

## Análise comparativa de Componentes Principais via simulação de variáveis de distribuições contínuas

Ianne Porfirio de Queiroz<sup>1 4</sup>

Hudson de Sousa e Silva<sup>2 4</sup>

Edwirde Luiz Silva<sup>3 4</sup>

Análise de Componentes Principais (ACP) consiste numa técnica multivariada de modelagem de estrutura de covariância, cujo objetivo é reduzir a dimensão dos dados, preservando grande porção da variância total explicada por estes (FERREIRA, 2008, p.395). Componentes Principais (CP) são variáveis latentes, ortogonais entre si, obtidas através de combinações lineares de um grupo de variáveis correlacionadas. Neste trabalho, buscou-se ilustrar a ACP, via simulação, a partir do uso do *software* R 3.0.1. Foram definidas seis variáveis correlacionadas – cada uma delas com 30 valores aleatórios – para algumas distribuições contínuas: Cauchy padrão; Gama; Exponencial; Qui-quadrado; Lognormal. À exceção da primeira, que possui esperança e variância indefinidas, as distribuições foram estabelecidas considerando média dois e variância quatro. Para estas, a ACP foi aplicada à matriz de correlação (em razão de não padronização). A partir das análises da proporção da variação explicada por cada CP e do estudo dos gráficos *biplot* (representação das variáveis e das observações de um conjunto multivariado) e *screeplot* (gráfico de autovalores *versus* CP), foram comparados os resultados por distribuição. Utilizando o critério de corte de 80%, adotado conforme sugestão de Ferreira (2008, p.403), verificou-se que os conjuntos oriundos das distribuições Cauchy, Gama, Exponencial e Qui-quadrado poderiam ser eficientemente explicados pelas duas primeiras CP. A *Lognormal*, por sua vez, concentrou aproximadamente 99% da variância total na primeira CP, que poderia substituir o conjunto original de dados, preservando grande porção da variância total explicada por estes.

**Palavras-chave:** ACP, simulação, distribuições contínuas.

### Referências:

[1] FERREIRA, Daniel Furtado. **Estatística Multivariada**. 1.ed. Lavras: Ed. UFLA, 2008.

[2] **R Core Team** (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

---

<sup>1</sup>UEPB/CCT/DE. Email: ianne.queiroz@gmail.com

<sup>2</sup>UEPB/CCT/DE. Email: hudsonsousasilva399@hotmail.com

<sup>3</sup>UEPB/CCT/DE. Email: edwirde1@hotmail.com

<sup>4</sup>Agradecimentos à UEPB pelo apoio financeiro