = c_{11}X_1 + c_{21}X_2 + \dots + c_{p1}X_p$$

$$Y_2 = \underline{c}'_2 \underline{X} = c_{12}X_1 + c_{22}X_2 + \dots + c_{p2}X_p$$

$$\vdots$$

$$Y_p = \underline{c}'_p \underline{X} = c_{1p}X_1 + c_{2p}X_2 + \dots + c_{pp}X_p$$

$$C_{p \times p} = \begin{pmatrix} c_{11} & c_{21} & \dots & c_{p1} \\ c_{12} & c_{22} & \dots & c_{p2} \\ \vdots & \vdots & \ddots & \dots \\ c_{1p} & c_{2p} & \dots & c_{pp} \end{pmatrix}$$

$$\text{Então, } \underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix} = \begin{pmatrix} c_{11} & c_{21} & \dots & c_{p1} \\ c_{12} & c_{22} & \dots & c_{p2} \\ \vdots & \vdots & \ddots & \dots \\ c_{1p} & c_{2p} & \dots & c_{pp} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = C_{p \times p} \underline{X}.$$

Proposição 2.29: Seja \underline{X} uma matriz multivariada e \underline{c}_i um vetor de constantes com dimensões apropriadas, então $V(Y_i) = \underline{c}'_i V(\underline{X}) \underline{c}_i$.

Demonstração 2.30:

$$\begin{aligned} V(Y_i) &= V(\underline{c}'_i \underline{X}) = E \left[(\underline{c}'_i \underline{X} - \underline{c}'_i \underline{\mu}) (\underline{c}'_i \underline{X} - \underline{c}'_i \underline{\mu})' \right] = E \left[\underline{c}'_i (\underline{X} - \underline{\mu}) (\underline{X} - \underline{\mu})' \underline{c}_i \right] = \\ &= \underline{c}'_i E \left[(\underline{X} - \underline{\mu}) (\underline{X} - \underline{\mu})' \right] \underline{c}_i = \underline{c}'_i V(\underline{X}) \underline{c}_i. \end{aligned}$$

■

O próximo resultado afirma que, o máximo que uma forma quadrática, para pontos na esfera unitária, pode assumir, a partir de uma matriz simétrica B , é exatamente o valor do maior autovalor relacionado aos seus autovetores.

Proposição 2.31 - Maximização de Formas Quadráticas para Pontos na Esfera Unitária: Seja a matriz B de ordem $p \times p$ simétrica positiva definida com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ e com os respectivos autovetores padronizados $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$. Então, o $\max = \frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \lambda_1$ que é alcançado em $\underline{x} = \underline{e}_1$ e o $\min = \frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \lambda_p$ que é alcançado em $\underline{x} = \underline{e}_p$.

Demonstração 2.32: Como B é positiva definida, temos que $\underline{x}' B \underline{x} > 0, \forall \underline{x} \neq 0$. Como B é simétrica, pelo Teorema da Decomposição Espectral (TDE) 2.26 podemos escrever $B = P \Lambda P'$, onde P é uma matriz ortogonal normalizada, formada pelos autovetores de B e Λ é uma matriz diagonal formada pelos autovalores de B . Assim note que,

$$\begin{aligned} (P \Lambda^{\frac{1}{2}} P') (P \Lambda^{\frac{1}{2}} P') &= P \Lambda^{\frac{1}{2}} P' P \Lambda^{\frac{1}{2}} P' = P \Lambda^{\frac{1}{2}} I \Lambda^{\frac{1}{2}} P' = P \Lambda P' = B \Rightarrow \\ (P \Lambda^{\frac{1}{2}} P')^2 &= B \Rightarrow B^{\frac{1}{2}} = P \Lambda^{\frac{1}{2}} P' \end{aligned}$$

onde,

$$\Lambda^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\Lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\Lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\Lambda_p} \end{pmatrix}$$

Assim,

$$\frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \frac{\underline{x}' B^{\frac{1}{2}} B^{\frac{1}{2}} \underline{x}}{\underline{x}' P P' \underline{x}} = \frac{\underline{x}' P \Lambda^{\frac{1}{2}} P' P \Lambda^{\frac{1}{2}} P' \underline{x}}{\underline{x}' P P' \underline{x}} = \frac{\underline{x}' P \Lambda P' \underline{x}}{\underline{x}' P P' \underline{x}} = \frac{\underline{Y}' \Lambda \underline{Y}}{\underline{Y}' \underline{Y}} = \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2}.$$

Isso ocorre, pois

$$(y_1 \ y_2 \ \dots \ y_p) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = y_1^2 \lambda_1 + y_2^2 \lambda_2 + \dots + y_p^2 \lambda_p.$$

Então,

$$\frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2} < \frac{\sum_{i=1}^p \lambda_1 y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_1 \frac{\sum_{i=1}^p y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_1.$$

Fazendo, $\underline{x} = \underline{e}_1$ tem-se $\underline{y} = P' \underline{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ e substituindo, temos

$$B = \frac{\underline{e}'_1 B \underline{e}_1}{\underline{e}'_1 \underline{e}_1} = \frac{\underline{Y}' \Lambda \underline{Y}}{\underline{Y}' \underline{Y}} = \frac{\underline{e}'_1 P \Lambda P' \underline{e}_1}{\underline{e}'_1 P P' \underline{e}_1} = \frac{(1 \ 0 \ \dots \ 0) \Lambda (1 \ 0 \ \dots \ 0)'}{(1 \ 0 \ \dots \ 0) (1 \ 0 \ \dots \ 0)'} = \frac{\lambda_1}{1} = \lambda_1.$$

Pois,

$$(1 \ 0 \ \dots \ 0) \Lambda (1 \ 0 \ \dots \ 0)' = (1 \ 0 \ \dots \ 0) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} =$$

$$(1 \ 0 \ \dots \ 0) \begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \lambda_1.$$

Analogamente, o $\min = \frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \lambda_p$ em $\underline{x} = \underline{e}_p$.

Como visto $\underline{x}' B \underline{x} > 0$ é positiva definida e como B é simétrica pelo teorema 2.26, podemos escrever $B = P \Lambda P'$ e $B^{\frac{1}{2}} = P \Lambda^{\frac{1}{2}} P'$. Assim,

$$\frac{\underline{x}' B \underline{x}}{\underline{x}' \underline{x}} = \frac{\underline{x}' P \Lambda^{\frac{1}{2}} P' P \Lambda^{\frac{1}{2}} P' \underline{x}}{\underline{x}' P P' \underline{x}} = \frac{\underline{x}' P \Lambda P' \underline{x}}{\underline{x}' P P' \underline{x}} = \frac{\underline{Y}' \Lambda \underline{Y}}{\underline{Y}' \underline{Y}} =$$

$$(y_1 \ y_2 \ \dots \ y_p) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = (y_1 \ y_2 \ \dots \ y_p) \begin{pmatrix} \lambda_1 y_1 \\ \lambda_2 y_2 \\ \vdots \\ \lambda_p y_p \end{pmatrix} =$$

$y_1^2 \lambda_1 + y_2^2 \lambda_2 + \dots + y_p^2 \lambda_p$, então,

$$\frac{\underline{Y}' \Lambda \underline{Y}}{\underline{Y}' \underline{Y}} = \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2} > \frac{\sum_{i=1}^p \lambda_p y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_p.$$

Fazendo $\underline{x} = \underline{e}_p$, tem-se $\underline{y} = P' \underline{e}_p = (0 \ 0 \ \dots \ 1)'$ substituindo temos

$$\frac{\underline{e}'_p B \underline{e}_p}{\underline{e}'_p \underline{e}_p} = \frac{\underline{Y}' \Lambda \underline{Y}}{\underline{Y}' \underline{Y}} = \frac{\underline{e}'_p P \Lambda P' \underline{e}_p}{\underline{e}'_p P P' \underline{e}_p} = \frac{(0 \ 0 \ \dots \ 1)' \Lambda (0 \ 0 \ \dots \ 1)}{(0 \ 0 \ \dots \ 1)' (0 \ 0 \ \dots \ 1)} = \frac{\lambda_p}{1} = \lambda_p.$$

■

Observação 2.33: Do resultado anterior, podemos afirmar que $\max_{\underline{x}' \underline{x}} = \lambda_1$ com $\underline{x} = \underline{e}_1$, pois a matriz Σ satisfaz as condições de 2.31, por esse motivo definimos as componentes principais como sendo $Y_i = \underline{e}'_i \underline{X}$. Dessa forma, a primeira componente principal possui variância máxima igual a λ_1 . De fato!

$$V(Y_i) = V(\underline{e}'_i \underline{X}) = \underline{e}'_i V(\underline{X}) \underline{e}_i = \frac{\underline{e}'_i \Sigma \underline{e}_i}{\underline{e}'_i \underline{e}_i}, \text{ mostremos que } \Sigma \underline{e}_k = \lambda_k \underline{e}_k.$$

$$\begin{aligned} \Sigma \underline{e}_k &= P \Lambda P' \underline{e}_k = P \Lambda (P' \underline{e}_k) = P \Lambda \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = P \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \lambda_k \\ \vdots \\ 0 \end{pmatrix} = \\ &= (e_1 \ e_2 \ \dots \ e_k \ \dots \ e_p) \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \lambda_k \\ \vdots \\ 0 \end{pmatrix} = \lambda_k \underline{e}_k. \end{aligned}$$

Portanto, $V(Y_i) = \underline{e}'_i \Sigma \underline{e}_i$, logo

$$V(Y_i) = \underline{e}'_i \sum \underline{e}_i = (e_{11} \ e_{21} \ \dots \ e_{p1}) \begin{pmatrix} e_{11} \lambda_1 \\ e_{21} \lambda_1 \\ \vdots \\ e_{p1} \lambda_1 \end{pmatrix} =$$

$$e_{11} e_{11} \lambda_1 + e_{21} e_{21} \lambda_1 + \dots + e_{p1} e_{p1} \lambda_1 = \lambda_1 [e_{11} e_{11} + e_{21} e_{21} + \dots + e_{p1} e_{p1}] = \lambda_1 \cdot 1 = \lambda_1.$$

O próximo resultado mostra que, se as componentes principais são definidas da forma $Y_i = \underline{e}'_i \underline{X}$, então elas são não correlacionadas.

Proposição 2.34: Sejam Y_i e Y_k componentes principais da matriz de covariância $\Sigma = V(\underline{X})$, então $Cov(Y_i, Y_k) = 0$.

Demonstração 2.35: Como $Y_i = \underline{e}'_i \underline{X}$, então

$$\begin{aligned} Cov(Y_i, Y_k) &= Cov(\underline{e}'_i \underline{X}, \underline{e}'_k \underline{X}) = E \left[(\underline{e}'_i \underline{X} - \underline{e}'_i \underline{\mu}) (\underline{e}'_k \underline{X} - \underline{e}'_k \underline{\mu})' \right] = \\ &= E \left[\underline{e}'_i (\underline{X} - \underline{\mu}) (\underline{X} - \underline{\mu})' \underline{e}'_k \right] = \underline{e}'_i E \left[(\underline{X} - \underline{\mu}) (\underline{X} - \underline{\mu})' \right] \underline{e}'_k = \underline{e}'_i V(\underline{X}) \underline{e}'_k = \underline{e}'_i \sum \underline{e}'_k \end{aligned}$$

Pelo Teorema da Decomposição Espectral 2.26 temos,

$$\begin{aligned} \underline{e}'_i \sum \underline{e}'_k &= \underline{e}'_i P \Lambda P' \underline{e}'_k = \underline{e}'_i P \Lambda \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \underline{e}'_i P \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \lambda_k \\ \vdots \\ 0 \end{pmatrix} = \underline{e}'_i \begin{pmatrix} e_{1k} \lambda_k \\ e_{2k} \lambda_k \\ \vdots \\ e_{pk} \lambda_k \end{pmatrix} = \\ & \begin{pmatrix} e_{1i} & e_{2i} & \dots & e_{pi} \end{pmatrix} \begin{pmatrix} e_{1k} \lambda_k \\ e_{2k} \lambda_k \\ \vdots \\ e_{pk} \lambda_k \end{pmatrix} = e_{1i} e_{1k} \lambda_k + e_{2i} e_{2k} \lambda_k + \dots + e_{pi} e_{pk} \lambda_k = \\ & \lambda_k [e_{1i} e_{1k} + e_{2i} e_{2k} + \dots + e_{pi} e_{pk}] = \lambda_k \cdot 0 = 0. \end{aligned}$$

■

O seguinte resultado mostra que o traço de matriz de covariância, coincide com a soma das variâncias das componentes principais.

Proposição 2.36: Sejam \underline{X} uma matriz multivariada e $\Sigma V(\underline{X})$ a matriz de covariância de \underline{X} , então $tr(\Sigma) = \lambda_1 + \lambda_2 + \dots + \lambda_p = V(Y_1) + V(Y_2) + \dots + V(Y_p)$.

Demonstração 2.37: Pelo Teorema da Decomposição Espectral 2.26, temos $\Sigma = P \Lambda P'$, então:

$$tr(\Sigma) = tr(P \Lambda P')$$

Pela propriedade de traço de matrizes podemos escrever:

$$\begin{aligned} tr(P \Lambda P') &= tr(P P' \Lambda) = tr(I_n \Lambda) = tr(\Lambda) = tr \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix} = \\ & \lambda_1 + \lambda_2 + \dots + \lambda_p = V(Y_1) + V(Y_2) + \dots + V(Y_p). \end{aligned}$$

■

Observação 2.38: Como o $tr(\Sigma)$ é a soma das variâncias das variáveis aleatórias envolvidas, o resultado anterior mostra que a variância total populacional é igual a soma das variâncias das componentes principais. E, conseqüentemente, a proporção da variância total explicada pela k -ésima componente principal é

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, k = 1, 2, \dots, p$$

Os resultados anteriores mostraram que as componentes principais possuem variância máxima e são não correlacionadas. Apesar dessas propriedades serem de grande valia, elas ainda não justificam a utilização das componentes principais para analisar o conjunto de variáveis em substituição das variáveis originais. Para tanto precisamos verificar se existem correlações entre as componentes principais e a variáveis aleatórias originais.

O próximo resultado afirma que essas correlações existem e, além disso, podem ser encontradas a partir dos autovetores e autovalores da matriz de covariância $\Sigma = V(\underline{X})$.

Proposição 2.39: Sejam X_k uma variável aleatória da matriz multivariada \underline{X} e Y_i uma componente principal, então $\rho(X_k, Y_i) = \frac{e_{ki}\sqrt{\lambda_i}}{\sigma_k}$.

Demonstração 2.40: Seja $\underline{c}'_k = (0 \quad \dots \quad 1 \quad \dots \quad 0)$ vetor de valores 0, com 1 na k -ésima posição, e $\underline{X}' = (X_1 \quad X_2 \quad \dots \quad X_k \quad \dots \quad X_p)$, então podemos escrever uma variável aleatória X_k como $X_k = \underline{c}'_k \underline{X}$, pois,

$$\underline{c}'_k \underline{X} = (0 \quad \dots \quad 1 \quad 0 \quad \dots \quad 0) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \\ \vdots \\ X_p \end{pmatrix} = X_k.$$

Além disso, pela proposição 2.16 item (ii)

$$\mu_k = E(X_k) = E(\underline{c}'_k \underline{X}) = \underline{c}'_k E(\underline{X}) = \underline{c}'_k \underline{\mu} \quad \text{e} \quad \mu_i = E(Y_i) = E(\underline{e}'_i \underline{X}) = \underline{e}'_i E(\underline{X}) = \underline{e}'_i \underline{\mu}.$$

Por definição tem-se que $\rho(X_k, Y_i) = \frac{\text{Cov}(X_k, Y_i)}{\sqrt{V(X_k)V(Y_i)}}$ e $\text{Cov}(X_k, Y_i) = E[(X_k - \mu_k)(Y_i - \mu_i)]$. Então, aplicando novamente os resultados proposição 2.16 itens (ii) e (iv), tem-se,

$$\begin{aligned} \text{Cov}(X_k, Y_i) &= \text{Cov}(\underline{c}'_k \underline{X}, \underline{e}'_i \underline{X}) = E\left[(\underline{c}'_k \underline{X} - \underline{c}'_k \underline{\mu})(\underline{e}'_i \underline{X} - \underline{e}'_i \underline{\mu})'\right] = \\ &= E\left[\underline{c}'_k (\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' \underline{e}_i\right] = \underline{c}'_k E\left[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\right] \underline{e}_i = \underline{c}'_k \sum \underline{e}_i. \end{aligned}$$

Sabe-se que $\sum \underline{e}_i = \lambda_i \underline{e}_i$, logo,

$$\text{Cov}(X_k, Y_i) = \underline{c}'_k \sum \underline{e}_i = \underline{c}'_k \lambda_i \underline{e}_i = \lambda_i e_{ki}$$

Pois,

$$\underline{c}'_k \lambda_i \underline{e}_i = \underline{c}'_k \begin{pmatrix} \lambda_i e_{1i} \\ \lambda_i e_{2i} \\ \vdots \\ \lambda_i e_{ki} \\ \vdots \\ \lambda_i e_{pi} \end{pmatrix} = (0 \quad \dots \quad 1 \quad 0 \quad \dots \quad 0) \begin{pmatrix} \lambda_i e_{1i} \\ \lambda_i e_{2i} \\ \vdots \\ \lambda_i e_{ki} \\ \vdots \\ \lambda_i e_{pi} \end{pmatrix} = \lambda_i e_{ki}$$

Como as variâncias de X_k e Y_i são $V(Y_i) = V(\underline{e}'_i \underline{X}) = \underline{e}'_i \sum \underline{e}_i = \underline{e}'_i \lambda_i \underline{e}_i = \lambda_i$ e $V(X_k) = \sigma_k^2$, tem-se que,

$$\rho(X_k, Y_i) = \frac{\text{Cov}(X_k, Y_i)}{\sqrt{V(X_k)V(Y_i)}} = \frac{\lambda_i e_{ki}}{\sigma_k \sqrt{\lambda_i}} = \frac{e_{ki} \sqrt{\lambda_i}}{\sigma_k}.$$

■

2.4 Análise Fatorial

Nesta seção são apresentados os conceitos teóricos da Análise Fatorial (AF). Todos os argumentos, bem como as demonstrações de alguns resultados, podem ser encontradas em (JOHNSON, 1987), (ARANHA, 2008) e (MINGOTI, 2007).

Segundo Lattin (2011), a análise fatorial diz respeito à identificação de fontes subjacentes de variação comuns a duas ou mais variáveis (chamadas fatores comuns). Um pressuposto explícito nesse modelo de fator comum é que a variação observada em cada variável é atribuível aos fatores comuns subjacentes e à um fator específico (frequentemente interpretável como erro de medida).

A Análise Fatorial teve início, modernamente, no princípio do século XX com K. Pearson e C. Spearman, que estudaram as medidas de inteligência. A dificuldade nos cálculos impediu um desenvolvimento maior da técnica. O surgimento dos computadores altamente velozes trouxe novamente o interesse nos aspectos teóricos e computacionais da Análise Fatorial.

Suponha que variáveis possam ser agrupadas por suas correlações, isto é, todas as variáveis dentro de um grupo particular são altamente correlacionadas entre si, mas têm correlações relativamente baixas com variáveis de um grupo diferente. É admissível que cada grupo de variáveis represente um fator, que é responsável pelas correlações observadas.

O objetivo da análise fatorial é agrupar as informações contidas de um grande número de variáveis originais, em um conjunto menor de fatores com o mínimo de perda de informação. Em Corrar (2009) encontram-se descritos os objetivos da análise fatorial. São eles:

1. Harmonizar ou condensar um grande número de observações em grupos;
2. Obter o menor número de variáveis a partir do material original e reproduzir toda a informação de forma resumida;
3. Obter os fatores que reproduzam um padrão separado de relações entre as variáveis;
4. Interpretar de forma lógica o padrão de relações entre as variáveis;
5. Identificar as variáveis apropriadas para uma posterior análise de regressão e correlação ou análise discriminante.

A análise fatorial é uma técnica estatística que pode encontrar uma forma resumida das informações contida na matriz de dados, transformando as muitas variáveis originais em um conjunto menor de novas variáveis estatísticas (fatores) com perda mínima de informações. Mais especificamente, as técnicas de análise fatorial atendem um entre dois objetivos:

- Identificar uma estrutura por meio do resumo dos dados: ao analisar as correlações entre as variáveis, torna-se possível identificar as relações estruturais existente entre essas variáveis;
- Redução de dados: por meio da análise fatorial, é possível identificar as variáveis representativas de um conjunto maior criando um novo conjunto de variáveis, muito menor que o original, que poderá substituir sem muito prejuízo, o conjunto original de variáveis.

Nos dois casos, o propósito é manter a natureza e o caráter das variáveis originais, reduzindo seu número para simplificar a análise multivariada a ser aplicada posteriormente sem comprometer o resultado da análise.

2.4.1 Suposições da Análise Fatorial

Teste de Esfericidade de Bartlett

Segundo Lattin (2011) o teste de esfericidade, desenvolvido por Bartlett (1950), coloca diretamente a questão: a matriz de correlação deve ser decomposta em fatores, em primeiro lugar? O teste é um teste de qui-quadrado aproximado com um teste estatístico que é função do determinante da matriz de correlação R , como apresentado a seguir:

$$\chi^2 \left[\frac{(p^2 - p)}{2} \right] = - \left[(n - 1) - \left(\frac{2p + 5}{6} \right) \right] \ln |R|$$

onde:

$\ln |R|$ = log natural do determinante da matriz de correlação;

$\frac{(p^2 - p)}{2}$ = número de graus de liberdade associado ao teste estatístico de qui-quadrado;

p = número de variáveis;

n = número de observações.

A lógica do teste é a seguinte: o determinante da matriz de correlação é uma medida generalizada da variância. Ele pode ser calculado tomando-se o produto dos autovalores da matriz, isto é,

$$|R| = \prod_{j=1}^p \lambda_j$$

Quando as variáveis na análise são de fato mutuamente independentes, espera-se que R se aproxime da matriz identidade I (isto é, elementos diagonais igual a 1 e elementos fora da diagonal perto de zero em valor absoluto). Quando esse é o caso, um diagrama de dispersão dos dados apresenta formato esférico em vez de oval, como uma bola de futebol americano (daí o nome do teste). Todos os autovalores da matriz são próximos de 1,0 (indicando que o eixo mais

longo do elipsoide não é significativamente mais longo do que qualquer outro eixo); portanto, $|R|$ é próximo de 1,0 e $\ln|R|$ é próximo de 0. À medida que o montante de correlação entre as variáveis aumenta, o diagrama de dispersão dos dados começa a parecer mais elipsoidal. Nesse caso, alguns dos autovalores de R são maiores que 1 e alguns são mais próximos de zero. O resultado é que, à medida que o nível de correlação aumenta, o produto dos autovalores torna-se mais próximo de zero, o que significa que $\ln|R|$ torna-se um número negativo maior.

Crítério de Kaiser-Meyer-Olkin – KMO

Kaiser (1959) recomendou somente a retenção dos componentes principais com autovalores que excedam a unidade (isto é, supondo que trata-se de variáveis padronizadas na análise). A regra reflete uma noção de bom-senso: qualquer componente principal, por se tratar de uma medida da variância comum, deveria explicar pelo menos tantas variações quanto qualquer uma das variáveis originais X . A regra de Kaiser, portanto, busca um julgamento absoluto em relação ao valor de variância explicada por cada componente principal.

O critério de Kaiser-Meyer-Olkin – KMO é outra forma para identificar se o modelo de análise fatorial que está sendo utilizado está adequadamente ajustado aos dados, isto se dá testando a consistência geral dos dados. O método verifica se a matriz de correlação inversa é próxima da matriz diagonal e consiste em comparar os valores dos coeficientes de correlação linear observados com os valores dos coeficientes de correlação parcial. A medida de adequabilidade que fundamenta esse princípio é dada pela seguinte expressão:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2}$$

em que r_{ij} é o coeficiente de correlação simples entre as variáveis X_i e X_j , e a_{ij} é o coeficiente de correlação parcial entre X_i e X_j , dados os outros X 's.

Para interpretação do critério de KMO, os valores vão variar de 0 a 1, pois, pequenos valores de KMO indicam que o uso da análise fatorial não é adequada, e quanto mais próximo de 1, mais adequada é a aplicação da análise fatorial nos dados. Assim, existe a seguinte referência conforme tabela a seguir.

Crítério de Kaiser-Meyer-Olkin - KMO	
Valor	Grau de Adequação da Amostra
> 0,90	Ótima
de 0,80 a 0,90	Boa
de 0,70 a 0,80	Razoável
de 0,60 a 0,70	Baixa
< 0,60	Inadequada

Quadro 2.1: Valores de Interpretação do Crítério de Kaiser-Meyer-Olkin – KMO

2.4.2 O Modelo Fatorial Ortogonal

Seja a variável aleatória observável \underline{X} , com p componentes, $X \sim (\mu, \Sigma)$. O modelo fatorial postula que \underline{X} é linearmente dependente sobre algumas variáveis aleatórias não observáveis (latentes) F_1, F_2, \dots, F_m ($m < p$) chamadas fatores comuns (m fatores extraídos dos p existentes) e p fontes de variação aditivas $1, 2, \dots, p$, chamadas erros ou, algumas vezes, fatores específicos,

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2$$

$$\vdots$$

$$X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \dots + l_{im}F_m + \varepsilon_i$$

$$\vdots$$

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p$$

ou, em notação matricial:

$$\underline{X}_{p \times 1} - \underline{\mu}_{p \times 1} = \underline{L}_{p \times m} \underline{F}_{m \times 1} + \underline{\varepsilon}_{p \times 1}.$$

Os coeficientes l_{ij} (entradas da matriz L) são chamados de pesos ou carregamentos e, especificamente, l_{ij} é o carregamento na i -ésima variável do j -ésimo fator, tal que a matriz $l_{p \times m}$ é a matriz de carregamentos (pesos) dos fatores. Note que o fator específico ou erro i é associado somente com a i -ésima variável original X_i . Os desvios $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ são expressos em termos de $p + m$ variáveis aleatórias: $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ que não são observáveis. Isto distingue o modelo fatorial do modelo de regressão multivariada, cujas variáveis independentes (cujas posições são ocupadas por F) podem ser observadas.

Agora será assumido que:

$$E(\underline{F}) = \underline{0}_{m \times 1}, Cov(\underline{F}) = E(\underline{F}\underline{F}') = I_m$$

$$E(\underline{\varepsilon}) = \underline{0}_{p \times 1}, Cov(\underline{\varepsilon}) = E(\underline{\varepsilon}\underline{\varepsilon}') = \Psi_{p \times p} = \begin{pmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{pmatrix}$$

e que \underline{F} e $\underline{\varepsilon}$ são independentes, assim

$$Cov(\underline{\varepsilon}, \underline{F}) = E(\underline{\varepsilon}\underline{F}') = \underline{0}_{p \times p}$$

com $m = p$.

Com estas suposições o relacionamento construído em $\underline{X}_{p \times 1} - \underline{\mu}_{p \times 1} = \underline{L}_{p \times m} \underline{F}_{m \times 1} + \underline{\varepsilon}_{p \times 1}$ é chamado de Modelo Fatorial Ortogonal e pode ser escrito

$$\underline{X}_{p \times 1} = \underline{\mu}_{p \times 1} + \underline{L}_{p \times m} \underline{F}_{m \times 1} + \underline{\varepsilon}_{p \times 1}.$$

Matriz de Covariância do Vetor \underline{X}

Observe que:

$$\begin{aligned} (\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})' &= (\underline{L}\underline{F} + \underline{\varepsilon})(\underline{L}\underline{F} + \underline{\varepsilon})' = (\underline{L}\underline{F} + \underline{\varepsilon})\left((\underline{L}\underline{F})' + \underline{\varepsilon}'\right) = \\ & \underline{L}\underline{F}(\underline{L}\underline{F})' + \underline{\varepsilon}(\underline{L}\underline{F})' + \underline{L}\underline{F}\underline{\varepsilon}' + \underline{\varepsilon}\underline{\varepsilon}' = (\underline{L}\underline{F}\underline{F}'\underline{L}' + \underline{L}\underline{F}\underline{\varepsilon}' + \underline{\varepsilon}\underline{F}'\underline{L} + \underline{\varepsilon}\underline{\varepsilon}') \end{aligned}$$

Logo, a matriz de covariância \underline{X} pode ser escrita como:

$$\begin{aligned} \sum &= Cov(\underline{X}) = E\left[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'\right] = E[(\underline{L}\underline{F}\underline{F}'\underline{L}' + \underline{L}\underline{F}\underline{\varepsilon}' + \underline{\varepsilon}\underline{F}'\underline{L} + \underline{\varepsilon}\underline{\varepsilon}')] = \\ & E(\underline{L}\underline{F}\underline{F}'\underline{L}') + E(\underline{L}\underline{F}\underline{\varepsilon}') + E(\underline{\varepsilon}\underline{F}'\underline{L}) + E(\underline{\varepsilon}\underline{\varepsilon}') = \\ & \underline{L}E(\underline{F}\underline{F}')\underline{L}' + \underline{L}E(\underline{F}\underline{\varepsilon}') + E(\underline{\varepsilon}\underline{F}')\underline{L} + E(\underline{\varepsilon}\underline{\varepsilon}') = \\ & \underline{L}\underline{I}_m\underline{L}' + \underline{L}\underline{0}_{p \times 1} + \underline{0}\underline{L}' + \underline{\Psi}_{p \times p} = \underline{L}\underline{L}' + \underline{\Psi}. \end{aligned}$$

Portanto,

$$\sum = Cov(\underline{X}) = \underline{L}\underline{L}' + \underline{\Psi} \quad (2.1)$$

Consequentemente, tem-se:

$$V(X_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \Psi_i, i = 1, 2, \dots, p$$

Sabe-se que:

$$\sum = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{pmatrix}$$

então, por 2.1 tem-se que

$$\begin{aligned}
\Sigma &= \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \dots & l_{p1} \\ l_{12} & l_{22} & \dots & l_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1m} & l_{2m} & \dots & l_{pm} \end{bmatrix} + \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{bmatrix} = \\
&\begin{bmatrix} l_{11}l_{11} + \dots + l_{1m}l_{1m} & \dots & l_{11}l_{p1} + \dots + l_{1m}l_{pm} \\ l_{21}l_{11} + \dots + l_{2m}l_{1m} & \dots & l_{21}l_{p1} + \dots + l_{2m}l_{pm} \\ \vdots & \ddots & \vdots \\ l_{p1}l_{11} + \dots + l_{pm}l_{1m} & \dots & l_{p1}l_{p1} + \dots + l_{pm}l_{pm} \end{bmatrix} + \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{bmatrix} = \\
&\begin{bmatrix} l_{11}^2 + \dots + l_{1m}^2 & \dots & l_{11}l_{p1} + \dots + l_{1m}l_{pm} \\ l_{21}l_{11} + \dots + l_{2m}l_{1m} & \dots & l_{21}l_{p1} + \dots + l_{2m}l_{pm} \\ \vdots & \ddots & \vdots \\ l_{p1}l_{11} + \dots + l_{pm}l_{1m} & \dots & l_{p1}^2 + \dots + l_{pm}^2 \end{bmatrix} + \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{bmatrix} = \\
&\begin{bmatrix} l_{11}^2 + \dots + l_{1m}^2 + \Psi_1 & \dots & l_{11}l_{p1} + \dots + l_{1m}l_{pm} \\ l_{21}l_{11} + \dots + l_{2m}l_{1m} & \dots & l_{21}l_{p1} + \dots + l_{2m}l_{pm} \\ \vdots & \ddots & \vdots \\ l_{p1}l_{11} + \dots + l_{pm}l_{1m} & \dots & l_{p1}^2 + \dots + l_{pm}^2 + \Psi_p \end{bmatrix}
\end{aligned}$$

Portanto,

$$V(X_i) = S_{ii} = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \Psi_i, i = 1, 2, \dots, p.$$

Assim, a matriz de covariância pode ser decomposta em duas partes (matrizes) LL' e Ψ . A matriz Ψ é chamada de matriz de variâncias específicas, que é uma matriz diagonal possuindo na diagonal principal as "variâncias específicas" Ψ_i das variáveis originais. Já a matriz produto LL' tem na diagonal principal as comunalidades

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2, i = 1, 2, \dots, p (j = 1, 2, \dots, m).$$

Com base nos resultados apresentados anteriormente, a covariância entre o vetor das variáveis originais \underline{X} e o vetor dos fatores \underline{F} é:

1. $Cov(\underline{X}, \underline{F}) = L$;
2. $Cov(X_i, X_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \dots + l_{im}l_{km}$;
3. $Cov(X_i, F_j) = l_{ij}$.

Demonstração 2.41:

1. $Cov(\underline{X}, \underline{F}) = L$.

Pela definição de covariância, tem-se:

$$Cov(\underline{X}, \underline{F}) = E \left[(\underline{X} - \underline{\mu}) (\underline{F} - 0)' \right] = E \left[(\underline{X} - \underline{\mu}) \underline{F}' \right]$$

Pela suposição do modelo fatorial, temos: $\underline{X} - \underline{\mu} = \underline{L}\underline{F} + \underline{\varepsilon}$, $E(\underline{F}\underline{F}') = I$ e $E(\underline{\varepsilon}\underline{F}') = 0$. Assim,

$$Cov(\underline{X}, \underline{F}) = E[\underline{L}\underline{F}\underline{F}' + \underline{\varepsilon}\underline{F}'] = \underline{L}E(\underline{F}\underline{F}') + E(\underline{\varepsilon}\underline{F}') = \underline{L}.I + 0 = \underline{L}$$

Portanto,

$$Cov(\underline{X}, \underline{F}) = \underline{L}.$$

$$2. Cov(X_i, X_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \dots + l_{im}l_{km};$$

Como $Cov(X_i, X_j) = S_{ij}$ e pelo produto $\underline{L}\underline{L}' + \Psi$, foi visto que na posição (ij) , tem-se a expressão $l_{i1}l_{j1} + l_{i2}l_{j2} + \dots + l_{im}l_{jm}$. Então,

$$Cov(X_i, X_k) = l_{i1}l_{k1} + l_{i2}l_{k2} + \dots + l_{im}l_{km}.$$

$$3. Cov(X_i, F_j) = l_{ij}.$$

Por definição, tem-se que

$$Cov(\underline{X}, \underline{F}) = \begin{bmatrix} Cov(X_1, F_1) & Cov(X_1, F_2) & \dots & Cov(X_1, F_m) \\ Cov(X_2, F_1) & Cov(X_2, F_2) & \dots & Cov(X_2, F_m) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_i, F_1) & Cov(X_i, F_2) & \ddots & Cov(X_i, F_m) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, F_1) & Cov(X_p, F_2) & \dots & Cov(X_p, F_m) \end{bmatrix}$$

No item 1, foi mostrado que

$$Cov(\underline{X}, \underline{F}) = \underline{L} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{i1} & l_{i2} & \dots & l_{im} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix}$$

Portanto, por igualdade de matriz, segue que

$$Cov(X_i, F_j) = l_{ij}.$$

■

Comunalidades e Variâncias Específicas

A porção da variância da i -ésima variável aleatória X_i advinda como contribuição dos m fatores comuns (extraídos) é chamada de comunalidade e a porção da $V(X_i) = \sigma_{ii} = \sigma_i^2$ oriunda do fator específico é a variância específica. Assim, tem-se:

Pela definição do modelo fatorial ortogonal, tem-se que:

$$\underline{X} = \underline{\mu} + \underline{L} \cdot \underline{F} + \underline{\varepsilon}$$

ou seja,

$$\underline{X}_j = \underline{\mu}_i + \underline{L}_i \underline{F}_j + \underline{\varepsilon}_j$$

então,

$$V(X_i) = \underline{\mu}_i + \underline{L}_i \underline{F} + \underline{\varepsilon}_j$$

onde

$$\underline{L}_i \underline{F} = l_{i1} F_1 + l_{i2} F_2 + \dots + l_{im} F_m$$

Usando a Propriedade da Variância 3 item (iii), tem-se:

$$V(X_i) = V(\mu_i) + V(l_{i1} F_1 + l_{i2} F_2 + \dots + l_{im} F_m) + V(\underline{\varepsilon}_j)$$

$$V(X_i) = V(\mu_i) + V(l_{i1} F_1) + V(l_{i2} F_2) + \dots + V(l_{im} F_m) + V(\underline{\varepsilon}_j)$$

Como $l_{i1}, l_{i2}, \dots, l_{im}$ são os carregamentos dos fatores, ou seja, são valores constantes, podemos utilizar a propriedade 1 da variância, onde $V(cX) = c^2 V(X)$, então:

$$V(\mu_i) + l_{i1}^2 V(F_1) + l_{i2}^2 V(F_2) + \dots + l_{im}^2 V(F_m) + V(\underline{\varepsilon}_j) =$$

$$0 + l_{i1}^2 \cdot 1 + l_{i2}^2 \cdot 1 + \dots + l_{im}^2 \cdot 1 + \Psi_i$$

Portanto,

$$V(X_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \Psi_i.$$

$V(X_i) = h_i^2 + \Psi_i$, onde $i = 1, 2, \dots, p$ e $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$ sendo a soma de quadrados dos carregamentos na i -ésima variável dos m fatores comuns (extraídos).

Estimação

Dadas as observações x_1, x_2, \dots, x_n de p variáveis geralmente correlacionadas, a Análise Fatorial procura responder a pergunta:

"Representará o modelo fatorial os dados adequadamente, com um número $m < p$ (baixo) de fatores?"

A matriz de covariância amostral S é um estimador da matriz de covariâncias populacional desconhecida. Se os elementos fora da diagonal de S são baixos, ou

equivalentemente na matriz de correlação amostral R eles são praticamente nulos, as variáveis não são relacionadas e a Análise Fatorial não é útil. Por outro lado quando é significativamente diferente de uma matriz diagonal, então o modelo fatorial pode ser usado e o problema inicial é o de estimar os carregamentos (pesos) l_{ij} e as variâncias específicas Ψ_i . Vamos considerar no nosso estudo a estimação pelo Método das Componentes Principais. Seja Σ a matriz de covariâncias de X , então, dado que Σ seja positiva definida, podemos decompô-la na forma abaixo, segundo a decomposição espectral:

$$\Sigma = \lambda_1 \underline{e}_1 \underline{e}'_1 + \lambda_2 \underline{e}_2 \underline{e}'_2 + \dots + \lambda_p \underline{e}_p \underline{e}'_p$$

onde

$\lambda_1, \lambda_2, \dots, \lambda_p$ são os autovalores da matriz de covariância Σ .

$\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$ são os respectivos autovetores (dos autovalores) da matriz de covariância.

Como Σ (matriz de covariância) é positiva definida, pode-se utilizar o Teorema da Decomposição Espectral 2.26:

$$\Sigma = P\Lambda P'$$

onde:

$P =$ é a matriz gerada pelos autovetores da matriz de covariância.

$\lambda =$ é a matriz diagonal formada pelos autovalores da matriz de covariância.

Então:

$$\Sigma = P\Lambda P'$$

$$\Sigma = \begin{bmatrix} e_{11} & e_{21} & \dots & e_{p1} \\ e_{12} & e_{22} & \dots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \dots & e_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \dots & e_{pp} \end{bmatrix} =$$

$$\begin{bmatrix} e_{11} & e_{21} & \dots & e_{p1} \\ e_{12} & e_{22} & \dots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \dots & e_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1 e_{11} & \lambda_1 e_{12} & \dots & \lambda_1 e_{1p} \\ \lambda_2 e_{21} & \lambda_2 e_{22} & \dots & \lambda_2 e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_p e_{p1} & \lambda_p e_{p2} & \dots & \lambda_p e_{pp} \end{bmatrix} =$$

$$\begin{bmatrix} e_{11}\lambda_1 e_{11} + e_{21}\lambda_2 e_{21} + \dots + e_{p1}\lambda_p e_{p1} & \dots & e_{11}\lambda_1 e_{1p} + e_{21}\lambda_2 e_{2p} + \dots + e_{p1}\lambda_p e_{pp} \\ e_{12}\lambda_1 e_{11} + e_{22}\lambda_2 e_{21} + \dots + e_{p2}\lambda_p e_{p1} & \dots & e_{12}\lambda_1 e_{1p} + e_{22}\lambda_2 e_{2p} + \dots + e_{p2}\lambda_p e_{pp} \\ \vdots & \ddots & \vdots \\ e_{1p}\lambda_1 e_{11} + e_{2p}\lambda_2 e_{21} + \dots + e_{pp}\lambda_p e_{p1} & \dots & e_{1p}\lambda_1 e_{1p} + e_{2p}\lambda_2 e_{2p} + \dots + e_{pp}\lambda_p e_{pp} \end{bmatrix} =$$

$$\begin{aligned}
& \begin{bmatrix} e_{11}\lambda_1 e_{11} & \dots & e_{11}\lambda_1 e_{1p} \\ e_{12}\lambda_1 e_{11} & \dots & e_{12}\lambda_1 e_{1p} \\ \vdots & \ddots & \vdots \\ e_{1p}\lambda_1 e_{11} & \dots & e_{1p}\lambda_1 e_{1p} \end{bmatrix} + \begin{bmatrix} e_{21}\lambda_2 e_{21} & \dots & e_{21}\lambda_2 e_{2p} \\ e_{22}\lambda_2 e_{21} & \dots & e_{22}\lambda_2 e_{2p} \\ \vdots & \ddots & \vdots \\ e_{2p}\lambda_2 e_{21} & \dots & e_{2p}\lambda_2 e_{2p} \end{bmatrix} + \dots + \begin{bmatrix} e_{p1}\lambda_p e_{p1} & \dots & e_{p1}\lambda_p e_{pp} \\ e_{p2}\lambda_p e_{p1} & \dots & e_{p2}\lambda_p e_{pp} \\ \vdots & \ddots & \vdots \\ e_{pp}\lambda_p e_{p1} & \dots & e_{pp}\lambda_p e_{pp} \end{bmatrix} = \\
& \lambda_1 \begin{bmatrix} e_{11}e_{11} & \dots & e_{11}e_{1p} \\ e_{12}e_{11} & \dots & e_{12}e_{1p} \\ \vdots & \ddots & \vdots \\ e_{1p}e_{11} & \dots & e_{1p}e_{1p} \end{bmatrix} + \lambda_2 \begin{bmatrix} e_{21}e_{21} & \dots & e_{21}e_{2p} \\ e_{22}e_{21} & \dots & e_{22}e_{2p} \\ \vdots & \ddots & \vdots \\ e_{2p}e_{21} & \dots & e_{2p}e_{2p} \end{bmatrix} + \dots + \lambda_p \begin{bmatrix} e_{p1}e_{p1} & \dots & e_{p1}e_{pp} \\ e_{p2}e_{p1} & \dots & e_{p2}e_{pp} \\ \vdots & \ddots & \vdots \\ e_{pp}e_{p1} & \dots & e_{pp}e_{pp} \end{bmatrix} = \\
& \lambda_1 \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1p} \end{bmatrix} [e_{11} \ e_{12} \ \dots \ e_{1p}] + \lambda_2 \begin{bmatrix} e_{21} \\ e_{22} \\ \vdots \\ e_{2p} \end{bmatrix} [e_{21} \ e_{22} \ \dots \ e_{2p}] + \dots + \\
& \lambda_p \begin{bmatrix} e_{p1} \\ e_{p2} \\ \vdots \\ e_{pp} \end{bmatrix} [e_{p1} \ e_{p2} \ \dots \ e_{pp}].
\end{aligned}$$

$$\therefore \Sigma = \lambda_1 \underline{e}_1 \underline{e}'_1 + \lambda_2 \underline{e}_2 \underline{e}'_2 + \dots + \lambda_p \underline{e}_p \underline{e}'_p.$$

Assim, se $\Sigma = L.L' + \Psi$, tem-se que $\Psi_{p \times p}$ no ajuste do modelo fatorial. Exceto pelo escalar $\sqrt{\lambda_j}$, os carregamentos no j -ésimo fator são os coeficientes populacionais na j -ésima componente principal. Embora a representação de $\Sigma = L.L' + 0 = LL'$ seja exata, ela não é particularmente útil, pois tem muitos fatores comuns. É preferível um modelo que explique a estrutura de covariância em termos de poucos fatores comuns. Uma aproximação, quando $p-m$ autovalores são baixos, é negligenciar a contribuição de $\lambda_{m+1} \underline{e}_{m+1} \underline{e}'_{m+1} + \lambda_{m+2} \underline{e}_{m+2} \underline{e}'_{m+2} + \dots + \lambda_p \underline{e}_p \underline{e}'_p$ para Σ na decomposição espectral.

Assim, tem-se:

$$\Sigma \approx [\sqrt{\lambda_1} \underline{e}_1 \quad \sqrt{\lambda_2} \underline{e}_2 \quad \dots \quad \sqrt{\lambda_m} \underline{e}_m] \begin{bmatrix} \sqrt{\lambda_1} \underline{e}'_1 \\ \sqrt{\lambda_2} \underline{e}'_2 \\ \vdots \\ \sqrt{\lambda_m} \underline{e}'_m \end{bmatrix} = LL', \text{ de ordem } p \times p.$$

Esta representação aproximada assume que os fatores específicos \underline{e} são de menor importância e podem, também, ser ignorados na fatorização de Σ . Se os fatores específicos \underline{e} são incluídos no modelo, suas variâncias são os elementos da diagonal da matriz diferença $\Sigma - LL'$ e conseqüentemente,

$$\Psi_i = \sigma_{ii} - \sum_{j=1}^m l_{ij}^2$$

para $i = 1, 2, \dots, p$.

Para aplicar esta abordagem aos dados amostrais $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ é usual, primeiro, centrar as observações subtraindo a média amostral \bar{x} . As observações centradas são:

$$[x_j - \bar{x}] = \begin{bmatrix} x_{1j} - \bar{x}_1 \\ x_{2j} - \bar{x}_2 \\ \vdots \\ x_{pj} - \bar{x}_p \end{bmatrix}, j = 1, 2, \dots, n.$$

Pode-se também, trabalhar com as variáveis padronizadas

$$\underline{z}_j = \begin{bmatrix} \frac{x_{1j} - \bar{x}_1}{\sqrt{S_{11}}} \\ \frac{x_{2j} - \bar{x}_2}{\sqrt{S_{22}}} \\ \vdots \\ \frac{x_{pj} - \bar{x}_p}{\sqrt{S_{pp}}} \end{bmatrix}, j = 1, 2, \dots, n$$

cujas matriz de correlação amostral é a matriz de correlação R das observações originais $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$. A representação $\Sigma \approx L.L' + \Psi$, quando se usa a matriz de covariância S ou, então, a matriz de correlação R , é conhecida como Solução por Componentes Principais.

2.5 Confiabilidade Estrutural

A Análise de Confiabilidade Estrutural baseia-se na relação entre as cargas impostas à um sistema e sua capacidade de suportá-las. Ambas as cargas e a resistência podem ser incertas, de modo que o resultado de sua interação também seja incerto. Atualmente, é comum expressar a condição de uma estrutura na forma de um índice de confiabilidade, podendo ser relacionado com a probabilidade de fracasso (BAECHER, 2003). Deve-se entender neste contexto que "fracasso" inclui não apenas falha catastrófica, como no caso de um deslizamento de terra, mas também, como qualquer diferença inaceitável entre o desempenho esperado e o observado (LEONARDS, 1975).

Com base na formulação convencional da confiabilidade estrutural, para a obtenção da probabilidade de falha de uma estrutura é essencial definir o vetor de variáveis aleatórias \underline{X} , dado por:

$$\underline{X} = (X_1, X_2, \dots, X_n)^t$$

que corresponde às incertezas associadas ao projeto, como por exemplo, à solicitação imposta à estrutura, à resistência, à geometria e aos materiais.

A função desempenho, apresentada a seguir,

$$H(\underline{X}) = H(X_1, X_2, \dots, X_n) \quad (2.3)$$

estabelece um limite entre o domínio de falha e o domínio seguro, ou seja,

$$D_f = \{\underline{X}; H(\underline{X}) \leq 0\}$$

é o domínio de falha, e

$$D_s = \{\underline{X}; H(\underline{X}) > 0\}$$

é o domínio de segurança.

A equação de estado limite é definida por $H(\underline{X}) = 0$. Consequentemente, a probabilidade de falha pode ser avaliada por:

$$P_f = \int \dots \int_{H(\underline{X}) \leq 0} f_X(X_1, X_2, \dots, X_n) dx_1 dx_2 \dots dx_n$$

onde $f_X(X_1, X_2, \dots, X_n)$ é a função densidade de probabilidade conjunta para os vetores de variáveis aleatórias e a integração é realizada sobre o domínio de falha, $H(\underline{X}) \leq 0$.

Para Santos (2012) o cálculo da probabilidade de falha por meio da avaliação da integral múltipla não é fácil, devido aos seguintes motivos:

1. Envolve uma integral multi-dimensional;
2. A forma exata da função densidade de probabilidade conjunta das variáveis aleatórias raramente é conhecida;
3. A equação de estado limite $H(\underline{X}) = 0$ nem sempre é dada de forma analítica, mas como a solução de algum algoritmo numérico.

Soluções aproximadas podem ser obtidas eficientemente usando os Métodos de Confiabilidade de Primeira Ordem (*First Order Reliability Method - FORM*), ou de Segunda Ordem (*Second Order Reliability Method - SORM*). Os conceitos desses métodos são encontrados em (SANTOS, 2012).

Para a realização desse trabalho, é utilizada uma função de desempenho multivariada, no qual as variáveis aleatórias envolvidas sejam apenas as obtidas através das leituras dos instrumentos. Segundo (SILVA, 2016) quando o método estatístico multivariado é utilizado para determinar a função de desempenho multivariada, a probabilidade de falha não exige a avaliação de uma integral múltipla e nem a utilização de técnicas de programação linear, para esse objetivo é necessário apenas que os escores fatoriais possuam uma distribuição normal de probabilidade. Por esse motivo, no próximo capítulo é detalhado tal método estatístico para que, posteriormente, o mesmo possa ser utilizado na definição de uma função de desempenho que satisfaça as exigências da análise de confiabilidade estrutural.

2.6 Estudo da Arte

A área de Sistemas Dinâmicos estuda processos cuja evolução é dada por uma lei matemática, como os que encontramos na Física, na Química, na Biologia, na Economia,

Meteorologia e em quase todos os ramos do conhecimento. A lei de evolução pode ser uma função ou uma equação diferencial. O objetivo é construir uma teoria matemática destes processos que permita compreender e prever sua evolução. São usados métodos das mais diversas áreas da Matemática e métodos computacionais (GERAIS, 2017).

Conforme Belato (2002) sistemas dinâmicos são encontrados em diversas áreas de aplicação, como por exemplo em estudos populacionais, econômicos, na biologia, na eletricidade, na mecânica e muitos outros; estudar sobre essas aplicações significa conseguir estimar a variação temporal das variáveis escolhidas para a sua representação a partir das perturbações existentes.

Segundo Pinto (2007) Jay W. Forrester, foi o responsável pela formação do *System Dynamics Group no MIT (Massachusetts Institute of Technology - Instituto de Tecnologia de Massachusetts)*, que iniciou, na década de 1960, os estudos voltados ao comportamento de sistemas dinâmicos complexos, através do uso de fluxos de retroalimentação.

No trabalho desenvolvido por Mello (2017), foram utilizados os conceitos de sistemas dinâmicos para realizar um estudo no setor de embalagens de uma empresa produtora de ervamate, localizada no norte do estado do Rio Grande do Sul, onde o objetivo foi aplicar o método de cronoanálise para o processo de embalagem, no qual fosse possível definir um tempo padrão para cada atividade operacional, visando uma redistribuição de atividades entre os operadores, bem como a eliminação de atividades que não agregam valor ao produto.

Outro trabalho, na área de Sistemas Dinâmicos, foi apresentado por Zanardi (2016), no qual foi realizado a modelagem matemática do processo de desidratação da soja após a colheita para diminuir o risco de deterioração e aumentar o tempo de armazenagem. Essa modelagem foi importante para realizar a simulação de novas condições operacionais. Com o objetivo de avaliar o perfil da umidade com o tempo de secagem, os autores propuseram duas estratégias numéricas para a solução da segunda lei de Fick com coeficiente de difusão variável. A primeira estratégia consiste em derivar a equação na direção radial empregando a regra do produto e então discretizá-la nesta direção por Diferenças Finitas. Na segunda estratégia, a equação original é discretizada diretamente na direção radial, sendo que para isto são empregadas interpolações lineares para o cômputo das propriedades entre dois pontos da malha.

O estudo apresentado por Oliveira (2015) teve como objetivo propôr uma metodologia para otimizar a distribuição de alunos em escolas públicas. Como estudo de caso, foram utilizados dados, obtidos via ferramentas computacionais, referentes às escolas e estudantes do ensino médio da rede estadual de educação, residentes no distrito de Neves, no município de São Gonçalo-RJ, tendo como referência o ano de 2014. Foi proposto neste trabalho um modelo de Programação Linear composto de duas fases: na primeira fase, o modelo visa maximizar o número de alunos que possam ir a pé até as suas escolas; na segunda fase, mantendo o número ótimo obtido na primeira fase de estudantes que podem ir a pé, o modelo redistribui os alunos minimizando a distância total percorrida pelos que não podem ir a pé. Os resultados apontam a possibilidade de mais de 90% dos alunos envolvidos no estudo de caso, serem alocados em escolas consideradas próximas às suas residências.

No trabalho desenvolvido por Nieradka (2016) é apresentada, de forma detalhada, a construção do algoritmo que permite o monitoramento de barragens usando o método estatístico multivariado, que tem como base a Análise Multivariada, Mineração de Dados e Estatística, sendo possível, identificar possíveis anomalias que possam estar ocorrendo ou na eminência de ocorrer. Essas anomalias podem ser, por exemplo, o Deslizamento ou Tombamento de um dos blocos que compõe a estrutura da barragem. O monitoramento consiste em conhecer os valores máximos e mínimos de leituras de cada instrumento dentro de uma faixa denominada Bom Comportamento e, com isso, usar uma função de desempenho que possa determinar se uma leitura indica um Bom Comportamento ou uma Mudança de Comportamento.

Krüger (2008) realizou um trabalho cujo objetivo principal foi desenvolver uma metodologia para a análise de confiabilidade estrutural de barragens de concreto, e compará-la com os procedimentos determinísticos normalmente utilizados nas análises de estabilidade, bem como elaborar programas computacionais genéricos em linguagem *Fortran*, para o cálculo da probabilidade de falha, ponto de projeto, etc., em situações semelhantes de barragens de concreto à gravidade.

Neste mesmo trabalho, foram desenvolvidas equações de estado limite para os vários modos de falha analisados em barragens de concreto à gravidade: flutuação, tombamento, deslizamento e tensões normais. Foram implementados algoritmos computacionais para aplicação dos métodos de confiabilidade de primeira ordem FOSM e AFOSM e para o método de segunda ordem SORM. Para fins de comparação de resultados, também foram realizadas simulações pelo método Monte Carlo. O procedimento desenvolvido foi testado e validado a partir de dados do concreto compactado com rolo (CCR) obtidos durante a construção e de levantamentos recentes, na barragem de Salto Caxias, situada no Rio Iguaçu, Estado do Paraná.

Com isso, percebe-se que as ideias e o método de sistemas dinâmicos são aplicáveis para sistemas naturais, humanos e técnicos, combinando a teoria e a simulação computacional com a aplicação prática em problemas reais.

Capítulo 3

Modelo Matemático Multivariado

O modelo estatístico multivariado que é apresentado, utiliza os resultados da análise fatorial aplicado em uma matriz multivariada, onde as variáveis aleatórias são os dados gerados pelas leituras dos instrumentos instalados em um bloco-chave da Barragem de Itaipu.

No capítulo 2, a análise fatorial foi apresentada de forma detalhada, mostrando suas propriedades, tais como o agrupamento em fatores das variáveis altamente correlacionadas, que resulta na diminuição da dimensão das variáveis. Além disso, mostrou-se que com esses fatores, é possível gerar os escores fatoriais que são as variáveis aleatórias que substituirão as variáveis originais com pouca perda de informação, que ainda são não correlacionadas entre si, o que permite a aplicação de diversas técnicas estatísticas em que essa hipótese é necessária.

Para o desenvolvimento desse método estatístico multivariado são definidos alguns objetos intrínsecos desse método, apresentados a seguir:

1. Falha na estrutura: considera-se uma falha na estrutura, uma mudança brusca no comportamento das leituras dos instrumentos que são capazes de monitorar uma determinada anomalia;
2. Vetor de Simulação: são vetores com as leituras dos instrumentos instalados no bloco em estudo, que são gerados dentro de intervalos específicos através de um algoritmo de geração de números aleatórios, que ocasionam uma mudança brusca nessas leituras, diferente do que é normalmente registrado para esses instrumentos, com o objetivo de simular uma anomalia na estrutura;
3. Transformações nos dados: é a aplicação de uma função nos escores fatoriais que possuem uma distribuição de probabilidade qualquer, com o objetivo de torná-los variáveis aleatórias com distribuição de probabilidade normal. Tal transformação nem sempre é possível de se determinar.
4. Mudança de comportamento: é um comportamento atípico nas leituras dos instrumentos de acordo com o proprietário da Barragem e, que merece atenção pelos profissionais caso esse comportamento se mantenha;
5. Regiões Críticas: é o intervalo onde os escores fatoriais se agrupam quando ocorre uma mudança de comportamento em determinados instrumentos, que são encontrados através de simulações das leituras relacionados com uma determinada anomalia;

6. Determinação dos modos de falha: é um algoritmo implementado no Matlab que permite determinar as regiões críticas para cada anomalia possível de ser determinada, de acordo com os resultados apresentados pela AF que mostra os agrupamentos dos instrumentos que são altamente correlacionados.

As etapas apresentadas no fluxograma abaixo têm por finalidade a definição de uma função de desempenho multivariada, onde as variáveis envolvidas sejam apenas aquelas obtidas através das leituras dos instrumentos instalados no bloco de concreto em questão.

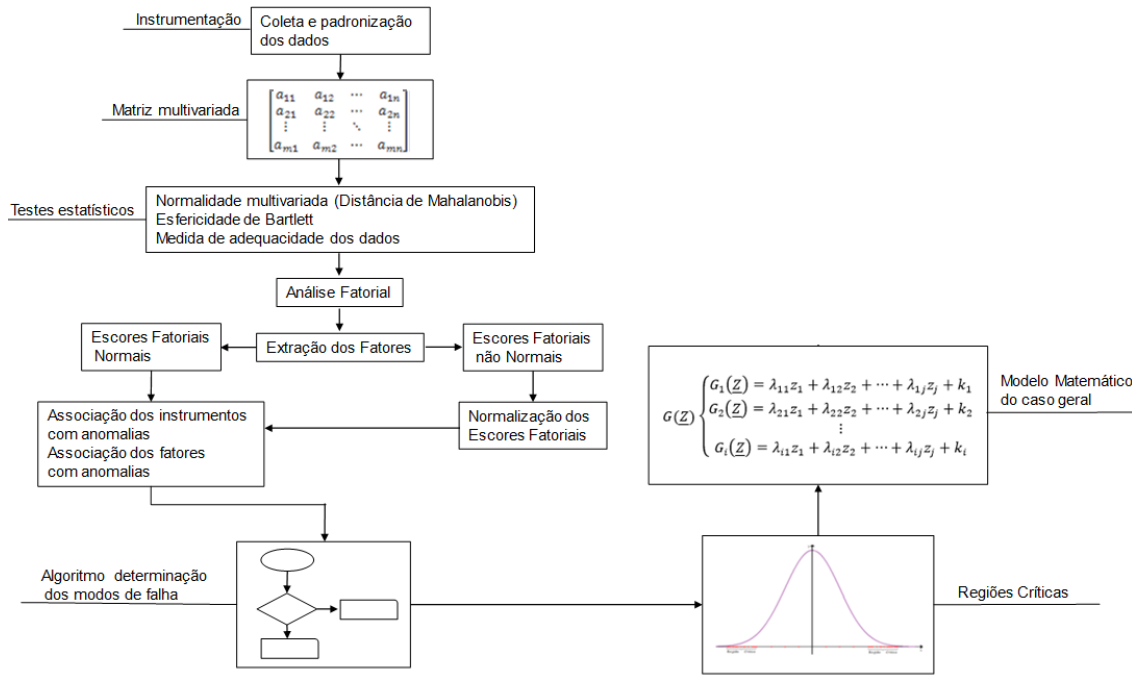


Figura 3.1: Fluxograma do Modelo.

O modelo descrito na figura 3.1 é proposto para realizar o monitoramento de um bloco de concreto altamente instrumentado, no qual as leituras dos instrumentos são variáveis aleatórias correlacionadas entre si e capazes de monitorar anomalias na estrutura.

Nas próximas seções, os objetos definidos anteriormente são estabelecidos e utilizados de forma ordenada para o desenvolvimento do algoritmo que possibilita a determinação dos modos de falha e da função de desempenho multivariada para o caso geral e aplicado.

3.1 Caracterização do Modelo Matemático para o Caso Geral

Como a análise fatorial agrupa os instrumentos altamente correlacionados à um mesmo fator, a ideia é utilizar as associações dos instrumentos com as anomalias apresentadas em (FILHO, 2006), (SILVEIRA, 2003) e (ELETROBRÁS, 2003) e o conhecimento prático dos profissionais em segurança de barragem da UHE de Itaipu para relacionar os fatores com as anomalias.

Se mais de um fator é necessário para monitorar certa anomalia, têm-se o caso de modos de falha de sistemas, onde a ocorrência de mais de um evento é necessária para ocasionar uma falha.

Como não são utilizadas as grandezas físicas, deve-se verificar o comportamento anormal da estrutura através das leituras dos instrumentos. De acordo com as literaturas citadas e do que foi repassado pelos profissionais em segurança de barragem da UHE de Itaipu, para que certas anomalias ocorram, é necessário que exista uma combinação de leituras fora do comportamento normal, simultaneamente. Assim, deve-se determinar em quais fatores esses instrumentos são agrupados.

Como os escores fatoriais dependem dos fatores obtidos na análise fatorial e, conseqüentemente, dependem dos instrumentos agrupados neles, quando houver alguma mudança brusca nas leituras desses instrumentos correlacionados com uma determinada anomalia, o resultado do escore fatorial também deverá sofrer alteração. Portanto, para realizar esse monitoramento busca-se determinar as regiões críticas desses escores fatoriais a qual indicará quais as anomalias ocorreram.

Para obter as regiões críticas de uma certa anomalia, antes é necessário definir os intervalos de mudança de comportamento para os instrumentos capazes de monitorá-la. Para determinar o intervalo de mudança de comportamento de um certo instrumento k são utilizados os limites de especificação de projeto e os intervalos de 99% de confiança da média dos últimos cinco anos desse mesmo instrumento, obtendo-se:

$$MC_k = (\min\{l_i, L_i\} - k_1s, \max\{l_i, L_i\} - k_1s) \cup (\min\{l_s, L_s\} + k_2s, \max\{l_s, L_s\} + k_2s) \quad (3.1)$$

Em (3.1), l_i e l_s são os limites inferiores e superiores do intervalo de confiança de 99% da média, respectivamente, L_i e L_s são os limites inferior e superior do intervalo de projeto, s é o desvio padrão amostral das leituras do instrumento k correspondente ao mesmo período de tempo do intervalo de confiança da média e, $\{k_1, k_2\}$ são constantes, maiores ou iguais à zero, que garantem que o intervalo de mudança de comportamento possa gerar valores extremos já obtidos pelas leituras do instrumento k , mesmo que esses valores sejam raros de ocorrer.

Observe que o intervalo MC_k assim definido, apresenta valores entre os intervalos de projeto e o intervalo de confiança da média e, ele não sugere valores impossíveis de serem alcançados, mas que não são normalmente registrados. Por esse motivo, recebe o nome de intervalo de mudança de comportamento. Valores obtidos dentro desse intervalo não apontam um problema eminente, já que podem estar dentro do intervalo de projeto ou de confiança, mas indicam uma mudança no seu comportamento e pode servir como um aviso para um monitoramento mais detalhado.

Pode-se então definir o intervalo de bom comportamento do instrumento k , para os casos onde os instrumentos apresentam intervalos de projeto, como mostra a equação a seguir:

$$BC_k = (\max\{l_i, L_i\} - k_1s, \min\{l_s, L_s\} + k_2s) \quad (3.2)$$

Em (3.2) l_i e l_s são os limites inferior e superior do intervalo de confiança de 99% da média, respectivamente, L_i e L_s são os limites inferior e superior do intervalo de projeto, respectivamente, s é o desvio padrão amostral das leituras do instrumento k correspondente ao mesmo período de tempo do intervalo de confiança da média, e, $\{k_1, k_2\}$ são as mesmas constantes que formaram o intervalo de mudança de comportamento do instrumento k correspondente.

De acordo com Silva (2016) não existe uma regra geral para a determinação dos valores de $\{k_1, k_2\}$, pois os mesmos dependem da característica de cada instrumento, de sua sazonalidade e principalmente, do nível de sensibilidade que o proprietário da barragem desejar. Por exemplo, valores próximos de zero para essas constantes, fornecem uma sensibilidade maior, pois incluem valores já ocorridos nos últimos anos como uma mudança no comportamento. Já valores maiores podem não acusar um comportamento já ocorrido como anormal, mesmo que essa leitura tenha sido muito rara e mereça atenção.

Usando um programa gerador de números aleatórios, já implementado no software *Matlab*, é possível gerar vetores de simulação das anomalias, que são possíveis de monitorar a partir do que foi apresentado pela análise fatorial. Por exemplo, se quatro instrumentos estão relacionados com certa anomalia, basta gerar leituras para esses instrumentos em seus respectivos intervalos *MC*, e para os demais instrumentos gerar leituras dentro dos intervalos *BC* obtidos através do intervalo de confiança de 99% da média, fazendo isso, tem-se a geração de um vetor que simula essa anomalia.

Ainda, segundo Silva (2016), a geração por mais que seja aleatória, deve seguir o seguinte princípio: se existem quatro instrumentos altamente relacionados com essa anomalia, então, antes é necessário verificar qual é o instrumento com maior influência no fator, essa informação é revelada pela análise fatorial, observando as cargas fatoriais de cada instrumento nos fatores. Além disso, a análise fatorial também revela qual é a correlação entre todos os instrumentos nesse estudo. Reordenando os instrumentos, de forma que todos aqueles relacionados com a anomalia fiquem nos primeiros lugares, e o primeiro instrumento deve ser o que possui maior influência no fator. Dessa forma, quando for gerado o primeiro número aleatório para o primeiro instrumento, é verificado em qual dos conjuntos do intervalo *MC* ele pertence, visto que o intervalo *MC* é formado pela união de dois conjuntos. Se por exemplo, for gerado um número do intervalo esquerdo do conjunto *MC*, quando for gerado o próximo número aleatório para o próximo instrumento, deve ser verificado qual é a correlação do segundo instrumento com o primeiro. Se a correlação for positiva, este número aleatório deve estar do mesmo lado esquerdo do intervalo *MC* do segundo instrumento; caso contrário, ou seja, se a correlação for negativa, o número aleatório deve estar do lado direito do intervalo *MC* do segundo instrumento.

Seguindo esse princípio para todos os instrumentos, mesmo aqueles que não têm relação com a anomalia, contudo, neste caso, deve ser verificado o lado esquerdo e/ou direito do intervalo de bom comportamento *BC* de tais instrumentos, visto que, todos intervalos contém a média das leituras e podem ser escritos como a união de dois conjuntos, ou seja, $BC = (b_i, \bar{x}) \cup (\bar{x}, b_s)$, onde b_i e b_s são os limites inferior e superior do respectivo intervalo de bom

comportamento e, x é a média amostral das leituras coletadas nos anos 2000 a 2016 dos respectivos instrumentos.

Esse processo tem a finalidade de respeitar o comportamento da estrutura, onde os instrumentos são altamente correlacionados. Dessa forma, é garantido um vetor de simulação condizente com a realidade. O algoritmo de geração do vetor de simulação e obtenção da região crítica para uma determinada anomalia é descrito a seguir.

Considere certa anomalia i , e suponha que existam m instrumentos correlacionados com a anomalia i . Considere também que desses m instrumentos, um número $m' < m$ esteja altamente correlacionado com um fator j .

Passos:

1. Ordene, pelo valor das cargas fatoriais, os m instrumentos em ordem decrescente, isto é, de forma que o instrumento com maior influência no fator esteja em primeiro lugar no vetor de simulação e sucessivamente;
2. gere um número aleatório para o primeiro instrumento no seu respectivo intervalo MC ;
3. gere números aleatórios para os instrumentos $2, \dots, m'$, dos seus respectivos intervalos MC , verificando suas correlações com o primeiro instrumento;
4. gere os demais números aleatórios para os demais instrumentos, dentro dos seus respectivos intervalos de bom comportamento BC , verificando as correlações com o primeiro instrumento;
5. gerar o vetor de simulação, padronizando todas as entradas, subtraindo as respectivas médias, e dividindo pelos respectivos desvios padrão das medidas obtidas por cada instrumento;
6. multiplique esse vetor pelo vetor de coeficientes fatoriais (que também é revelado pela análise fatorial, e que permite a geração dos escores fatoriais) do fator j , some os resultados para obter o valor do escore fatorial correspondente E_{ij} ;
7. padronize E_{ij} obtendo z_{ij} , se necessário aplique antes a transformação em E_{ij} para torná-lo uma variável aleatória normal;
8. plote z_{ij} no eixo x ;
9. repita esse processo para um número suficientemente grande de vezes.

Nesse algoritmo, E_{ij} representa o valor encontrado do escore fatorial j quando simulado a anomalia i , z_{ij} é a transformação desse escore em variável normal padrão.

Como o escore fatorial j é uma variável aleatória normal, e o valor z_{ij} foi obtido simulando uma situação atípica, o valor obtido para ele se distanciará da média 0. Se for realizada essa simulação para um número suficientemente grande de vezes, então é obtido um número suficientemente grande de pontos no eixo x , que normalmente distanciarão da média, formando uma região em uma ou ambas as caldas da curva de Gauss, assim essas regiões são utilizadas para determinar a região crítica da anomalia i , monitorada a partir do fator j . A figura a seguir, apresenta um fluxograma desse algoritmo.

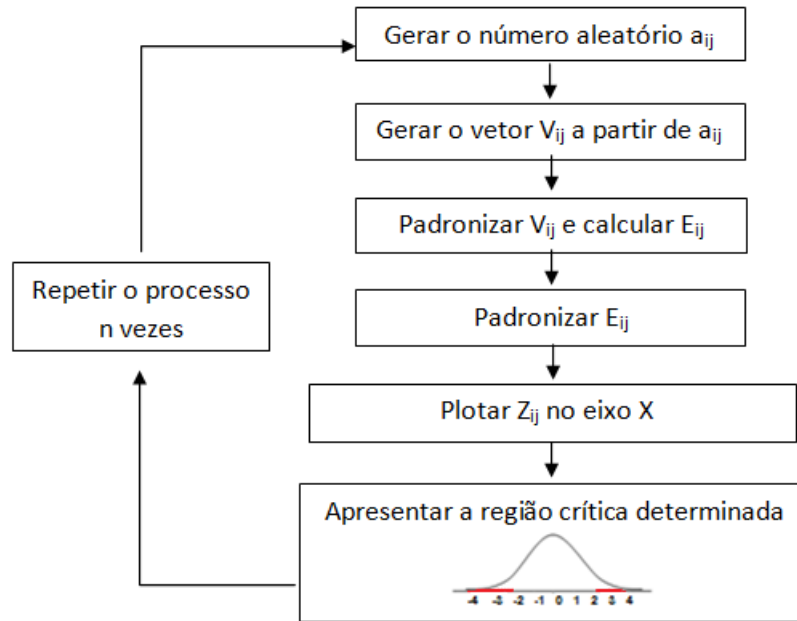


Figura 3.2: Fluxograma do Algoritmo para Determinar os Modos de Falha.

Para determinar as regiões críticas, são analisados os pontos plotados no eixo x . Escolhe-se a região onde eles mais se agrupam, ignorando um número de pontos fora dela, dependendo do percentual de acerto que se deseja. Se um percentual de 90% for satisfatório ao simular um número suficientemente grande de vetores, pode-se rejeitar 10% dos pontos que estejam distantes da região onde eles mais se agrupam, e para essa região, determinar um intervalo, considerando ele como um intervalo da reta real, mesmo sabendo que esse conjunto possui pontos discretos de simulações, porém, em 90% das vezes em que foi simulada aquela anomalia, os pontos se acumularam nesse intervalo.

3.2 Modelo Matemático para o Caso Geral

No trabalho desenvolvido por Silva (2016) foi definido um modelo matemático como função de desempenho que teve o objetivo de realizar o monitoramento de um bloco de concreto de uma barragem altamente instrumentada, onde os instrumentos são correlacionados. Para o caso geral foi assumida algumas hipóteses, que são:

1. A matriz de dados amostrais é proveniente de uma população normal multivariada;
2. Todos os testes estatísticos mencionados anteriormente foram bem sucedidos;
3. A Análise Fatorial foi bem sucedida, garantindo uma explicação da variabilidade superior à 75%;

Supondo que existam n instrumentos instalados no bloco, e após a análise fatorial obteve-se j fatores. Suponha também, que os instrumentos instalados no bloco sejam capazes de monitorar i anomalias, e para cada anomalia exista um número $m \leq n$ instrumentos capazes de identificar i .

Se um único fator possuir os m instrumentos, com alta contribuição no fator, capazes de monitorar a anomalia i , então apenas esse fator será necessário para diagnosticar essa anomalia.

Se existir $m \leq n$ instrumentos capazes de identificar a anomalia i , com esses instrumentos distribuídos em alguns fatores com alta contribuição, então todos esses fatores são necessários para monitorar i .

Por exemplo, sejam as anomalias $i = \{3, 4\}$, tais que, todos os instrumentos capazes de monitorar a anomalia 3 estejam todos agrupados no fator 2, e, todos instrumentos capazes de monitorar a anomalia 4 estejam agrupados em dois fatores, a saber, 3 e 4. Logo, para o primeiro caso, utiliza-se o algoritmo descrito na figura 3.1.1, para fazer simulações com os instrumentos associados a anomalia 3, obtendo valores do escore fatorial 2 no eixo x , obtendo RC_{32} , definida como a região crítica da anomalia 3 monitorada a partir do fator 2. Para o caso da anomalia 4, utiliza-se o algoritmo apresentado na figura 3.1.1 fazendo simulações com os instrumentos capazes de monitorar essa anomalia em ambos os fatores, obtendo RC_{43} e RC_{44} , definidas com região crítica da anomalia 4 a partir do fator 3 e região crítica da anomalia 4 a partir do fator 4. Portanto, dado um vetor de leituras mensais dos instrumentos instalados no bloco de concreto, é dito que a anomalia 3 pode estar ocorrendo quando o valor do escore fatorial 2 pertencer a RC_{32} . E para o caso da anomalia 4, é dito que pode estar ocorrendo se o escore fatorial 3 pertencer a RC_{43} e o escore fatorial 4 pertencer a RC_{44} .

Portanto, considerando um bloco-chave de uma barragem de concreto, foi determinado as anomalias que os instrumentos instalados nesse bloco são capazes de monitorar e, as regiões críticas dessas anomalias que já foram identificadas através do algoritmo descrito anteriormente, pode-se definir uma função de desempenho capaz de realizar o monitoramento, contudo, antes deve-se definir:

1. G_i é a componente da função de desempenho que monitora a anomalia i ;
2. K_{ij} é um conjunto formado apenas pelos escores fatoriais j que se relacionam com a anomalia i ;
3. k_i é o número de fatores relacionados com i menos um, ou seja, $k_i = \#\{K_{ij}\} - 1$.

Portanto, a função de desempenho, é definida como sendo $G(\underline{Z})$, dada por:

$$G(\underline{Z}) = \begin{cases} G_1(\underline{Z}) = \lambda_{11}z_1 + \lambda_{12}z_2 + \dots + \lambda_{1j}z_j + k_1 \\ G_2(\underline{Z}) = \lambda_{21}z_1 + \lambda_{22}z_2 + \dots + \lambda_{2j}z_j + k_2 \\ \vdots \\ G_i(\underline{Z}) = \lambda_{i1}z_1 + \lambda_{i2}z_2 + \dots + \lambda_{ij}z_j + k_i \end{cases} \quad (3.3)$$

Com

$$\begin{cases} \lambda_{ij} = 0 & j \notin K_{ij} \quad \text{ou} \quad z_j \notin RC_{ij} \\ & \text{se} \\ \lambda_{ij} = \frac{-1}{z_j} & z_j \in RC_{ij} \end{cases}$$

onde λ_{ij} representa o coeficiente da anomalia i do escore fatorial j , z_j representa o valor do escore fatorial j padronizado e RC_{ij} representa as regiões críticas definidas para a anomalia i a partir do escore fatorial j . λ_{ij} também assume o valor 0 se o fator j não tiver influência sobre a anomalia i .

Note que o valor assumido pela constante λ_{ij} depende se o valor z_j pertence ou não a sua respectiva região crítica, que foi definida na seção anterior. Logo, dessa função tira-se as seguintes conclusões:

1. $G_i(\underline{Z}) \geq 0, \forall i$ apenas quando nenhuma anomalia ocorreu;
2. $G_i(\underline{Z}) < 0$ para algum i quando uma ou mais anomalias ocorreram;
3. Os índices i 's indicam quais anomalias podem ser monitoradas pela função de desempenho e os índices j 's indicam quais fatores foram responsáveis pela ocorrência da anomalia e, conseqüentemente, quais instrumentos apresentaram valores nos seus respectivos intervalos de mudança de comportamento.

Obteve-se uma função de desempenho multivariada, onde, caso geral, consideramos que cada componente da função é uma combinação linear de variáveis aleatórias normais padrão e, portanto, também possui essa distribuição de probabilidade. Apesar dessa função não utilizar as variáveis físicas normalmente utilizadas na análise de confiabilidade estrutural para determinação da função desempenho, ela preserva a condição de assumir valores menores que zero apenas quando ocorre uma ou mais anomalias.

Nos casos aplicados não é possível garantir que todos os escores fatoriais possuam uma distribuição de probabilidade normal e, nesses casos, a função de desempenho não poderá determinar a probabilidade de falha de cada anomalia mas, ainda sim, poderá prever quando uma falha pode estar ocorrendo. No próximo capítulo do trabalho, inclui-se alguns conceitos na função de desempenho que será definida para o bloco-chave A-7 da Barragem de Itaipu, conceitos como semi função e sinais de alerta para tornar o sistema mais próximo possível do que é atualmente utilizado pelos profissionais de Barragem Da Usina Hidrelétrica (UHE) de Itaipu. Essas novas informações que a função de desempenho fornecerá serão algumas das contribuições científicas que esse trabalho propôs a cerca do tema.

Capítulo 4

Testes e Resultados

Neste capítulo é aplicado o modelo descrito anteriormente, no bloco-chave A-7 da UHE de Itaipu, cujo os dados dos instrumentos estão disponíveis em um banco de dados disponibilizado pelos profissionais em segurança de barragens da UHE de Itaipu.

É criada uma matriz $M_{202 \times 40}$ multivariada, com as leituras geradas pelos instrumentos instalados no bloco em estudo, que estão apresentados no quadro 4.1, incluindo outros dois instrumentos, que medem a temperatura ambiente e o nível do reservatório, que não estão instalados no bloco-chave A-7 mas influenciam todo o sistema. Foi considerado um período de tempo a partir de janeiro de 2000 até outubro de 2016, correspondendo a 202 meses. As leituras no geral não estavam padronizadas, então para possibilitar a análise, os dados mensais foram padronizados. Para os instrumentos que possuíam mais de uma leitura mensal, foram calculadas as médias, que é o caso da maioria. No entanto, alguns instrumentos possuíam apenas leituras trimestrais, o caso dos Termômetros Internos, neles foi aplicado a técnica de interpolação Spline Cúbica, que já se encontra disponível implementada no programa matemático *Matlab*.

Instrumentos Bloco-Chave A-7			
Instrumentos	Quantidades	Unidade de Leitura	Função
Pêndulo (PD)	2	mm	Mede os deslocamentos na direção da jusante e de montante.
Extensômetro (EM)	6	mm	Mede o deslocamento e deformações que ocorrerão entre pontos no maciço da fundação.
Base de Alongômetro (JS)	4	mm	Medidor superficial de Junta para o Concreto.
Piezômetro Elétrico (PG)	1	msnm	Mede a Subpressão dos poros no núcleo e no contato núcleo-fundação.
Tensômetro de Concreto (TN)	1	kgf/cm ²	Para medição direta das tensões de compressão no concreto.
Termômetro Superficial (TS)	1	°C	Termômetro embutidos no concreto para medição da temperatura durante a construção e operação.
Medidor de Vasão	1	l/s	Para medição de vasão de percolação pelas estruturas e suas fundações.
Piezômetro Standpipe (PS)	7	msnm	Piezômetros para medir a pressão dos poros e a subpressão hidráulica em juntas ou contatos escolhidos no maciço rochoso ou no contato barragem-fundação.
Termômetro Interno (TI)	4	°C	Termômetros embutidos no concreto para medição da temperatura durante a construção e operação.
Roseta Deformação (RD)	11	kgf/cm ²	Embutidas no concreto para medição das variações de volume autógenas, fluência e deformação elástica e temperatura.

Quadro 4.1: Instrumentos Instalados no Bloco-Chave A-7.3.1

4.1 Aplicação do Método Estatístico Multivariado ao Bloco-Chave A-7 da Barragem de Itaipu

Os materiais empregados nesse estudo são as variáveis aleatórias geradas pelas leituras dos instrumentos instalados no bloco-chave A-7 da Barragem de Itaipu, bloco esse que pertence ao complexo do vertedouro. No total foram 40 instrumentos analisados, na qual, a grande maioria, estão instalados apenas no bloco-chave A-7, porém alguns instrumentos também pertencem aos blocos A-8 e A-6, pois estão instalados na divisa desses blocos para monitorar aberturas, recalques e deslizamentos.

Na figura a seguir é apresentado o bloco-chave que foi utilizado para o estudo de caso desse trabalho.

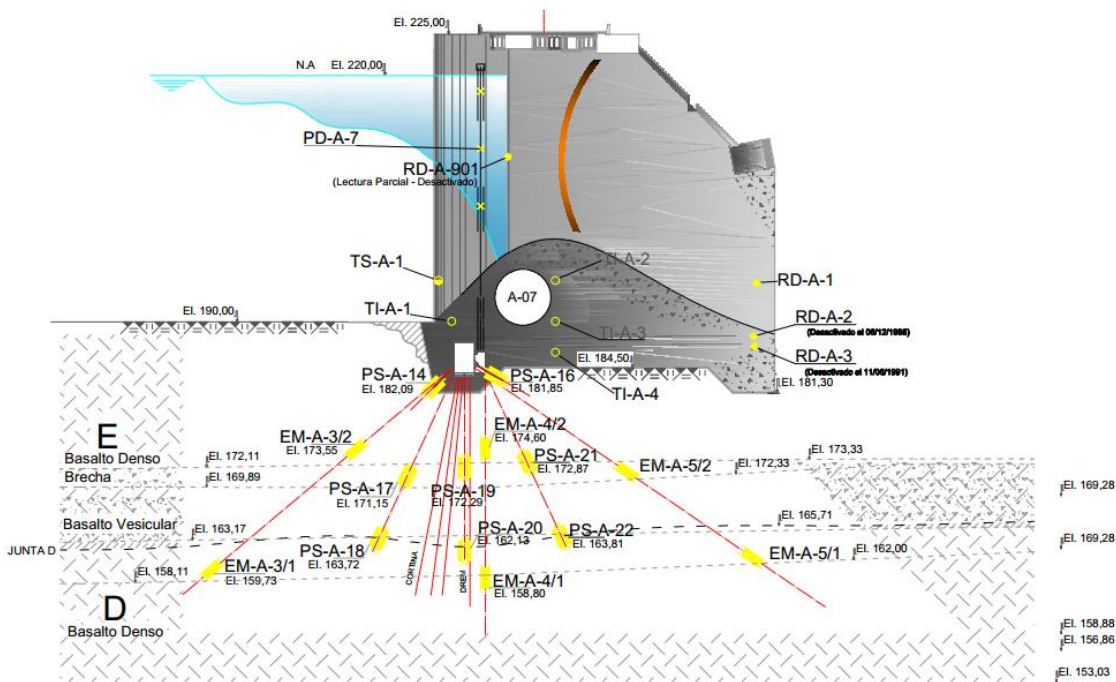


Figura 4.1: Bloco-Chave A-7. Fonte: Binacional (2003)

Os instrumentos instalados na Barragem de Itaipu e as funções que desempenham na estrutura são encontrados em (BINACIONAL, 2003), porém pode ser visto no quadro 4.1 uma breve descrição dos instrumentos que são utilizados nesse estudo, apenas aqueles que estão instalados no bloco de interesse, e estão disponíveis nos manuais Sistema de Operação e Manutenção de Itaipu (SOM) que foram fornecidos pela equipe técnica de Itaipu.

4.1.1 Testes Estatísticos

Antes de aplicar a análise fatorial na matriz $M_{202 \times 40}$ utilizando o *Software Statistica*, verificou-se alguns testes estatísticos, indispensáveis para aplicação do método.

Teste de Normalidade Multivariada

Inicialmente foi testada a normalidade multivariada da matriz $M_{202 \times 40}$. Foi utilizado o teste computacionalmente intensivo baseado na distância de Mahalanobis (BIASE, 2012). A hipótese nula desse teste afirma:

H_0 : A matriz de dados amostrais é proveniente de uma população normal multivariada.

A aplicação do teste, programado no Matlab, resultou em $p = 0.7552$ (obtido com 100.000 simulações Monte Carlo). Portanto, ao nível de significância de 0,05, a hipótese de normalidade multivariada não pode ser descartada.

Teste de Esfericidade de Bartlett

Sabendo-se da normalidade multivariada dos dados, aplicou-se o Teste de Esfericidade de Bartlett para testar as hipóteses:

H_0 : A matriz de correlação é igual a matriz identidade, ou seja, $\rho = I$;

H_1 : A matriz de correlação é diferente da matriz identidade, ou seja, $\rho \neq I$.

A aplicação do teste resultou em $X^2 = 7980$, $p = 0,9920$, ou seja, rejeitou-se a hipótese H_0 . Portanto, a amostra multivariada é proveniente de uma população com matriz de correlação significativamente diferente da matriz identidade.

Cálculo de KMO

O cálculo da medida de adequacidade dos dados para realização de uma análise fatorial resultou em $KMO = 0,8349$. Como o valor mínimo exigido é 0,50, a amostra multivariada satisfaz essa condição.

Extração dos Fatores

Comprovados os resultados dos testes estatísticos, aplica-se a análise fatorial na matriz $M_{202 \times 40}$. O método adotado para obtenção dos coeficientes fatoriais foi o Método das Componentes Principais com rotação Varimax e aplicando o critério do mínimo autovalor, visto que esse método obteve uma maior porção da variância explicada, comparado com o método da Máxima Verossimilhança.

Obteve-se 5 fatores, que explicam 88,07% de toda variância, entretanto, duas variáveis (instrumentos) apresentaram comunalidades baixas, menores que 0,70, e devido a isso, uma nova análise foi realizada removendo essas variáveis. A nova análise envolveu todos os testes estatísticos realizados anteriormente, mas agora em uma matriz $M_{202 \times 38}$ para reconfirmar adequação dos dados. Novamente todos os testes estatísticos foram bem sucedidos.

Com a nova análise, foi obtido uma variância explicada de 89,43%, sem nenhuma variável com baixa comunalidade. Assim, considerando o apresentado na literatura e o conhecimento prático dos profissionais em segurança de barragens da UHE de Itaipu, foi estabelecido ser possível com a determinação destes grupos, monitorar as seguintes anomalias: escorregamentos e subpressão, fluência e tombamento.

De acordo com a Análise Fatorial realizada, o quadro a seguir apresenta os cinco grupos de instrumentos que são altamente correlacionados, denominados fatores.

Grupos de Instrumentos Altamente Correlacionados				
Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
JS-A-13/desliz	JS-A-13/abert	EM-A-3/2	RD-A-1/tr1	PS-A-22/msnm
EM-A-3/1	JS-A-14/abert	CO-A-1/X Asoc	RD-A-1/tr2	RD-A-1/dp1
EM-A-4/1	JS-A-14/asent	CO-A-1/Y Asoc	RD-A-1/tr3	RD-A-1/dp2
EM-A-4/2	EM-A-5/1	PG-A-8/msnm	RD-A-1/tr4	RD-A-1/dp3
EM-A-5/2	TN-A-1	RD-A-1/dp5	RD-A-1/tr5	RD-A-1/dp4
PS-A-14/msnm	TI-A-1	TS-A-1	Nível do Reserv	RD-A-1/dp6
PS-A-17/msnm		Temp Amb		TI-A-3
PS-A-18/msnm				TI-A-4
PS-A-19/msnm				
PS-A-20/msnm				
PS-A-21/msnm				

Quadro 4.2: Grupos de Instrumentos Altamente Correlacionados.

Os escores fatoriais gerados a partir dos coeficientes fatoriais são utilizados para a determinação dos modos de falha. Neles são realizados os testes estatísticos para comprovar se são variáveis aleatórias normais ou se serão necessárias transformações nos escores para transformá-los em variáveis aleatórias normais. No entanto, mesmo aplicando essas transformações nem sempre é possível obter uma condição de normalidade nos escores e, nesse caso, é realizado apenas a determinação das regiões críticas de cada anomalia e seu monitoramento, sem se preocupar com o cálculo da probabilidade de falha.

Adequação dos Escores Fatoriais

Após obter os fatores e gerar os escores fatoriais da forma especificada no capítulo 3, foram obtidos cinco escores fatoriais, que são as variáveis aleatórias que substituirão as originais, com pouca perda de informação. Essas novas variáveis aleatórias, de acordo com a teoria da análise fatorial, são independentes entre si.

O quadro a seguir, apresenta os resultados obtidos após os testes estatísticos e as transformações aplicadas.

Resultados obtidos após Teste Lilliefors			
Escores	Estimativas	Transformação	Decisão Teste
1	d= 0.17910, p<0.01	$\sqrt{ x }$	Não Normal
2	d=0.08531, p<0.15	$\sqrt{ x }$	Não Normal
3	d=0.10556, p<0.05	$\sqrt{ x }$	Não Normal
4	d=0.07156, p>0.20	Não	Normal
5	d=0.20658, p<0.01	$\sqrt{ x }$	Não Normal

Quadro 4.3: Resultados Obtidos após o Teste Lilliefors.

Analisando o quadro 4.3, observa-se que apenas o escore fatorial 4 já possui a condição de normalidade. Já os escores 1, 2, 3 e 5 não são normais e não foi encontrado até o momento nenhuma transformação para eles, se o objetivo desse trabalho fosse o cálculo da probabilidade de falha, os mesmos não seriam utilizados no monitoramento da estrutura. No entanto, neste trabalho o objetivo é determinar os modos de falha possíveis de serem monitorados e suas regiões críticas, sem a preocupação da estimativa dessas probabilidades de falha e, por esse motivo eles puderam ser incluídos.

4.2 Determinação dos Modos de Falha

Define-se o primeiro vetor de simulação da anomalia i a partir do fator j como sendo

$$V_{ij} = (a_{1,1}; a_{1,2}; \dots; a_{1,k}; a_{1,k+1}; a_{1,k+2}; \dots; a_{1,n})$$

onde, $a_{1,m}$ é o primeiro número aleatório gerado para o instrumento m do seu respectivo intervalo MC e k é a quantidade de instrumentos altamente correlacionados com a anomalia i que pertencem ao fator j , a_{pk} são os números aleatórios gerados dos intervalos de bom comportamento BC , dos instrumentos $p + k$ com $p = 1, \dots, n - k$ e sendo n o número total de instrumentos. Para que esse vetor de simulação seja eficiente para cada fator j que é utilizado, ele deve ser reordenado de forma a deixar os instrumentos principais nas primeiras entradas e os demais na sequência, e além disso, a geração de todos os números aleatórios desse vetor, depende das correlações com o primeiro número aleatório gerado. O primeiro número aleatório é gerado do instrumento com maior representação no fator j que tenha influência com a anomalia i .

Na análise fatorial realizada, foi obtido uma matriz com os coeficientes fatoriais, que são essenciais na geração dos escores fatoriais. Dessa matriz obtêm-se o vetor V_{cj} capaz de gerar o escore fatorial j , define-se agora o vetor

$$W = \begin{cases} \frac{a_{11} - \bar{x}_1}{s_1} \\ \frac{a_{12} - \bar{x}_2}{s_2} \\ \vdots \\ \frac{a_{1,k+1} - \bar{x}_{k+1}}{s_{k+1}} \\ \frac{a_{1,k+2} - \bar{x}_{k+2}}{s_{k+2}} \\ \vdots \\ \frac{a_{1n} - \bar{x}_n}{s_n} \end{cases}$$

como sendo o vetor de simulação padronizado, onde \bar{x}_n e s_n são, respectivamente, as médias e os desvios padrão da amostra das variáveis aleatórias, desde Janeiro de 2000 a outubro de 2016.

Para esse caso o vetor V_{cj} possui dimensões 38×1 . Fazendo $(V_{cj})_{38 \times 1}^t \cdot W_{38 \times 1}$ obtém-se E_{1ij} que é o primeiro valor do escore fatorial j quando simulada a anomalia i . Esse valor E_{1ij} pode ser padronizado de acordo com sua média e variância, obtendo-se z_{1ij} .

Após, esses cálculos plota-se z_{1ij} no gráfico, obtêm-se um ponto no eixo x . Esse é o primeiro ponto da região crítica que é definida para anomalia i . Repetindo esse processo para um grande número de vezes obtêm-se, esse número de pontos no eixo x . Como esses pontos são gerados simulando a mesma anomalia, eles se concentram em regiões do plano cartesiano.

Na sequência do trabalho, usa-se com frequência o vetor W_a , denominado vetor de amostras das leituras das medidas padronizadas dos instrumentos instalados no bloco em um certo mês, que pretende-se verificar se houve ou não alguma anomalia na estrutura.

As anomalias que pretende-se monitorar e as regiões críticas para cada uma delas, são apresentadas a seguir.

4.2.1 Escorregamentos na Brecha D

As forças horizontais mobilizam tensão de cisalhamento ao longo de uma superfície de baixa resistência ao cisalhamento, preferencialmente horizontal no corpo da barragem, em sua base ou em um plano de fraqueza da fundação. O resultado pode ser um deslizamento, ou seja um deslocamento para jusante (JAIME, 2006). O escorregamento de um bloco de uma barragem de concreto é quando há um movimento de abertura e deslizamento entre os blocos vizinhos, fazendo com que ele se movimente para uma determinada direção, no caso do bloco-chave A-7 esse movimento pode ocorrer na direção vertical oposta ao reservatório.

Os escores fatoriais dos fatores 1 e 3 podem monitorar o escorregamento na brecha D, esse escorregamento pode ocorrer quando existe uma mudança no comportamento das bases de alongâmetro instaladas no bloco-chave A-7, em conjunto com uma subpressão detectada pelos piezômetros instalados na brecha D.

Para determinar a região crítica dessa anomalia utiliza-se o algoritmo representado na figura 3.2. Faz-se simulações, onde apenas os instrumentos relacionados com o escorregamento e com a subpressão na brecha D apresentam valores dentro dos intervalos MC , e nesse caso determina-se a região crítica da anomalia 1 a partir do fator 1, $RC_{11} = (7,3; 11,4)$ e também a região crítica da anomalia 1 a partir do fator 3, $RC_{13} = (-1,5; 5,5)$. A figura a seguir apresenta as regiões críticas para essa anomalia.

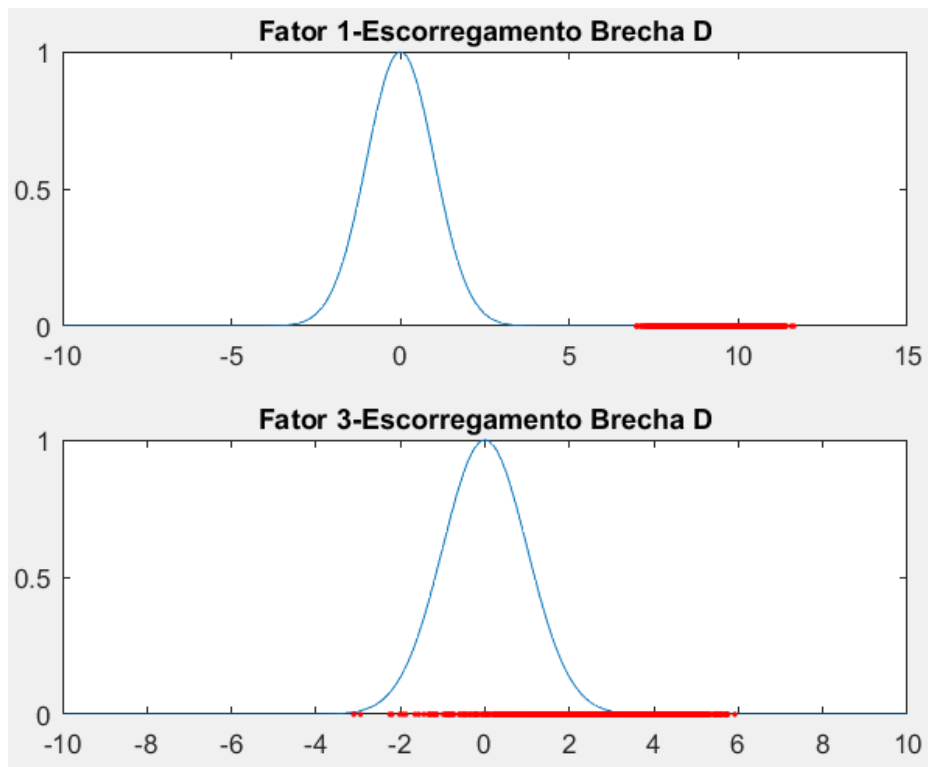


Figura 4.2: Região Crítica do Escorregamento na Brecha D.

Portanto, para realizar o monitoramento, dado um vetor W_a definido anteriormente, é dito que essa anomalia pode ocorrer quando:

$$\{z_1, z_3 \in RC_{11}, RC_{13}\}$$

onde z_j é o valor do escore fatorial padronizado gerado a partir de W_a e do fator j , e RC_{ij} é a região crítica determinada da anomalia i a partir do escore fatorial j .

4.2.2 Subpressão

A subpressão na brecha D é quando os piezômetros instalados nessas regiões que se encontra a uma determinada profundidade, apresentam valores fora da normalidade. Esses instrumentos estão associados ao fator 1.

A anomalia apresentada neste momento é a subpressão na Brecha D, que pode ocorrer sem que haja um escorregamento. Como os piezômetros que monitoram tal anomalia estão associados ao fator 1, utiliza-se o algoritmo descrito na figura 3.2 e faz-se simulações onde apenas os instrumentos relacionados com a subpressão na Brecha D apresentam valores dentro

dos intervalos MC , e nesse caso determina-se a região crítica de anomalia 2 a partir do fator 1, $RC_{21} = (6,5; 10)$. A figura a seguir apresenta a região crítica determinada.

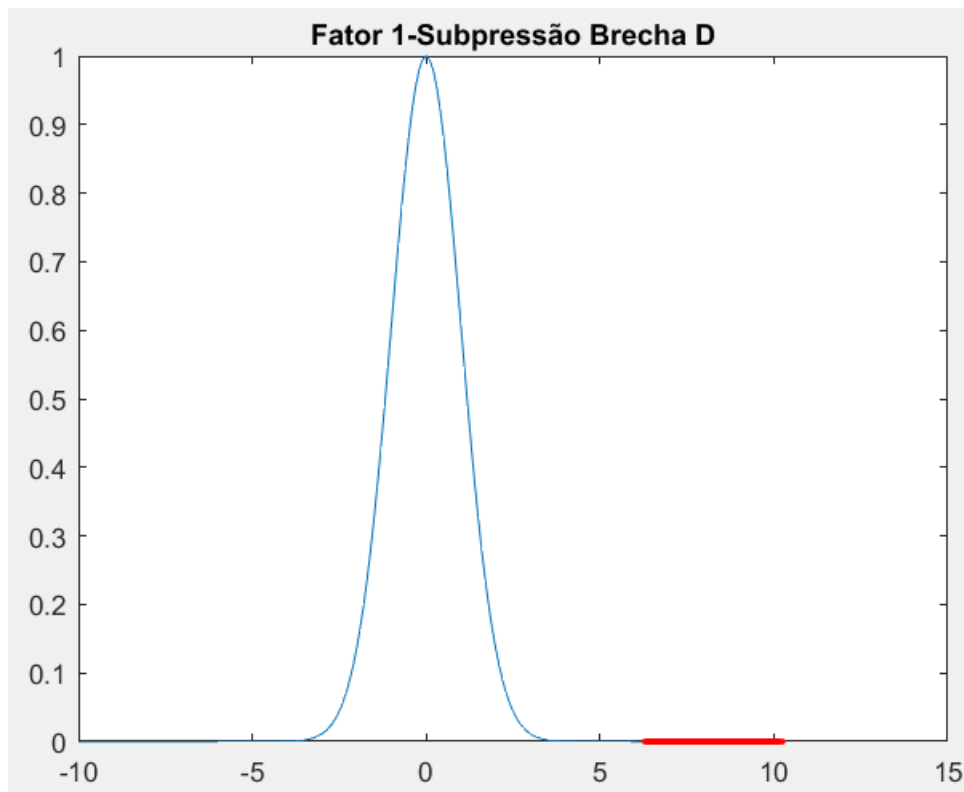


Figura 4.3: Região Crítica da Subpressão na Brecha D.

Portanto, para realizar o monitoramento, dado um vetor W_a definido anteriormente, é dito que essa anomalia pode ocorrer quando:

$$\{z_1 \in RC_{21}\}$$

onde z_j é o valor do escore fatorial padronizado gerado a partir de W_a e do fator j , e RC_{ij} é a região crítica determinada da anomalia i a partir do escore fatorial j .

4.2.3 Escorregamentos na Junta D

Os escores fatoriais dos fatores 1 e 2 podem monitorar o escorregamento na junta D, esse escorregamento pode ocorrer, com uma mudança de comportamento das bases de alongâmetro em conjunto com uma subpressão detectada pelos piezômetros instalados na junta D.

Para determinar a região crítica dessa anomalia utiliza-se o algoritmo descrito na figura 3.2 e faz-se simulações onde apenas os instrumentos relacionados com o escorregamento e com a subpressão na junta D apresentam valores dentro dos intervalos MC , e nesse caso determina-se a região crítica de anomalia 3 a partir do fator 1, $RC_{31} = (7; 9,9)$ e também a região crítica da anomalia 3 a partir do fator 2, $RC_{32} = (-3; 3,1)$. A figura a seguir apresenta as regiões críticas para essa anomalia.

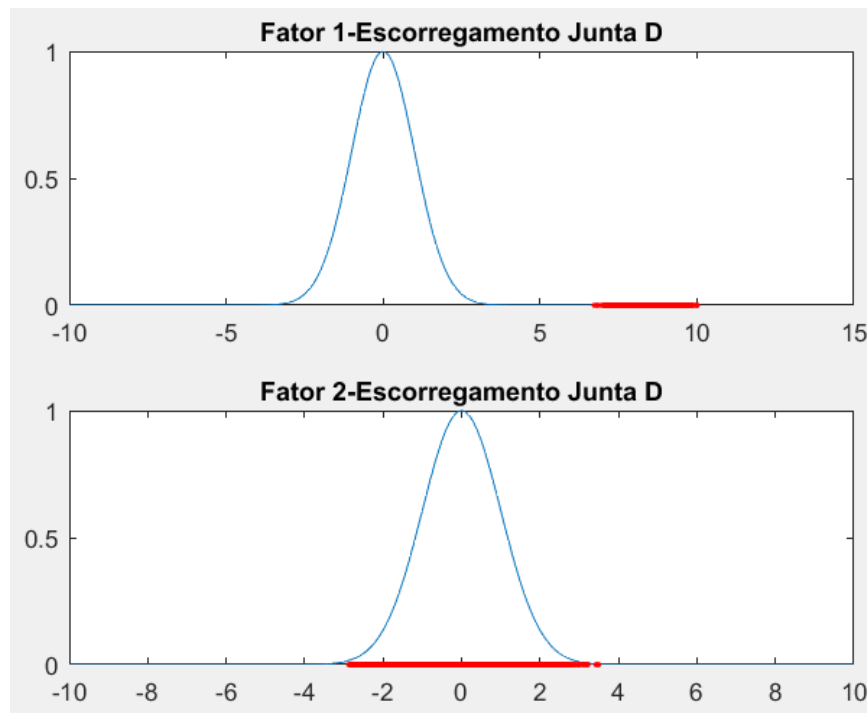


Figura 4.4: Região Crítica do Escorregamento na Junta D.

Portanto, para realizar o monitoramento, dado um vetor W_a definido anteriormente, é dito que essa anomalia pode ocorrer quando:

$$\{z_1, z_2 \in RC_{31}, RC_{32}$$

onde z_j é o valor do escore fatorial padronizado gerado a partir de W_a e do fator j , e RC_{ij} é a região crítica determinada da anomalia i a partir do escore fatorial j .

4.2.4 Fluência

Nas estruturas de concreto massa, onde se têm grandes volumes de concreto, deve-se ter especial preocupação com o estudo e controle da geração de calor, ocasionada pela reação de hidratação do cimento. Esta geração de calor pode causar efeitos mecânicos que atingindo certos limites, pode levar a fissuração e posterior dano da estrutura. Somando-se ao controle de geração de calor, deve-se ter atenção também a um outro fenômeno mecânico causado pela acomodação dos agregados, através da ação constante de uma carga, chamado de fluência (JAIME, 2006).

Os instrumentos capazes de monitorar uma fluência são as rosetas de deformímetro que estão altamente correlacionadas com os fatores 3 e 4. Utilizando o algoritmo descrito na figura 3.2 faz-se simulações onde os valores desses instrumentos apresentam leituras dentro de seus respectivos intervalos MC , em todos fatores envolvidos. Após um número de simulações obtém-se as regiões críticas $RC_{43} = (8,7; 12,2)$ e $RC_{44} = (0,2; 2,9)$ denominada região crítica da anomalia 4 do fator j . A figura que segue apresenta as regiões críticas obtidas para essa anomalia.

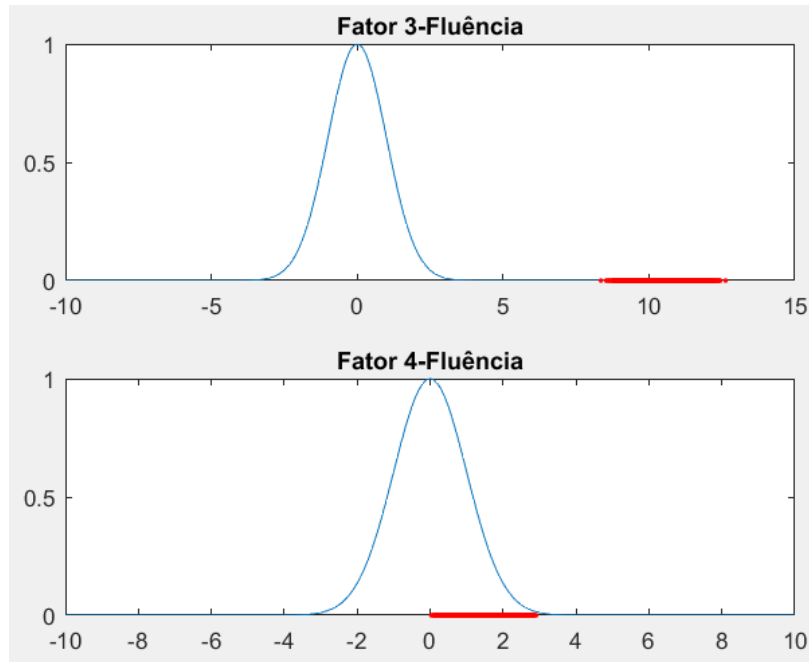


Figura 4.5: Região Crítica da Fluência.

Logo, dado um vetor W_a é dito que pode estar ocorrendo uma Fluência quando ambas situações acontecem:

$$\{z_3, z_4 \in RC_{43}, RC_{44}$$

onde z_j é o valor do escore fatorial padronizado obtido através de W_a e do fator j .

4.2.5 Tombamento

A anomalia que pretende-se monitorar agora é o tombamento, que pode acontecer na direção da montante ou da jusante, como resultado das forças atuantes na estrutura. O tombamento é o mecanismo de ruptura da barragem como um todo, tendendo a fazê-la girar em torno de um ponto A . O tombamento é, no entanto, um mecanismo raro de ocorrer. Antes da ruptura por tombamento, desenvolvem-se esforços de tração e aumento de subpressão na rocha, à montante. Ocorrem, também, aumentos das tensões de compressão à jusante. As forças horizontais H estão presentes e o resultado é a ruptura por deslizamento antes da ocorrência do tombamento. A barragem desliza no seu contato inferior, ao invés de girar em torno do ponto A . O mal desempenho da barragem é portanto o deslizamento, um problema que pode ser solucionado por injeções de cimento.

Para realizar esse monitoramento deve-se envolver todos os fatores que possuam os instrumentos que possam identificar um tombamento. De acordo com as funções e localidades dos instrumentos instalados nesse bloco, identificou-se os fatores que contribuem com essa tarefa, nesse caso, são os fatores 1, 2, 3 e 5. Define-se como um possível estado de tombamento, quando todos esses fatores apresentarem a resultante dos seus escores fatoriais nas suas respectivas regiões críticas. Isso pode ser representado como um sistema em paralelo do evento tombamento.

As regiões críticas dos fatores são obtidas utilizando o algoritmo descrito na figura 3.2 simulando valores para os instrumentos dentro de seus respectivos intervalos MC , para todos os escores fatoriais envolvidos. A figura abaixo apresenta as regiões críticas $RC_{51} = (1,5; 8,7)$, $RC_{52} = (2,8; 8,5)$, $RC_{53} = (-3; 5)$ e $RC_{55} = (9,5; 14,5)$ obtidas.

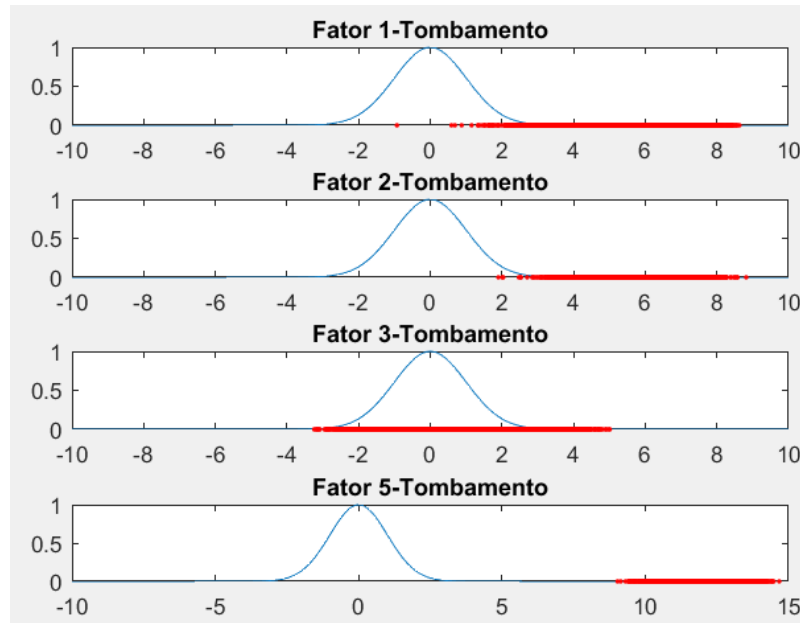


Figura 4.6: Região Crítica do Tombamento.

Portanto, dado um vetor W_a é dito que existe a possibilidade de ocorrer um tombamento quando ambas situações acontecem:

$$\{z_1, z_2, z_3, z_5 \in RC_{51}, RC_{52}, RC_{53}, RC_{55}\}$$

onde z_j é o valor do escore fatorial padronizado obtido a partir de W_a e do fator j .

4.2.6 Sinais de Alerta

O trabalho desenvolvido por Silva (2016), apresenta um modelo matemático multivariado, cujo objetivo foi possibilitar o monitoramento das possíveis anomalias que podem ocorrer em um bloco de concreto altamente instrumentado, sem a preocupação de definir um sistema de alerta que fosse semelhante ao que é utilizado pelos profissionais em segurança de barragens da Usina Hidrelétrica de Itaipu.

Por esse motivo, neste trabalho é proposto um sistema de alerta onde seja possível a comparação com o sistema que é aplicado pelos profissionais de Itaipu para que, futuramente, através de um banco de dados criado, seja possível comparar os dois métodos. De acordo com esses profissionais, são utilizados três tipos de alertas para o monitoramento, o alerta verde que indica um bom estado de comportamento, um alerta amarelo que indica alguma mudança no comportamento mas nada sério e por fim, um alerta vermelho que indica uma mudança em um grupo específico de instrumentos importantes, que quando ocorre pode gerar um problema na estrutura, sendo necessário uma vistoria e, permanecendo esse estado, uma ação para corrigir o problema.

Para tornar esse sistema de alerta o mais condizente possível com o atual sistema utilizado pelos profissionais em segurança de barragens da Itaipu, o mesmo será definido da seguinte forma:

1. Sinal Verde: quando a função de desempenho apresentar a $G_i > 0$, ou seja, quando não apresentar o valor de nenhum escore fatorial z_j dentro do intervalo que indica a região crítica para a anomalia i ;
2. Sinal Amarelo: quando a função de desempenho apresentar a $G_i = 0$, ou seja, quando uma certa anomalia i , apresentar uma quantidade de escores fatoriais $j - 1$ vezes dentro de suas respectivas regiões críticas.
3. Sinal Vermelho: quando a função de desempenho apresentar a $G_i < 0$, ou seja, quando o valor de todos os escores fatoriais z_j estejam dentro do intervalo que indica a região crítica para a anomalia i ;

4.3 Semi-Componentes da Função de Desempenho

De acordo com os profissionais em segurança de barragens de Itaipu, para que uma certa anomalia ocorra é necessário que uma combinação de leituras de certos instrumentos apresente uma mudança de comportamento. Essa propriedade é o que possibilita a utilização da análise fatorial no monitoramento da Barragem. No entanto, algumas anomalias necessitam de alguns instrumentos que foram agrupados com pouca contribuição nos fatores e de forma isolada pela análise fatorial, o que impossibilita a identificação completa dessa anomalia.

Para aproveitar os agrupamentos da análise fatorial o máximo possível, é definido como semi-componente da função de desempenho a anomalia que pretende-se monitorar, mesmo que esteja incompleta, isso significa que nem todos instrumentos necessários para seu monitoramento foram utilizados e, conseqüentemente, essa componente não poderá indicar um alerta vermelho para tal anomalia, nesses casos será possível detectar apenas um alerta amarelo, sendo que o profissional em segurança de barragens terá que consultar os instrumentos faltantes nessa análise para poder constatar um alerta vermelho.

4.3.1 Recalque ou Distensão Pé Montante na Junta D

Entende-se por Recalque ou Distensão Pé Montante o movimento vertical que afeta a estrutura, tendo por causa o terreno. Sabendo que toda a estrutura tem peso, o solo recebe a carga que vai provocar movimento e não há como evitar os deslocamentos dos apoios, caso esse movimento seja uma compressão da estrutura a anomalia ocorrida será o recalque, caso contrário, ou seja, se a estrutura distender, a anomalia registrada será uma distensão pé montante. Na prática, o recalque ou a distensão podem ser tão pequenos que se tornam desprezíveis em terrenos especiais. Costuma-se, então, definir o recalque e a distensão admissíveis, para que não cause problemas ao desempenho de uma estrutura (FILHO, 2006).

As anomalias definidas anteriormente serão acrescentadas como semi componentes da função de desempenho para o bloco-chave A-7, pois em ambos os casos um extensômetro e um

piezômetro que contribuem com seus monitoramentos foram agrupados de maneira isolada pela AF. Logo, para verificação de um sinal vermelho, os profissionais em segurança de barragens de Itaipu deverão consultar as leituras desses específicos instrumentos faltantes, caso ocorra um sinal amarelo para essas componentes.

Os instrumentos capazes de monitorar um recalque ou uma distensão pé montante na Junta D são, de acordo com os profissionais em Segurança de Barragens de Itaipu, os extensômetros e piezômetros que estão instalados nessa feição da fundação, ambos estão altamente correlacionadas com o fator 1. Portanto, utilizando o algoritmo descrito na figura 3.2 faz-se simulações onde os valores desses instrumentos apresentam leituras dentro de seus respectivos intervalos MC , no fator 1. Após um número de simulações obtém-se as regiões críticas $RC_{61} = (5; 8.7)$ e $RC_{71} = (-5.5; -3.3)$, denominadas regiões críticas das anomalias Recalque na Junta D e Distensão Pé Montante na Junta D, respectivamente. A figura adiante, apresenta as regiões críticas obtidas para essas anomalias.

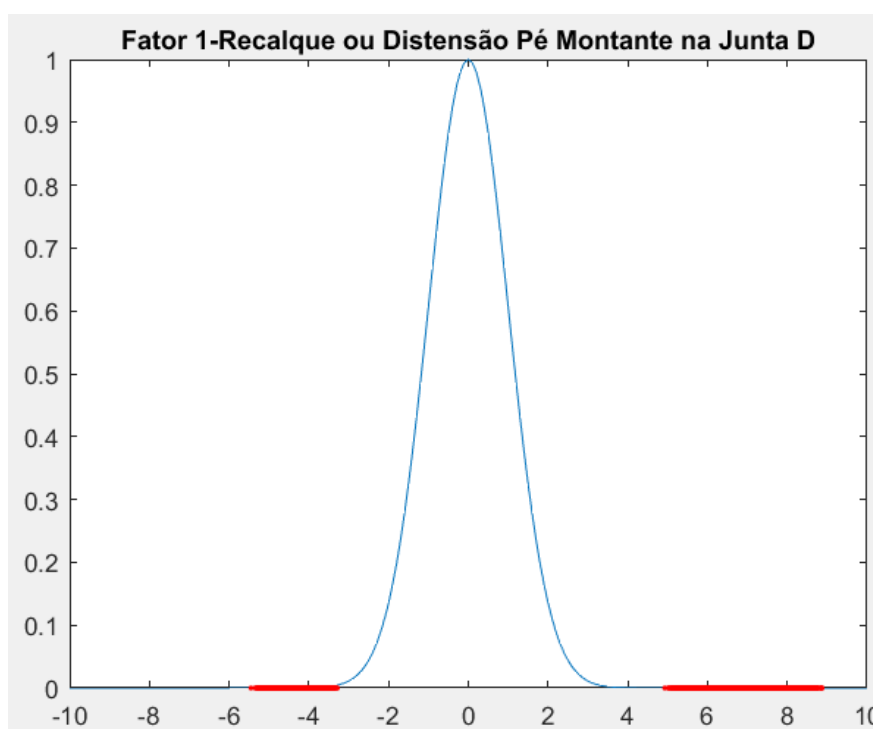


Figura 4.7: Região Crítica do Recalque ou Distensão Pé Montante na Junta D.

4.3.2 Recalque ou Distensão Pé Montante na Brecha D

Os instrumentos capazes de monitorar um recalque ou uma distensão pé montante na Brecha D são, de acordo com os profissionais em Segurança de Barragens de Itaipu, os extensômetros e piezômetros que estão instalados nessa feição da fundação, ambos estão altamente correlacionadas com o fator 1. Utilizando o algoritmo descrito na figura 3.2 faz-se simulações onde os valores desses instrumentos apresentam leituras dentro de seus respectivos intervalos MC , no fator 1. Após um número de simulações obtém-se as regiões críticas $RC_{81} = (3.6; 5.5)$ e $RC_{91} = (-6.5; -4.3)$, denominadas regiões críticas das anomalias Recalque na

Brecha D e Distensão Pé Montante na Brecha D, respectivamente. A figura a seguir, apresenta as regiões críticas obtidas para essas anomalias.

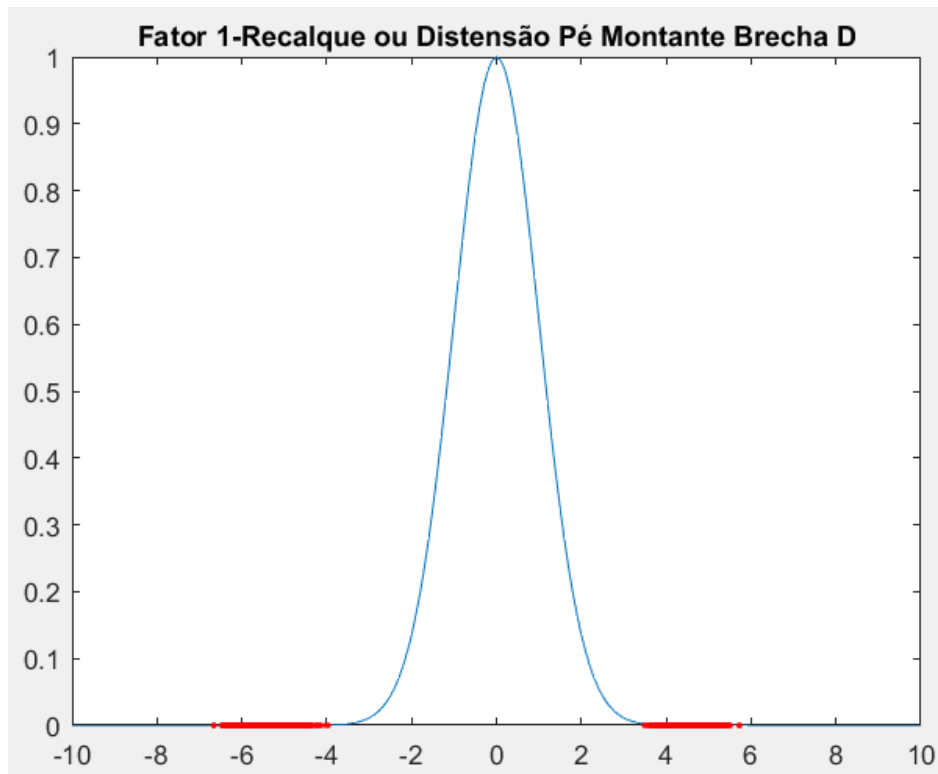


Figura 4.8: Região Crítica do Recalque ou Distensão Pé Montante na Brecha D.

Observação 4.1: As regiões críticas das anomalias Recalque e Distensão Pé Montante podem ser verificadas em um mesmo gráfico pelo fato delas serem monitoradas pelos mesmos instrumentos, a única diferença é em qual dos intervalos de mudança de comportamento as leituras foram simuladas. Se as leituras forem simuladas no intervalo esquerdo do MC significa que houve um Recalque, caso contrário, a anomalia registrada é uma Distensão. Essa característica das anomalias pode ser observada nos gráficos acima, pois observa-se duas regiões bem definidas em ambos os lados. No entanto, leituras geradas para essas anomalias nos seus respectivos intervalos de *MC*, geram escores fatoriais em lados opostos, ou seja, apesar das leituras serem geradas do lado esquerdo do intervalo de mudança de comportamento para a anomalia recalque, sua região crítica está definida do lado direito do gráfico apresentado anteriormente.

4.4 Função de Desempenho Multivariado para o Caso Aplicado

Para gerar o modelo matemático para o caso aplicado é conveniente apresentar a associação dos fatores com as anomalias. O quadro a seguir enumera e apresenta cada uma das anomalias com os fatores envolvidos em cada uma delas.

Associação dos Fatores com as Anomalias		
Índice	Anomalia	Fatores Envolvidos
1	Escorregamento na Brecha D	1, 3
2	Subpressão na Brecha D	1
3	Escorregamento na Junta D	1, 2,
4	Fluência	3, 4
5	Tombamento	1, 2, 3, 5
6	Recalque na Junta D	1
7	Distensão na Junta D	1
8	Recalque na Brecha D	1
9	Distensão na Brecha D	1

Quadro 4.4: Associação dos Fatores com as Anomalias.

Para definir a função de desempenho para o modelo aplicado utiliza-se as informações apresentados no quadro acima. Nesse quadro observa-se que para o bloco-chave A-7 de Itaipu foi possível monitorar 9 anomalias e quais fatores estão encarregados de monitorá-las.

Portanto, para esse caso a função de desempenho tem a seguinte forma:

$$G(\underline{Z}) = \begin{cases} G_1(\underline{Z}) = \lambda_{11}z_1 + \lambda_{13}z_3 + 1 \\ G_2(\underline{Z}) = \lambda_{21}z_1 \\ G_3(\underline{Z}) = \lambda_{31}z_1 + \lambda_{32}z_2 + 1 \\ G_4(\underline{Z}) = \lambda_{43}z_3 + \lambda_{44}z_4 + 1 \\ G_5(\underline{Z}) = \lambda_{51}z_1 + \lambda_{52}z_2 + \lambda_{53}z_3 + \lambda_{55}z_5 + 3 \\ G_6(\underline{Z}) = \lambda_{61}z_1 \\ G_7(\underline{Z}) = \lambda_{71}z_1 \\ G_8(\underline{Z}) = \lambda_{81}z_1 \\ G_9(\underline{Z}) = \lambda_{91}z_1 \end{cases} \quad (4.1)$$

Com

$$\begin{cases} \lambda_{ij} = 0 & j \notin K_{ij} \text{ ou } z_j \notin RC_{ij} \\ & \text{se} \\ \lambda_{ij} = \frac{-1}{z_j} & z_j \in RC_{ij} \end{cases}$$

onde λ_{ij} representa o coeficiente da anomalia i do escore fatorial j , z_j representa o valor do escore fatorial j padronizado e RC_{ij} representa as regiões críticas definidas para a anomalia i a partir do escore fatorial j .

A função de desempenho dada pela equação 4.1 definida anteriormente, é capaz de diagnosticar quando uma das anomalias descritas no quadro 4.4.1 podem estar ocorrendo, ou na eminência de ocorrer, para tanto, basta verificar os resultados apresentados pelas

componentes quando aplicados nelas os escores fatoriais obtidos a partir do vetor de leituras mensais padronizado.

4.5 Validação do Modelo Matemático

Nessa seção é proposto um teste para verificar a eficiência da função de desempenho 4.1 apresentada na seção anterior. Para tanto são gerada cinco matrizes $M_{500 \times 38}^i$ onde cada matriz apresenta 500 vetores com dados sintéticos simulando uma determinada anomalia i . O método de geração dos dados sintéticos segue os critérios apresentados no capítulo 3 seção 1, onde leva-se em consideração os instrumentos relacionados com cada anomalia, seus intervalos de mudança de comportamento e o instrumento de maior importância no fator.

Utiliza-se também nesse processo de validação uma matriz $R_{17 \times 38}$ com dados reais. A dimensão dessa matriz não é igual as demais matrizes pelo fato de depender das leituras realizadas dos instrumentos a partir de novembro do ano de 2016 até março de 2018.

Pretende-se verificar a porcentagem de acertos do modelo matemático, para isso, é realizado o cálculo dos escores fatoriais para cada vetor de cada matriz que simule uma determinada anomalia. E, após, aplica-se os escores correspondentes na função de desempenho 4.1, verificando quais componentes dessa função apresentam valores menores que zero, o que caracteriza uma determinada anomalia. No caso da matriz com dados reais, é verificado se o modelo acusou alguma anomalia em algum determinado mês, caso isso ocorra, é verificado se os instrumentos correspondentes àquela anomalia apresentam dados em seus respectivos intervalos de mudança de comportamento, isso justificaria o fato de tal anomalia ser detectada. Utiliza-se também uma matriz $R_{17 \times 38}$ com dados reais onde algumas linhas foram modificadas afim de provocarem uma anomalia.

O quadro 4.5 apresentado na próxima página, representa os resultados da validação com dados sintéticos.

Na primeira linha do quadro 4.5 estão os resultados referentes a anomalia Escorregamento na Brecha D. Observa-se que após 500 simulações a função de desempenho diagnosticou 493 vezes essa anomalia, no entanto, em 7 vezes a função não conseguiu monitorá-la e indicou valores maiores que zero para a componente G_i . Observa-se também que em 342 e 277 vezes a função de desempenho indicou a ocorrência de outras anomalias, respectivamente para $j = 2$ e $j = 3$. No caso de $j = 2$ foi indicado uma Subpressão na Brecha D. Esse fato não se caracteriza um erro, pois de acordo com a literatura, para que um escorregamento ocorra em uma determinada região é necessário que ocorra uma subpressão nessa mesma região, além disso, os instrumentos responsáveis no monitoramento de subpressões, também são responsáveis em monitorar um escorregamento. No caso de $j = 3$, a função de desempenho indicou a ocorrência de um escorregamento na Junta D, isso pode ser justificado pelo fato dos instrumentos conhecidos como bases de alongâmetro serem responsáveis em monitorar tanto o escorregamento na Brecha D quanto o escorregamento na Junta D, a única alteração que ocorre

entre essas anomalias são os diferentes piezômetros que são utilizados nesse monitoramento. Portanto, pode-se considerar normal quando a função de desempenho acusar essas duas anomalias quando um escorregamento na Brecha D ocorrer. Para esse caso obteve-se 98,6% de acertos.

Validação do Modelo Matemático com Dados Sintéticos						
Anomalias	Simulação	$G_i < 0$	$G_i > 0$	$G_j < 0$ $j \neq i$	% de Erros	% de Acertos
Escorregamento Brecha D	500	493	7	j=2; 342 j=3; 277	1,4	98,6
Subpressão Brecha D	500	497	3	j=1; 340 j=3; 327	0,6	99,4
Escorregamento Junta D	500	497	3	j=1; 476 j=2; 500	0,6	99,4
Fluência	500	493	7	j=7; 5	1,4	98,6
Tombamento	500	320	180	j=1; 59 j=2; 165 j=6; 275 j=7; 3	36	64
Recalque e Distensão Pé Montante na Junta D	500	498	2	j=1; 98 j=2; 168 j=7; 164	0,4	99,6
Recalque e Distensão Pé Montante na Brecha D	500	496	4	j=6; 161	0,8	99,2

Quadro 4.5: Resultados da Validação com Dados Sintéticos.

A segunda linha do quadro 4.5, apresenta os resultados referentes a Subpressão na Brecha D, para esse caso a função de desempenho conseguiu diagnosticar 497 vezes a ocorrência dessa anomalia, e em apenas 3 vezes ela não conseguiu identificá-la, isso garantiu um percentual de acerto de 99,4%. Observa-se também que em 340 e 327 vezes a função de desempenho indicou a ocorrência de outras anomalias, respectivamente para $j = 1$ e $j = 3$. No caso de $j = 1$ foi indicado um Escorregamento na Brecha D. Esse fato não se caracteriza um erro, pois de acordo com a literatura, para que uma subpressão ocorra em uma determinada região é necessário que ocorra um escorregamento nessa mesma região, além disso, os instrumentos responsáveis no monitoramento de subpressões nessa região, também são responsáveis em monitorar um escorregamento. No caso de $j = 3$, a função de desempenho indicou a ocorrência de um escorregamento na Junta D, isso pode ser justificado pelo fato dos instrumentos conhecidos como bases de alongâmetro serem responsáveis em monitorar tanto a Subpressão na Brecha D quanto o escorregamento na Junta D, a única alteração que ocorre entre essas anomalias são os diferentes piezômetros que são utilizados nesse monitoramento.

Na terceira linha estão os resultados referentes a um Escorregamento na Junta D. Nesse caso a função de desempenho conseguiu diagnosticar em 497 vezes a ocorrência dessa anomalia, e em apenas 3 vezes ela não conseguiu identificá-la, isso garantiu um percentual de acerto de 99,4%. Observa-se também que em 476 e 500 vezes a função de desempenho indicou

a ocorrência de outras anomalias, respectivamente para $j = 1$ e $j = 2$. No caso de $j = 1$ foi indicado um Escorregamento na Brecha D, pois alguns instrumentos são responsáveis também pelo monitoramento do Escorregamento na Junta D. No caso de $j = 2$ foi indicado uma Subpressão na Brecha D, isso pode ser justificado pelo fato dos instrumentos associados ao fator 1 serem responsáveis em monitorar tanto a Subpressão na Brecha D como também auxiliam no monitoramento do Escorregamento na Junta D.

Na quarta linha estão os resultados referentes a Fluência. A função de desempenho conseguiu diagnosticar em 493 vezes a ocorrência dessa anomalia, mas em 12 vezes não identificou essa ocorrência. Para esse caso obteve-se 98,6% de acertos. Observa-se também que em 5 vezes a função de desempenho indicou a ocorrência de outra anomalia, no caso $j = 7$, que representa a anomalia Recalque ou Distensão Pé Montante na Brecha D, pode ser justificado pelo fato dos instrumentos como extensômetros e piezômetros serem responsáveis em monitorar tanto a Fluência como também auxiliam no monitoramento do Recalque ou Distensão.

Na quinta linha estão os resultados referentes a um Tombamento. Nesse caso, a função de desempenho conseguiu diagnosticar em 320 vezes a ocorrência dessa anomalia, mas em 180 vezes não identificou essa ocorrência. Para esse caso obteve-se 64% de acertos. Observa-se também que em 59, 165, 275 e 3 vezes a função de desempenho indicou a ocorrência de outras anomalias, respectivamente para $j = 1, j = 2, j = 6$ e $j = 7$. Para que um tombamento ocorra é necessário que varias outras anomalias ocorram, como escorregamentos e subpressão, por esse motivo o fato da função acusar ambas anomalias é justificado. No caso de $j = 6$ e $j = 7$ que representa a anomalia Recalque ou Distensão Pé Montante na Brecha D e Recalque ou Distensão Pé Montante na Junta D, respectivamente, pode ser justificado pelo fato dos instrumentos como extensômetros e piezômetros serem responsáveis em monitorar tanto o tombamento como também auxiliam no monitoramento do Recalque ou Distensão.

Na sexta linha estão os resultados referentes ao Recalque ou Distensão Pé Montante na Junta D. Nesse caso, a função de desempenho conseguiu diagnosticar em 498 vezes a ocorrência dessa anomalia, mas em 2 vezes não identificou essa ocorrência. Para esse caso obteve-se 99,6% de acertos. Observa-se que em 98, 168 e 164 vezes a função de desempenho indicou a ocorrência de outras anomalias, respectivamente para $j = 1, j = 2$ e $j = 7$. No caso de $j = 1$, que representa a anomalia Escorregamento na Brecha D, isso pode ser justificado pelo fato dos instrumentos associados ao fator 1 serem responsáveis em monitorar tanto o Escorregamento na Brecha D como também auxiliam no monitoramento do Recalque ou Distensão Pé Montante na Junta D. No caso de $j = 2$ foi indicado uma Subpressão na Brecha D, isso pode ser justificado, também, pelo fato dos instrumentos associados ao fator 1 serem responsáveis em monitorar tanto a Subpressão na Brecha D como também auxiliam no monitoramento da Distensão Pé Montante na Junta D. No caso de $j = 7$, que representa o Recalque ou Distensão Pé Montante na Brecha D, esse fato pode ser justificado pelo fato dos extensômetros envolvidos serem os mesmos, a diferença é em que parte da fundação estão os sensores, e por esse motivo

a mudança de comportamento das leituras dos extensômetros da instalados na Junta D também influenciam os extensômetros localizados na Brecha D.

Na sétima linha estão os resultados referentes a Recalque ou Distensão Pé Montante Brecha D. Nesse caso, a função de desempenho conseguiu diagnosticar em 496 vezes a ocorrência dessa anomalia, mas em 4 vezes não identificou essa ocorrência. Para esse caso obteve-se 99,2% de acertos. Observa-se também que em 161 vezes a função de desempenho indicou a ocorrência da anomalia Recalque ou Distensão Pé Montante na Junta D, isso pode ser justificado pelo fato dos extensômetros envolvidos serem os mesmos, a diferença é em que parte da fundação estão os sensores, e por esse motivo a mudança de comportamento das leituras dos extensômetros da instalados na Junta D também influenciam os extensômetros localizados na Brecha D.

Portanto, para validação do modelo matemático a partir de dados sintéticos, os resultados são considerados satisfatórios, pois em apenas um caso obteve-se um desempenho inferior a 70%.

O quadro que segue apresenta os resultados referentes aos dados reais, que foram coletados e padronizados a partir de Novembro de 2016 até Março de 2018. O objetivo dessa análise é verificar se nesse período houve a ocorrência de alguma anomalia possível de ser monitorada pela função de desempenho 4.1 e, se essa função foi capaz de identificá-la.

Validação do Modelo Matemático com Dados Reais		
$M_{17 \times 38}$	$G_i < 0$	$G_i > 0$
G1	0	17
G2	0	17
G3	0	17
G4	0	17
G5	0	17
G6	0	17
G7	0	17
G8	0	17
G9	0	17

Quadro 4.6: Resultados da Validação com Dados Reais.

Ao analisar o quadro acima observa-se que no período de Novembro de 2016 até Março de 2018 a função de desempenho 4.1 não constatou a ocorrência de nenhuma das anomalias apresentadas na figura 4.4.

Cada linha da matriz $R_{17 \times 38}$ corresponde a um vetor que apresenta o comportamento do bloco-chave em um determinado mês. Ficou constatado que não ocorreu, em nenhum mês, a combinação de leituras fora da normalidade, o que justifica os resultados apresentados pelo modelo.

Para confirmar se a função de desempenho 4.1 é capaz de realizar o monitoramento do bloco-chave A-7 é proposto um novo teste. Escolhe-se nove meses de forma aleatória na matriz de dados reais e modifica-os de forma que cada mês apresente uma determinada anomalia. Isso é feito alterando as leituras dos instrumentos relacionados com cada anomalia, após isso, é realizada uma padronização dessa nova matriz, para que a partir dela possa ser calculado os escores fatoriais.

Dessa forma pode-se aplicar esses escores na função de desempenho 4.1, afim de verificar se o modelo foi capaz de identificar essas alterações nos dados. Foram selecionados os seguintes meses, Novembro de 2016, Janeiro de 2017, Março de 2017, Maio de 2017, Julho de 2017, Setembro de 2017, Outubro de 2017, Novembro de 2017 e Dezembro de 2017, e, em cada mês foram adicionadas respectivamente, as seguintes anomalias: Escorregamento Brecha D, Subpressão Brecha D, Escorregamento Junta D, Fluência, Tombamento, Recalque Junta D, Distensão Pé Montante Junta D, Recalque Brecha D e Distensão Pé Montante Brecha D.

O quadro abaixo apresenta os resultados obtidos após aplicação dos escores no modelo matemático.

Validação do Modelo Matemático com Dados Reais Modificados		
$M_{17 \times 38}$	$G_i < 0$	$G_i > 0$
G1	2	15
G2	1	16
G3	1	16
G4	1	16
G5	1	16
G6	2	15
G7	1	16
G8	2	15
G9	1	16

Quadro 4.7: Resultados da Validação com Dados Reais Modificados.

O quadro acima apresenta no período de 17 meses, quantas vezes cada componente da função de desempenho apresentou valores menores que zero e, além disso, também informa quantas vezes as componentes da função de desempenho apresentaram valores maiores que zero.

A primeira anomalia do quadro 4.7 é o Escorregamento na Brecha D, pode-se perceber que o algoritmo detectou dois valores menores que zero para essa anomalia, de acordo com a literatura, esse fato é justificado, pois para que um escorregamento ocorra em uma determinada região é necessário que ocorra uma subpressão nessa mesma região, além disso, os instrumentos responsáveis no monitoramento de subpressões, também são responsáveis em monitorar um escorregamento.

Merecem destaque as componentes G_6 e G_8 que apresentaram valores menores que zero em dois meses. Isso se deve pelo fato de que o mês de Setembro de 2017 já apresentava essa

anomalia (Recalque na Junta D), que corresponde a linha 6 da matriz $R_{17 \times 38}$, na linha 8 foi adicionada a anomalia Recalque na Brecha D e como visto anteriormente, nesses casos a função de desempenho também pode constatar um Recalque na Brecha D, pelo fato dos extensômetros envolvidos serem os mesmos, a diferença é em que parte da fundação estão os sensores, e por esse motivo a mudança de comportamento das leituras dos extensômetros instalados na Junta D também influenciam os extensômetros localizados na Brecha D.

Portanto, pode-se constatar a eficiência do modelo quando aplicado os dados reais modificados.

Capítulo 5

Conclusão

Nesse trabalho foi utilizado um modelo com um método alternativo para o monitoramento de anomalias em um bloco de concreto altamente instrumentado. Por ser um método estatístico, sua eficiência depende da qualidade e padronização dos dados e das comprovações de algumas hipóteses, que essencialmente são a confirmação de todos os testes estatísticos apresentados no capítulo 2.

Para aplicação desse modelo, primeiramente coletou-se as leituras de todos os instrumentos instalados no bloco-chave A-7 da Barragem de Itaipu, com o propósito de padronizar todos esses dados. A partir da matriz multivariada gerada por essas leituras padronizadas, realizou-se a análise fatorial, que gerou grupos de instrumentos altamente correlacionados.

Notou-se que apenas um dos cinco escores fatoriais possuía a condição de normalidade e, até o momento, não foi encontrado nenhuma transformação nos dados capaz de transformá-los em variáveis aleatórias com distribuição normais. Como sugestão para trabalhos futuros, pretende-se encontrar uma forma de calcular a estimativa de probabilidade de falha, para quando os escores fatoriais não possuem distribuição de probabilidade normal.

Pode-se considerar como contribuições científicas apresentadas neste trabalho, as demonstrações matemática dos principais resultados da análise multivariada, análise de componentes principais e análise fatorial, que foram as ferramentas fundamentais para a aplicação da metodologia proposta. Destaca-se também que, com as informações obtidas para esse bloco-chave, foi possível definir um sinal de alerta para tornar o monitoramento mais parecido com o que é atualmente utilizado pelos profissionais em Segurança de Barragens da UHE de Itaipu e, além disso, para utilizar o máximo possível as informações fornecidas pela AF, foi criada as semi componentes da função de desempenho, que são anomalias com informações incompletas de alguns instrumentos, mas que permitem um diagnostico de seus estados mesmo que parciais.

Finalmente, conclui-se que o objetivo inicial do trabalho foi alcançado, visto que, com aplicação do modelo proposto, foi possível obter uma lei matemática através da função de desempenho multivariada capaz de descrever o comportamento do bloco-chave, mostrando que o mesmo pode ser considerado como um sistema dinâmico. Verificou-se também que blocos-chaves diferentes, implicam em funções de desempenho diferentes, isso se deve as características de cada bloco, uma vez que os blocos-chaves da barragem de concreto podem possuir instrumentos e quantidades diferentes, ocasionando em matrizes multivariadas diferentes, que são os principais objetos para aplicação do modelo.

Referências Bibliográficas

- Aranha, F., Zambaldi, F. (2008). *Análise Fatorial em Administração*. São Paulo: Cengage Learning.
- Baecher, G. B., Christian, J. T. (2003). *Reliability and Statistics in Geotechnical Engineering*. Massachusetts: Wiley.
- Belato, D. (2002). *Análise Não Linear de Sistemas Dinâmicos Holônômos Não Ideais*. Tese de Doutorado, Universidade Estadual de Campinas, Campinas.
- Biase, A. G., Ferreira, D. F. (2012). *Teste Computacionalmente Intensivo Baseado na Distância de Mahalanobis para Normalidade Multivariada*. Bras. Biom. (30), pp.1-22.
- Binacional. (2003). *Itaipu- Usina Hidrelétrica- Aspectos de Engenharia*. Foz do Igauçu: TAB Marketing Editorial.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. New York: IE-Wiley.
- Corrar, L. J., Paulo E. & Dias, J. F. (2009). *Análise Multivariada: para os cursos de Administração, Ciências Contábeis e Economia*. São Paulo: Atlas.
- Eletrobrás. (2003). *Critérios de Projeto Civil de Usinas Hidrelétricas*. Rio de Janeiro.
- Filho, J. G. (2006). *Desempenho de Obras Geotécnicas*. Recife: Editora Universitária UFPE.
- Gerais, U. F. de M. (2017). *Sistemas Dinâmicos*. Acesso em 16 de Abril de 2018, disponível em: www.mat.ufmg.br/site/pesquisa/linhasdepesquisas/sistemas-dinamicos/>.
- Jaime, G. F. (2006). *Desempenho de Obras Geotécnicas*. Recife: Editora Universitária UFPE.
- Johnson, R. A., Wichern, D. W. (1987). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Kaiser, H. (1959). *The varimax criterion for analytic Rotation in Factor Analysis*. Amsterdã: Elsevier.
- Krüger, C. M. (2008). *Análise de Confiabilidade Estrutural Aplicada às Barragens de Concreto*. Tese de Doutorado, Universidade Federal do Paraná, Curitiba.
- Lattin, J. D., Carroll, P. E. (2011). *Análise de Dados Multivariados*. São Paulo: Cengage Learning.

- Leonards, G. A. (1975) *Discussion on Differencial Settlement of Buildings*. Journal of Geotechnical Engineering Division (101), pp. 700-702.
- Liliefors, H. (1967). *On the Kolmogorov-smirnov test for normality with mean and variance unknown*. American Statistical Association Stable, pp. 399-402.
- Matemática, L. (2017). *Sistemas Dinâmicos e Teoria Ergódica*. Acesso em 05 de Fevereiro de 2018, disponível em: <http://conteudo.icmc.usp.br/Portal/Pesquisa/pesquisaDinamico.php?id_laboratorio=80/>
- Mello, L. A. (2017). *Análise de Tempos e Métodos no Setor de Embalagem em uma Empresa produtora de Erva-Mate*. Engevista (2), pp. 1198-1212.
- Mendes, F. T. (2010). *Probabilidade para Engenharias*. Rio de Janeiro: LTC.
- Mendes, M., Pala, A. (2003). *Type in Error Rate and Power oh Three Normality Tests*. Journal of Information and Technologu (2), pp. 135-139.
- Mingoti, S. A. (2007). *Análise de Dados Através de Métodos de Estatística Multivariada*. Belo Horizonte: Livraria da Física.
- Monteiro, L. H. A. (2006). *Sistemas Dinâmicos*. São Paulo: Livraria da Física.
- Nieradka, I. P. (2016). *Sistema de Monitoramento da Instrumentação de Segurança: Um Estudo de Caso Considerando a Usina de Itaipu*. Dissertação de Mestrado. Universidade Estadual do Oeste do Paraná. Foz do Iguaçu, Paraná.
- Oliveira, A. A., Pessoa, M. C. (2015). *Aplicação de um Modelo de Otimização para a Distribuição de Estudantes em Escolas Públicas: Um Estudo de Caso em um Distrito no Município de São Gonçalo*. Engevista (17), pp. 540-553.
- Pinto, J. A. (2007). *Uso da Simulação de Sistemas Dinâmicos para Análise de Logística Hospitalar: Gestão Integrada de Estoque e Capacidade*. Dissertação de Mestrado. Universidade Estadual de Campinas. Campinas, São Paulo.
- Razali, N. M., Wah, Y. B. (2011). *Power Comparisons of Shapiro-Wilk, Kolmorogov-Smirnov, Lilliefors and Anderson-DARling Tests*. Journal of Statistical Modeling and Analitycs (2), pp. 21-33.
- Santos, S. R. (2012) *Análise de Confiabilidade Estrutural Utilizando o Método de Monte Carlo e Redes Neurais*. Tese de Doutorado. Universidade Federal do Paraná. Curitiba, Paraná.
- Silva, O. C. (2016). *Análise Multivariada e Confiabilidade Estrutural no Estudo de Riscos em Blocos de Concreto de Barragens Altamente Instrumentadas*. Tese de Doutorado. Universidade Federal do Paraná. Curitiba, Paraná.

Silveira, J. F. (2003). *Instrumentação e Comportamento de Fundações de Barragens de*. São Paulo: Oficina de Textos.

Steven, J. (2011). *Álgebra Linear com Aplicações*. Rio de Janeiro: LTC.

Zanardi, M. S., Turetta, L. F., Costa, A. O., Costa, E. F. (2016). *Simulação de Secagem da Soja com Coeficiente de Difusão Variável*. *Engevista* (18), pp. 207-224

Apêndice A

Artigo submetido e aceito para ser apresentado no Dam World 2018 em 17 à 21 de Setembro de 2018.

DETERMINAÇÃO DE MODOS DE FALHA EM UMA BARRAGEM DE CONCRETO ALTAMENTE INSTRUMENTADA

A. Chaves*, C. Rocha*, O. C. Silva* and I. P. Nieradka¹

*Universidade Estadual do Oeste do Paraná (UNIOESTE)
Av. Tarquínio Joslin dos Santos, 1300 - Lot. Universitario das Americas, Foz do Iguaçu - PR
e-mail: alinechavesrs@yahoo.com.br

Palavras-Chave: Segurança de Barragens; Blocos-chaves; Análise Multivariada.

Resumo. *A segurança de barragens é um assunto que ganhou destaque nos últimos anos devido a acidentes ocorridos pelos mais diversos motivos. Levando em consideração a existência de normas que visam o controle da segurança de barragens no país, neste trabalho é proposto um estudo da variabilidade dos blocos-chaves de uma barragem de concreto altamente instrumentada através das leituras dos instrumentos instalados nos blocos, com a aplicação de uma metodologia capaz de determinar os modos de falha possíveis de serem monitorados, utilizando apenas as variáveis aleatórias geradas das leituras dos instrumentos.*

O procedimento descrito nesse trabalho fornece uma alternativa para determinação de modos de falha que pode servir como uma ferramenta para o monitoramento e, auxiliar na tomada de decisão. Com a aplicação desse procedimento, é possível identificar com antecedência anomalias na estrutura. Por ser um método estatístico, sua eficiência depende da qualidade da coleta e padronização dos dados e da comprovação de algumas hipóteses, que essencialmente são a confirmação de testes estatísticos necessários para aplicação da Análise Fatorial.

1 INTRODUÇÃO

A segurança das barragens constitui uma preocupação permanente para as entidades governamentais, tanto por sua importância econômica específica como pelo risco potencial que representa a possibilidade de ruptura ou outro acidente grave, em termos de vidas humanas, impacto ambiental, prejuízos materiais e os reflexos econômico-financeiros [1].

Nos últimos anos, acidentes ocasionados por barragens tiveram destaque na mídia nacional, resultando em danos ambientais e financeiros jamais vistos no país. No Brasil, a lei apresentada em [2] estabelece a Política Nacional de Segurança de Barragens destinada à acumulação de água para quaisquer usos, e cria o Sistema Nacional de Informações sobre Segurança de Barragens. Em [3] são estabelecidas diretrizes para a implantação da Política Nacional de Segurança de Barragens, para a aplicação de seus instrumentos e atuação do Sistema Nacional de Informação sobre Segurança de Barragens. Ambas as leis tem como finalidade reduzir para números extremamente baixos o risco de eventuais acidentes em barragens brasileiras.

* Universidade Estadual do Oeste do Paraná (UNIOESTE)

¹ Instituto Federal do Paraná (IFPR)

Monitorar uma barragem requer métodos tecnológicos que permitem uma avaliação mais precisa do comportamento de uma estrutura, para tanto, vários instrumentos devem ser instalados de tal maneira que permitam leituras que avalie se a estrutura está em boas condições.

Diante de normas que visam o controle da segurança de barragens, neste artigo é utilizada uma metodologia desenvolvida e apresentada em [4] que é capaz de determinar modos de falha, ocasionados por algum comportamento atípico da estrutura, em um bloco-chave de concreto de uma barragem altamente instrumentada.

Blocos-chaves são seções representativas do comportamento típico de um determinado trecho da barragem, levando-se em consideração, basicamente, o tipo de fundação e as características geométricas da estrutura, que recebem uma instrumentação mais completa.

Anomalias em uma estrutura nada mais são que movimentos não admissíveis, que podem ocasionar danos mais graves como riscos de acidentes [5]. Os instrumentos destinados a garantir a segurança de uma barragem, são capazes de identificar movimentos não admissíveis, fornecendo leituras que comprovem esse comportamento.

Nas próximas seções são apresentadas as fundamentações teóricas utilizadas para a realização deste trabalho, uma análise descritiva e quantitativa dos resultados que foram obtidos com a aplicação da metodologia, e por fim as conclusões.

2 FUNDAMENTAÇÃO TEÓRICA

A barragem tem esforços de peso próprio e de água represada, que geram deformação e, em consequência, um movimento. Este movimento pode ser admissível ou não, conforme o tipo de barragem e se está bem projetada ou não [5].

A Figura 1 apresenta uma correlação entre os tipos de instrumentos usualmente empregados na auscultação de barragens, conjuntamente com as inspeções visuais e os principais tipos de deterioração [6].

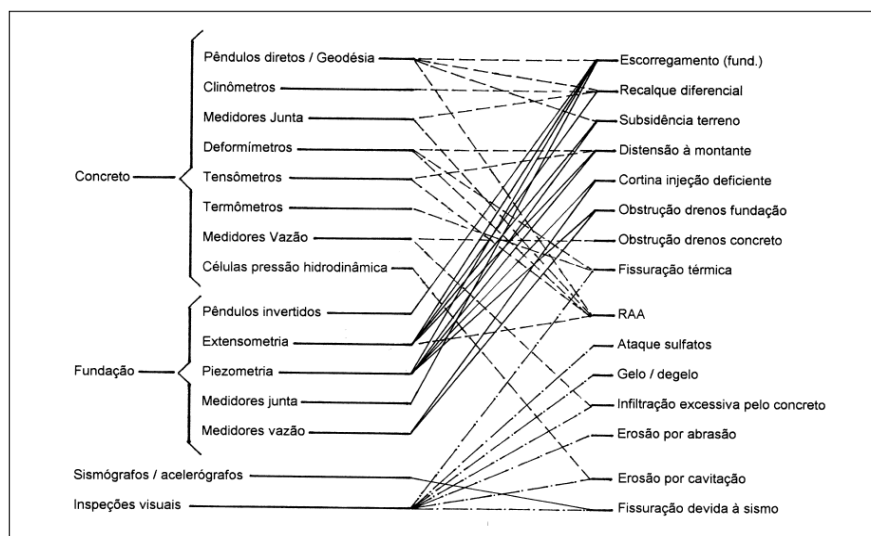


Figura 1: Instrumentação de barragens de concreto

Com a aplicação da Análise Fatorial são determinados os grupos de variáveis aleatórias que estão altamente correlacionadas, estabelecendo a relação entre instrumentos e uma mesma anomalia.

A determinação dos modos de falha de uma estrutura de concreto altamente instrumentada foi um problema abordado no trabalho apresentado em [7], no qual foi discutido o conceito de segurança de barragens, apresentando os instrumentos mais utilizados no mundo todo para garantir a segurança de uma barragem.

Com a aplicação de uma técnica da Análise Multivariada conhecida como Análise Fatorial, foi possível determinar grupos de instrumentos que estão altamente correlacionados. Esses grupos são conhecidos como fatores e através deles é possível gerar os escores fatoriais, que são as variáveis aleatórias que substituem as variáveis aleatórias iniciais (leituras dos instrumentos), simplificando a análise.

Com o conhecimento adquirido após a Análise Fatorial, e das correlações dos instrumentos com possíveis anomalias, foi desenvolvido um método capaz de determinar as regiões críticas de cada modo de falha.

Nesse contexto, foi considerada como "falha na estrutura" uma mudança brusca em seu comportamento, não sendo necessariamente uma ação catastrófica, mas sim um alerta de que algo de anormal esteja ocorrendo, e que, se não for reparado, tal anomalia poderá ocorrer.

Em [8] foi apresentado um algoritmo para a determinação das regiões críticas de cada modo de falha. A seguir é apresentado um resumo do procedimento para a confecção deste algoritmo.

A aplicação da metodologia foi na barragem da Usina de Itaipu, e como a Análise Fatorial agrupa os instrumentos altamente correlacionados a um mesmo fator, a ideia é utilizar as associações dos instrumentos com as anomalias apresentadas em [5], [6] e [9] e o conhecimento prático dos profissionais que atuam na barragem da Usina de Itaipu, para relacionar os fatores com as anomalias.

De acordo com as literaturas citadas e do que foi repassado por profissionais que têm experiência em barragens, para que certas anomalias ocorram, é necessária que exista simultaneamente uma combinação de leituras fora do comportamento normal.

Como os escores fatoriais dependem dos fatores obtidos na Análise Fatorial e, conseqüentemente, dependem dos instrumentos que estão agrupados neles, quando houver alguma mudança brusca nas leituras desses instrumentos correlacionados com uma determinada anomalia, o resultado do escore fatorial também deverá sofrer alteração. Portanto, para realizar esse monitoramento busca-se determinar as regiões críticas desses escores fatoriais, que indicarão quais as anomalias ocorreram.

Para determinar as regiões críticas, são utilizados os intervalos de projeto dos instrumentos instalados no bloco e os intervalos de 99% de confiança da média dos últimos dois anos, e, determina-se um intervalo que representa uma mudança de comportamento do instrumento k (MC_k), obtendo-se a equação:

$$MC_k = (\min\{l_i, L_i\} - k_1s, \max\{l_i, L_i\} - k_1s) \cup (\min\{l_s, L_s\} + k_2s, \max\{l_s, L_s\} + k_2s) \quad (1)$$

Na equação (1), l_i e l_s são os limites inferiores e superiores do intervalo de confiança de 99% da média, respectivamente, L_i e L_s são os limites inferiores e superiores do intervalo de projeto, s é o desvio padrão amostral das leituras do instrumento correspondente aos últimos cinco anos, e, $\{k_1, k_2\}$ são constantes, maiores ou iguais a zero, que garantam que o intervalo de mudança de comportamento possa gerar valores extremos já obtidos pelas leituras de cada instrumento, mesmo que esses valores sejam raros de ocorrer.

Não existe uma regra geral para a determinação dos valores de $\{k_1, k_2\}$, pois os mesmos dependem da característica de cada instrumento, de sua sazonalidade e principalmente do nível de sensibilidade que o proprietário da barragem desejar.

Usando um programa gerador de números aleatórios, é possível gerar vetores de simulação das anomalias, que são possíveis de monitorar a partir do que foi apresentado pela Análise Fatorial.

É importante destacar, que essa geração por mais que seja aleatória deve seguir o princípio de que se existirem três instrumentos altamente relacionados com essa anomalia, então antes é necessário verificar qual é o instrumento com maior influência no fator. Esta informação é

revelada pela Análise Fatorial. Além disso, a Análise Fatorial também revela qual é a correlação entre todos os instrumentos nesse estudo.

Considere certa anomalia i , e suponha que existam m instrumentos, associados com a anomalia i . Considere também que desses m instrumentos, um número $m' < m$ esteja altamente correlacionado com um fator j . A Figura 2, representada a seguir, apresenta um fluxograma deste algoritmo. Uma descrição dos passos necessários para a execução deste algoritmo pode ser encontrada, mais detalhadamente em [8].

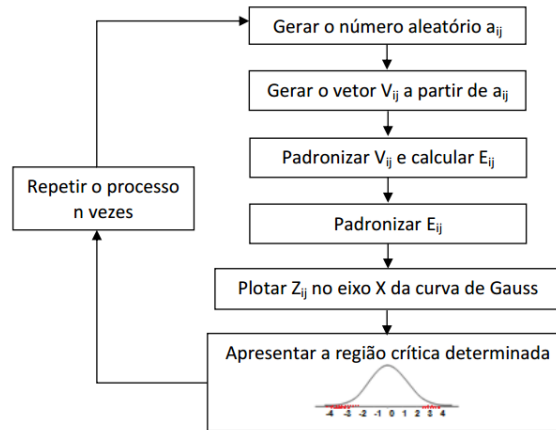


Figura 2: Fluxograma do algoritmo para determinação dos modos de falha

Na Figura 2, j representa uma variável aleatória normal, o valor z_{ij} foi obtido simulando uma situação atípica. Se for realizada essa simulação para um número suficientemente grande de vezes, então é obtido um número suficientemente grande de pontos no eixo x , formando uma região em uma ou ambas as caldas da curva de Gauss, assim é utilizado essas regiões para determinar a região crítica da anomalia i , monitorada a partir do fator j .

Uma vez que são coletadas as leituras dos instrumentos instalados em algum bloco-chave de uma barragem de concreto, feita as transformações necessárias de tal forma que cada instrumento possua apenas um representante mensal, e, calculado os escores fatoriais que sejam capazes de monitorar as anomalias, pode-se verificar se o valor desses escores fatoriais pertencem à uma determinada região crítica dessas anomalias. Caso ocorra, é dito que existem indícios estatísticos para acreditar que essa anomalia pode estar ocorrendo.

3 RESULTADOS OBTIDOS

Nesta seção são apresentados os resultados obtidos com a aplicação do método descrito na seção anterior, no bloco-chave A-7 da Barragem de Itaipu, e é apresentada uma análise destes resultados. Para a aplicação foram utilizados os softwares MatLab e Statistica.

Considerando as leituras geradas pelos instrumentos apresentados no quadro representado através da Figura 3, incluindo as leituras de outros dois instrumentos (não instalados no bloco A-7), que medem a temperatura ambiente e o nível do reservatório que influenciam todo o sistema, é estabelecida uma matriz $M_{202 \times 40}$ multivariada. Foi considerado um período de tempo a partir de janeiro de 2000 até outubro de 2016, correspondendo a 202 meses. As leituras no geral não estavam padronizadas e para possibilitar a análise, os dados mensais foram padronizados. Para os instrumentos que possuíam mais de uma leitura mensal, foram calculadas as médias. No entanto, alguns instrumentos possuíam apenas leituras trimestrais, como o caso dos Termômetros (TI). Neste caso, foi aplicado a técnica de interpolação Spline Cúbica.

Instrumentos Bloco-Chave A-7			
Instrumentos	Quantidades	Unidade de Leitura	Função
Pêndulo (PD)	2	mm	Mede os deslocamentos na direção da jusante e de montante.
Extensômetro (EM)	6	mm	Mede o deslocamento e deformações que ocorrerão entre pontos no maciço da fundação.
Base de Alongâmetro (JS)	4	mm	Medidor superficial de Junta para o Concreto.
Piezômetro Elétrico (PG)	1	msnm	Mede a Subpressão dos poros no núcleo e no contato núcleo-fundação.
Tensômetro de Concreto (TN)	1	kgf/cm ²	Para medição direta das tensões de compressão no concreto.
Termômetro Superficial (TS)	1	°C	Termômetro embutidos no concreto para medição da temperatura durante a construção e operação.
Medidor de Vasão	1	l/s	Para medição de vasão de percolação pelas estruturas e suas fundações.
Piezômetro Standpipe (PS)	7	msnm	Piezômetros para medir a pressão dos poros e a subpressão hidráulica em juntas ou contatos escolhidos no maciço rochoso ou no contato barragem-fundação.
Termômetro Interno (TI)	4	°C	Termômetros embutidos no concreto para medição da temperatura durante a construção e operação.
Roseta Deformação (RD)	11	kgf/cm ²	Ebutidas no concreto para medição das variações de volume autógenas, fluência e deformação elástica e temperatura.

Figura 3: Instrumentos instalados no bloco-chave A-7 do vertedouro

Informações sobre os instrumentos instalados na barragem da Usina de Itaipu e as funções que desempenham na estrutura são encontradas em [10]. A Figura 3 apresenta uma breve descrição dos instrumentos que foram utilizados nesse estudo.

Antes de aplicar a Análise Fatorial na matriz $M_{202 \times 40}$, foi verificado alguns testes estatísticos, indispensáveis para aplicação do método.

Assim, inicialmente foi testada a normalidade multivariada da matriz $M_{202 \times 40}$. Foi utilizado o teste computacionalmente intensivo baseado na distância de Mahalanobis [11]. A hipótese nula desse teste afirma:

H_0 : A matriz de dados amostrais é proveniente de uma população normal multivariada.

A aplicação do teste resultou em $p = 0.7552$ (obtido com 100.000 simulações Monte Carlo). Portanto, ao nível de significância de 0,05, a hipótese de normalidade multivariada não pode ser descartada.

Sabendo-se da normalidade multivariada dos dados, aplicou-se o Teste de Esfericidade de Bartlett para testar as hipóteses:

H_0 : A matriz de correlação é igual a matriz identidade, ou seja, $\rho = I$;

H_1 : A matriz de correlação é diferente da matriz identidade, ou seja, $\rho \neq I$.

A aplicação do teste resultou em $X^2 = 7988,4$, $p = 0,9920$, ou seja, rejeitou-se a hipótese H_0 . Portanto, a amostra multivariada é proveniente de uma população com matriz de correlação significativamente diferente da matriz identidade.

O cálculo da medida de adequabilidade dos dados para realização da Análise Fatorial resultou em $KMO = 0,8349$. Como o valor mínimo exigido é 0,50, a amostra multivariada satisfaz essa condição.

Comprovados os resultados dos testes, aplicou-se a Análise Fatorial na matriz $M_{202 \times 40}$. O método adotado para obtenção dos coeficientes fatoriais foi o Método das Componentes Principais com rotação Varimax, visto que esse método obteve uma maior porção da variância explicada, comparado com o método da Máxima Verossimilhança.

Aplicando o Critério de Kaiser do mínimo autovalor, foi obtido 5 fatores, que explicam 88,07% de toda variância, entretanto, duas variáveis apresentaram comunalidades baixas, menores que 0,70, e devido a isso, uma nova análise foi realizada removendo essas variáveis. A nova análise envolveu todos os testes estatísticos realizados anteriormente, mas agora em uma matriz $M_{202 \times 38}$ para reconfirmar adequação dos dados. Novamente todos os testes foram bem sucedidos.

Com a nova análise obteve-se uma variância explicada de 89,43%, sem nenhuma variável com baixa comunalidade.

Os escores fatoriais gerados a partir dos coeficientes fatoriais são utilizados para a determinação dos modos de falha. Neles é aplicado o teste de normalidade univariada conhecido como Teste de Lilliefors, para comprovar se são variáveis aleatórias com distribuição de probabilidade normais, caso contrário, será necessária à aplicação de transformações nos escores. Mais informações sobre o teste de Lilliefors podem ser encontrados em [8].

O quadro representado através da Figura 4 apresenta os resultados obtidos após os testes estatísticos e as transformações aplicadas.

Resultados obtidos após Teste Lilliefors			
Escores	Estimativas	Transformação	Decisão Teste
1	d= 0.17910, p<0.01	$\sqrt{ x }$	Não Normal
2	d=0.08531, p<0.15	$\sqrt{ x }$	Não Normal
3	d=0.10556, p<0.05	$\sqrt{ x }$	Não Normal
4	d=0.07156, p>0.20	Não	Normal
5	d=0.20658, p<0.01	$\sqrt{ x }$	Não Normal

Figura 4: Resultados obtidos após Teste Lilliefors

Analisando a Figura 4, observa-se que o escore fatorial 4 já possui a condição de normalidade. Já os escores 1, 2, 3 e 5 não são normais e não se encontrou até o momento uma transformação para eles. Se o objetivo desse trabalho fosse o cálculo da probabilidade de falha os mesmos não seriam utilizados no monitoramento da estrutura. No entanto, nesse artigo o objetivo é determinar os modos de falha possíveis de serem monitorados e suas regiões críticas, sem a preocupação da estimativa dessas probabilidades de falha e, por esse motivo eles puderam ser incluídos.

De acordo com a Análise Fatorial realizada foram gerados grupos de instrumentos que estão altamente correlacionados. O quadro representado através da Figura 5 apresenta esses grupos.

Assim, considerando o apresentado na literatura e o conhecimento prático dos profissionais em segurança de barragem da Usina de Itaipu, foi estabelecido ser possível com a determinação destes grupos, monitorar as seguintes anomalias: escorregamentos e subpressão, fluência e tombamento.

Grupos de Instrumentos Altamente Correlacionados				
Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
JS-A-13/desliz	JS-A-13/abert	EM-A-3/2	RD-A-1/tr1	PS-A-22/msnm
EM-A-3/1	JS-A-14/abert	CO-A-1/X Asoc	RD-A-1/tr2	RD-A-1/dp1
EM-A-4/1	JS-A-14/asent	CO-A-1/Y Asoc	RD-A-1/tr3	RD-A-1/dp2
EM-A-4/2	EM-A-5/1	PG-A-8/msnm	RD-A-1/tr4	RD-A-1/dp3
EM-A-5/2	TN-A-1	RD-A-1/dp5	RD-A-1/tr5	RD-A-1/dp4
PS-A-14/msnm	TI-A-1	TS-A-1	Nível do Reserv	RD-A-1/dp6
PS-A-17/msnm		Temp Amb		TI-A-3
PS-A-18/msnm				TI-A-4
PS-A-19/msnm				
PS-A-20/msnm				
PS-A-21/msnm				

Figura 5: Grupos de instrumentos altamente correlacionados

O escorregamento de um bloco em uma barragem de concreto é quando há um movimento de abertura e deslizamento entre os blocos vizinhos, fazendo com que ele se movimente para uma determinada direção.

A subpressão é um efeito causado pela pressão exercida pela água que percorre à fundação, aliviando o peso da barragem, contribuindo para o escorregamento da estrutura em uma determinada camada da fundação.

As deformações dependentes do tempo, são convencionalmente separadas em duas: a fluência e a retração. A fluência é o acréscimo contínuo das deformações que ocorre mesmo para uma tensão constante. A retração é a redução de volume do material na ausência de uma carga externa. Tanto a fluência, quanto a retração, diminuem com a redução do fator água-cimento e do consumo de cimento [12].

O tombamento é um mecanismo raro de ocorrer, o fato é que antes da ruptura por tombamento, desenvolvem-se esforços de tração e aumentos de subpressão na rocha à montante e aumentos das tensões de compressão à jusante, e com a existências das forças horizontais o resultado é a ruptura por deslizamento antes da ocorrência do tombamento [5].

O tombamento é um movimento quase que total da estrutura, e por esse motivo praticamente todos os instrumentos instalados são afetados quando essa anomalia ocorrer.

Para gerar o modelo para o caso aplicado é conveniente apresentar a associação dos fatores com as anomalias. O quadro representado através da Figura 6 enumera cada anomalia e apresenta essa associação.

Associação dos Fatores com as Anomalias		
Índice	Anomalia	Fatores Envolvidos
1	Escorregamento na Brecha D	1, 3
2	Subpressão na Brecha D	1
3	Escorregamento na Junta D	1, 2
4	Fluência	3, 4
5	Tombamento	1, 2, 3, 5

Figura 6: Associação dos fatores com as anomalias

De acordo com as informações obtidas através dos resultados, é possível monitorar as anomalias apresentadas através da representação da Figura 6. Para determinar as regiões críticas dessas anomalias foi utilizado o algoritmo representado pelo fluxograma apresentado na Figura 2. Para determinar cada região de falha foram realizadas simulações, e apenas os instrumentos

relacionados com uma determinada anomalia apresentaram valores dentro dos intervalos de Mudança de Comportamento, indicados por fatores que possuíam influência para aquela anomalia, determinando a região crítica correspondente.

Os limites de especificação de projeto utilizados para gerar os intervalos de mudança de comportamento para cada instrumento foram cedidos pelos profissionais de segurança de barragens de Itaipu. Os valores utilizados para as constantes k_1 e k_2 , que definem o nível de sensibilidade do sistema, são fornecidos pelo proprietário da barragem, e devem garantir que os vetores realmente simulam uma mudança de comportamento.

Dessa forma, as regiões críticas para as anomalias apresentadas através da representação da Figura 6 foram obtidas após 2500 simulações. As figuras 7, 8, 9, 10 e 11 ilustram estas regiões.

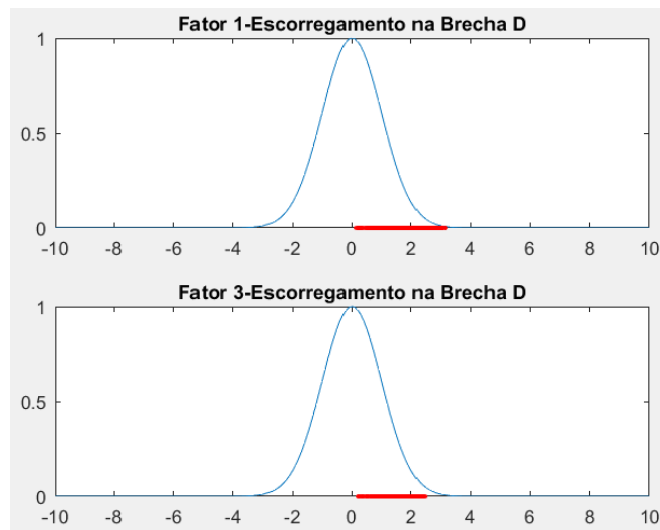


Figura 7: Escorregamento Brecha D

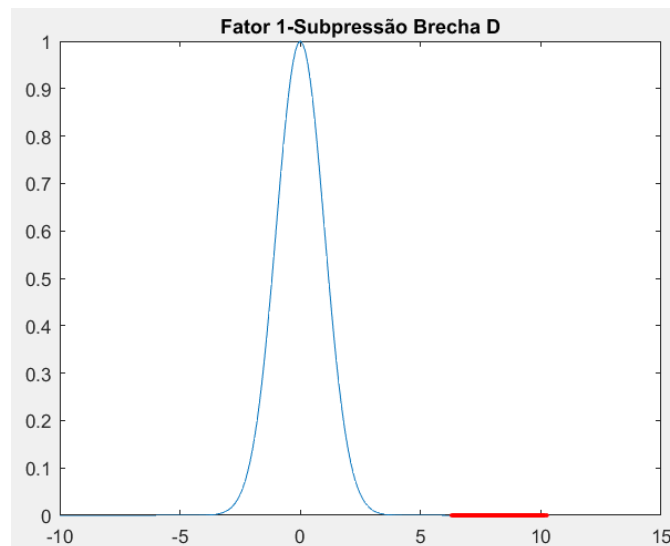


Figura 8: Subpressão na Brecha D

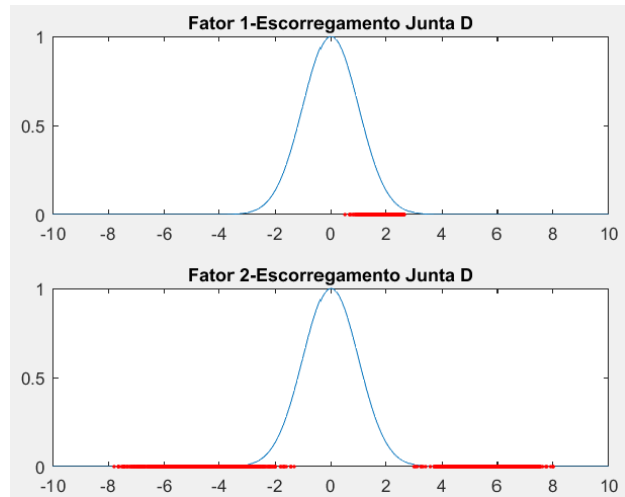


Figura 9: Escorregamento Junta D

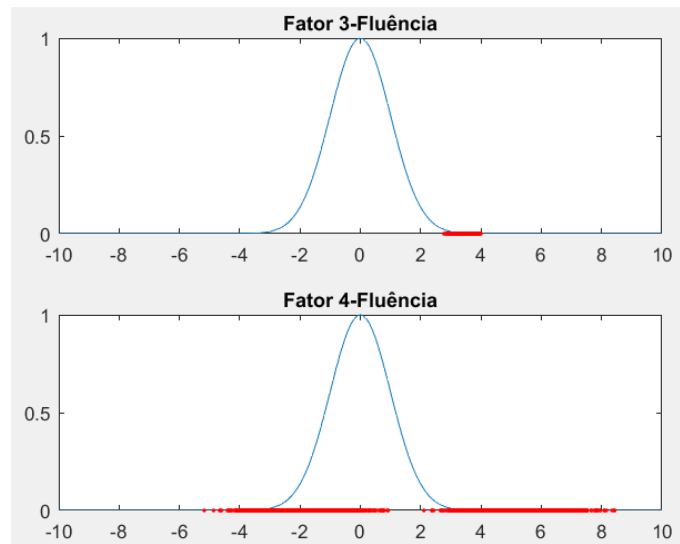


Figura 10: Fluência

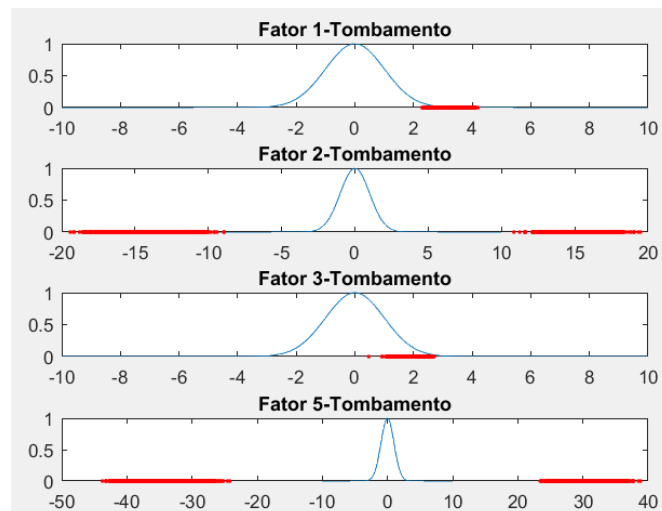


Figura 11: Tombamento

CONCLUSÃO

O procedimento apresentado forneceu os modos de falha de anomalias que são possíveis de serem monitoradas no bloco-chave A-7 da barragem da Usina de Itaipu de acordo com a metodologia adotada, e que pode servir como uma ferramenta para o monitoramento, e também auxiliar na tomada de decisão. Esses resultados poderão ser utilizados futuramente para a determinação de uma função de desempenho multivariada que satisfaça as exigências da análise de confiabilidade estrutural, que pode ser implementada computacionalmente e utilizada para simplificar o monitoramento do bloco.

REFERÊNCIAS

- [1] ICOLD, *Critérios de Projeto Civil de Usinas Hidrelétricas*, Centrais Elétricas Brasileiras S.A, Eletrobrás (2003).
- [2] Brasil. Lei nº 12.334, de 20 de setembro de 2010.
- [3] Brasil. Resolução nº 144 de 10 de julho de 2012.
- [4] O. C. Silva, *Análise Multivariada e Confiabilidade Estrutural no Estudo de Riscos em Blocos de Concreto de Barragens Altamente Instrumentadas*, 127 f, Tese de Doutorado, Método Numéricos, Curitiba, Paraná (2016).
- [5] J. G. Filho, *Desempenho de Obras Geotécnicas*, 523 p, Editora Universitária da UFPE, Recife (2006).
- [6] Eletrobrás, *Critérios de Projeto Civil de Usinas Hidrelétricas*, vol. 1, Rio De Janeiro (2003).
- [7] O. C. Silva and J. M. Marques, *Estudo de Variabilidade dos Instrumentos Instalados em Blocos Chaves da Barragem de Itaipu para Determinação de Modos de Falha*, Em: XXX SNGB – Seminário Nacional de Grandes Barragens, Maio, 10-12, Comitê Brasileiro de Grandes Barragens – CBDB, Foz do Iguaçu (2015).
- [8] O. C. Silva, J. M. Marques and I. P. Nieradka, *Testes Estatísticos e Análise Multivariada na Determinação dos Modos de Falha Monitorados pelos Instrumentos em Blocos de Concreto da Barragem de Itaipu*, vol. 5, pag. 3-21, Revista Holos (2015).
- [9] J. F. A. Silveira, *Instrumentação e Segurança de Barragens de Terra e Enrocamento*, 208 p, Editora Oficina de Textos, São Paulo (2003).
- [10] Itaipu, *Usina Hidrelétrica de Itaipu – Aspectos Teóricos de Engenharia*, 784 f, TAB Marketing Editorial, Foz do Iguaçu, (2003).
- [11] A. G. Biase and D. F. Ferreira, *Tese Computacionalmente Intensivo Baseado na Distância de Mahalanobis para Normalidade Multivariada*, 22 f, Bras, Biom, Universidade Federal de Lavras, vol. 30, nº1 (2002).
- [12] J. M. Araújo, *Estruturas de Concreto: Modelos de Previsão da Fluência e da Retração do Concreto*, nº4, Rio Grande, Dunas (2002).