Nome do artigo: Spatial autocorrelation analysis of multivariate rural insurance data

Link para o artigo: <https://biometria.ufla.br/index.php/BBJ/article/view/642/385>

Tópico abordado: Reconhecimento de Padrões.

O artigo busca utilizar técnicas multivariadas como análise de componentes principais e Análise Exploratória de Dados Espaciais para estudar dependência de seguros rurais de acordo com regiões geográficas brasileiras.

A parte computacional do artigo foi feita em Python, mas poderia ser facilmente realizada em R e talvez em SAS também. A graduação e estatística da UnB oferta matérias para capacitação em todas essas três linguagens. A técnica de análise de componentes principais também é estudada na disciplina Análise Multivariada 1 na graduação, enquanto técnicas de estatística espacial é por vezes ofertada, enquanto disciplina optativa.

Acredito ser plenamente possível um aluno de graduação de estatística da UnB replicar este artigo.

Nome do artigo: Ordinal data and residual analysis: Review and application

Link para o artigo: <https://biometria.ufla.br/index.php/BBJ/article/view/637/380>

Tópicos abordados: Análise Estatística Paramétrica, Análise Estatística Não Paramétrica e Aprendizado de Máquinas e/ou Estatístico não Supervisionado

O artigo busca fazer inferência paramétrica para um modelo logístico utilizando estimadores de máxima verossimilhança, seguido de análise de resíduos utilizando o teste não-paramétrico de Kolmogorov-Smirnov. O objetivo do estudo é estudar o dano no fígado de Tambaquis, de acordo com o genoma pré conhecido do animal.

Nominalmente, todas as técnicas são ensinadas nas disciplinas obrigatórias da graduação de estatística: Inferência estatística; Análise de dados categorizados; Métodos estatísticos 2. A aplicação computacional também seria possível cursando as matérias obrigatórias CE1 e CE2. Um reforço no aprendizado da técnica de modelagem logística seria possível com o curso optativo oferecido Modelos Lineares Generalizados.

Acredito que, respeitada a complexidade do artigo, seria possível um aluno de graduação de estatística da UnB replicar este artigo.

Nome do artigo: Classification and Analysis of Patients with COVID-19 Using Machine Learning

Link para o artigo: <https://biometria.ufla.br/index.php/BBJ/article/view/588/362>

Tópico abordado: Aprendizado de Máquinas e/ou Estatístico não Supervisionado.

O artigo busca classificar pacientes com COVID-19 utilizando algoritmos de aprendizado estatístico, como SVM e Naive-Bayes, entre outros. Técnicas computacionais sobre utilização desses algoritmos são citados nos cursos de CE2 – R e CE2 – Python, além de Estatística Computacional. Algumas matérias de Tópicos (1 e 2) também abordam estes.

Com alguma pesquisa, é possível um aluno de graduação elaborar um artigo como este. Inclusive, eu mesmo cheguei a fazer algo relacionado a pedido do meu estágio, criando um código para realizar processamento de linguagem natural e posterior classificação com alguns algoritmos. Disponível em

https://github.com/penasta/NLP/blob/main/funcao\_nlp.ipynb