



# SEMINÁRIO de BIOESTATÍSTICA

RESUMOS DOS SEMINÁRIOS APRESENTADOS PELOS ALUNOS

UC DO MESTRADO EM BIOESTATÍSTICA, 2022-2023

-----  
9 de fevereiro (quinta-feira)

9:30 – 10:00

**Avaliação dos potenciais corredores ecológicos para os carnívoros nos Neotrópicos**

**Gabriel Borges**

**Resumo:** Os carnívoros desempenham um papel fundamental nas relações tróficas dos ecossistemas, logo uma melhor compreensão da sua distribuição e das condições dos seus habitats são cruciais para se conseguir estabelecer planos de gestão e de conservação de forma a salvaguardar a saúde ecológica dos ecossistemas neotropicais no futuro.

Este trabalho consiste numa análise estatística de dados sobre avistamentos de carnívoros obtidos em vários países da América do Sul e da América Central. Os dados estão disponíveis num artigo de dados recentemente publicado por Nagy-Reis et al (2020).

O estudo tem como objetivo principal identificar corredores ecológicos para espécies de carnívoros nos Neotrópicos. A primeira parte do estudo será mapear o habitat adequado para seis famílias de carnívoros com base em variáveis ambientais, usando quatro modelos: Poisson GLM, N-mixture model, N-mixture with conditional autoregressive (CAR) random effects e N-mixture with splines. Segunda parte do estudo passará por identificar potenciais corredores de movimento utilizando o software circuitscape.

10:00 – 10:30

**Using supervised machine learning to quantify cleaning behaviour**

**Raúl Oliveira**

**Resumo:** In coral reefs, cleaner wrasses, Labroides spp, are known for their cleaning activity. These cleaner fish inspect client fish to mainly remove and feed on ectoparasites. Cleaners can interact with more than 2,000 clients daily. Cleaning behaviour includes a vast repertoire from cleaning bites, client jolts, tactile stimulation, chasing, advertising dances, punishment and manipulation – all measures of cleaning motivation and interaction quality. The analyses of these interspecific interactions among the fish have usually been made through the manual processing of video recordings. Apart from being very time-consuming, manual processing can be unreliable due to variation and bias across observers. Automatic posture tracking and behavioural classification are rapidly becoming indispensable to studying animal behaviour accurately. Within this context, using supervised and unsupervised machine learning pipeline, I will build a neural network that record and classify cleaning mutualistic behaviour events automatically.

10:30 – 11:00

**Análise estatística de dados longitudinais: efeito do ciclo menstrual e dos contraceptivos orais na composição corporal em estudantes universitárias do sexo feminino**

**Beatriz Ferreira**

**Resumo:** O estudo da composição corporal é determinante no diagnóstico e acompanhamento de situações de risco nutricional, bem como, uma variável determinante na performance dos atletas. As possíveis alterações na composição corporal consequentes de alterações fisiológicas características do ciclo menstrual não são claras, sendo ainda inconclusiva a realização de avaliações físicas a mulheres nas diferentes fases do ciclo menstrual. Os objetivos do projeto de mestrado são estudar a variação da composição corporal, da água corporal e da compressibilidade das pregas cutâneas durante o ciclo menstrual e durante o ciclo de toma de contraceptivo oral. Neste seminário apresenta-se o planejamento experimental, o cálculo da dimensão da amostra (constituída por estudantes universitárias do sexo feminino, dos 17 aos 25 anos) e alguma informação sobre os dados recolhidos até ao momento. Apresentam-se ainda as metodologias estatísticas que se planeia aplicar, como por exemplo: Análise da Variância (ANOVA) para medidas repetidas e Análise da Variância Multivariada (MANOVA).

11:00 – 11:30

**Redução de dimensionalidade em modelos de sobrevivência baseados em dados de expressão genética: Uma aplicação sobre cancro cerebral**

**João Brandão**

**Resumo:** O glioblastoma é um dos cancros cerebrais malignos mais agressivos, apresentando uma sobrevida muito baixa, sendo necessário encontrar biomarcadores de prognóstico, que expliquem a sobrevivência nestes pacientes e possam ser usados como potenciais alvos terapêuticos. Neste sentido, será realizada uma análise de sobrevivência, com seleção de variáveis incorporada, em dados sobre o tempo de vida de indivíduos com glioblastoma e as expressões genéticas presentes nos tumores.

Devido à elevada dimensionalidade dos dados, serão consideradas técnicas de regularização, particularmente Elastic Net e técnicas de regularização baseada em redes, obtendo-se diferentes modelos de Cox, consoante a abordagem de redução de dimensionalidade adotada. A partir dos vários modelos construídos, utilizar-se-á o Rank Product test como método de consenso. Espera-se identificar um subconjunto de variáveis (genes) importantes na sobrevivência dos pacientes e identificar *outliers*, correspondentes a pacientes com tempo de vida anómalo, dada a expressão genética, possibilitando a deteção de novos biomarcadores de prognóstico.

Foi realizada uma análise descritiva onde serão apresentados alguns resultados.

11:30 – 12:00

**Índice de exposição ambiental para o ELEAT - Estudo longitudinal da exposição ambiental a tóxicos**

**Beatriz Costa**

**Resumo:** A Perturbação do Espectro do Autismo (PEA) é uma doença do neurodesenvolvimento que implica dificuldades sociais e de comunicação e comportamentos restritos e repetitivos, sendo caracterizada por uma grande heterogeneidade clínica. A etiologia desta doença é complexa, sendo os fatores genéticos responsáveis pela explicação de 10-20% da mesma. No entanto, estimativas recentes de heritabilidade apontam para um efeito substancial de fatores ambientais, com vários estudos a sugerirem um papel da exposição a tóxicos durante a gravidez e primeiro ano de vida.

Com o intuito de compreender a complexidade da PEA e a ligação entre exposição ambiental e neurodesenvolvimento foi desenvolvido por uma equipa de investigadores da Universidade da Califórnia um questionário, o ELEAT. Este recolhe informação detalhada sobre a exposição ambiental durante o desenvolvimento inicial das crianças.

este seminário pretende-se apresentar o objetivo do trabalho de mestrado, um pouco da metodologia estatística proposta para o mesmo, a amostra a recolher e uma breve análise descritiva dos dados já disponíveis.

**12:00 – 12:30**

**Estimação de parâmetros biológicos do *Pagellus bogaraveo* - Aplicação de modelos de espaço de estados bayesianos**

**Érica Ribeiro**

**Resumo:** O Goraz é uma espécie de peixe cujos ciclos de vida e reprodutivo são relativamente complexos, o que complica a estimação dos seus parâmetros biológicos e de abundância. Os biólogos acreditam que a espécie de Goraz presente na Costa Portuguesa tem algumas características diferentes das do Goraz existente nas costas Espanhola e Marroquina. Contudo, a quantidade de dados para realizar inferência é escassa.

Os grandes objetivos deste trabalho são, criar um modelo capaz de estimar a abundância e os parâmetros populacionais deste peixe na Costa Portuguesa na presença de dados escassos e compará-los com os dos dois países mencionados.

Propõe-se para a resolução deste problema uma abordagem bayesiana através da utilização de modelo de espaço de estados. O modelo tem como objetivo, partindo de distribuições *a priori* adequadas, gerar distribuições *a posteriori*, condicionais aos dados observacionais, dos estados da população e dos parâmetros desconhecidos.

O modelo será desenvolvido e implementado no Nimble. Será depois ajustado com dados de pesca recolhidos pelo IPMA.

**12:30 – 13:00**

**Estimação do tempo até à ocorrência de lesões em jogadores de futebol**

**Daniela Sousa**

**Resumo:** A prevenção e predição de lesões em contexto desportivo é um tema bastante estudado atualmente, dada a importância que a ocorrência de uma lesão tem para o atleta e para o seu clube. Ao analisar um padrão anterior, tal como a intensidade de treino e/ou a existência de lesão prévia, pensa-se ser possível inferir se um jogador terá maior ou menor probabilidade de se lesionar futuramente.

O objetivo deste trabalho passa por estimar o tempo até à ocorrência da primeira lesão (e subsequentes) decorrente de sobrecarga, ao nível dos membros inferiores, em jogadores de futebol dos seguintes escalões do Sport Lisboa e Benfica: equipa B, sub-23 e sub-19. A análise dos dados referentes às lesões sofridas pelos jogadores compreenderá as épocas 20/21, 21/22 e 22/23 (até ao mês de Dezembro) e será realizada recorrendo a técnicas de análise de sobrevivência (e.g. análise de sobrevivência de eventos recorrentes), tendo em consideração a tipologia/localização da lesão.

-----  
**10 de fevereiro (sexta-feira)**

**9:30 – 10:00**

**Correction for baseline covariates in clinical trials and observational studies**

**Matilde Francisco**

**Resumo:** In longitudinal studies, individual measures are taken at several time points to allow the understanding of the degree and the direction of the change of a certain characteristic over time. Baseline measurements are taken at the beginning of the study with several purposes, including the assessment of treatment effect based on the change from baseline. In the case of observational studies, where baseline characteristics of patients in treatment groups may differ, it is necessary to correct for these systematic differences. In clinical trials, randomization is a fundamental process, since it ensures that it is possible to compare the effect of the different treatments between groups. In this seminar different approaches on how to account for baseline measurements in longitudinal studies will be presented: Constrained Longitudinal Data Analysis, where the baseline is modelled as a dependent variable, and Longitudinal Analysis of Covariance where the baseline is used as a covariate.

**10:00 – 10:30**

**Microbial diversity and discrimination of queijo de Azeitão and queijo de Nisa PDO cheeses based on metagenomic data**

**Carlota Teles**

**Resumo:** Studies related to the characterization and discrimination of the microbiota present in Protected Designation of Origin (PDO) cheeses play a very important role in the understanding of its organoleptic properties, as well as its safety, quality and typicity.

The data for this project was collected from 7 different cheese factories in the regions of Azeitão and Nisa, in Portugal, from 2016 to 2022. A series of multivariate statistical analyses and methodologies, such as Principal Components Analysis (PCA) and classification trees, have been carried out to find differences between the cheese samples, considering the following variables: region of origin, cheese factory and year of production. These results will be presented in the seminar, together with statistical methods like Linear Discriminant Analysis Effect Size (LEfSe) and PERmutational Multivariate ANalysis Of VAriance (PERMANOVA), which will be applied in the near future to identify discriminant characteristics of these PDO cheeses that could be associated with specific samples and, thus, acting as cheese-ID cards.

**10:30 – 11:00**

**Continuous analysis of the survival rate and sex ratio in zebrafish**

**Mariana Duarte**

**Resumo:** Os peixes-zebra são amplamente utilizados em variados estudos científicos devido aos múltiplos benefícios que trazem, como a grande quantidade de embriões que produzem, fácil manutenção e por partilharem cerca de 70% dos seus genes com o ser humano.

O projeto do IGC surgiu da necessidade da criação de um sistema de monitorização com o intuito de gerir a sua cultura de peixes-zebra, de forma que o seu fornecimento cumpra padrões exigidos quanto à quantidade e rácio sexual.

Serão analisadas as taxas de sobrevivência do desenvolvimento inicial dos peixes e o rácio sexual. De seguida, vai-se desenvolver um sistema de apoio à decisão que permita a identificação de clusters de linhas genéticas, a deteção de linhagens problemáticas no que se refere à taxa de sobrevivência e ao rácio sexual e a identificação dos fatores que contribuem para a ocorrência de situações anormais. Para tal, ir-se-á recorrer a uma análise exploratória dos dados, técnicas de *clustering* e modelos estatísticos.

11:00 – 11:30

### **Impacto da medicina tradicional chinesa na qualidade de vida dos doentes com lombalgia**

**Maria João Pinela**

**Resumo:** A Medicina Tradicional Chinesa (MTC) tem mostrado resultados eficazes no tratamento da dor, sendo recomendada como primeira linha nos tratamentos não invasivos. Ao fazer uma avaliação do paciente, propondo um diagnóstico e tratamento diferencial, a MTC tem como objetivo não só promover a redução ou eliminação da dor, como também melhorar aspetos estruturais e fisiológicos abrangentes, promotores da harmonização vital do paciente.

Este estudo tem como objetivo principal avaliar a eficácia dos tratamentos da MTC na qualidade de vida dos pacientes com lombalgia. Tentaremos ainda perceber se os tratamentos de MTC nos pacientes com lombalgia contribuíram para uma recuperação eficaz do paciente, avaliar as melhorias observadas na diminuição da dor, no equilíbrio emocional, na qualidade do sono e na função digestiva, diminuição de ingestão de fármacos, etc. Queremos ainda aferir se os pacientes com lombalgia que fazem tratamentos de MTC fizeram previamente tratamentos de Medicina Alopática, averiguando os resultados obtidos e a existência de comorbidades associadas aos tratamentos realizados.

11:30 – 12:00

### **Otimização da avaliação de stocks limitados por dados de raias nos portos portugueses através de estatística bayesiana e uma regra empírica de captura**

**Erick Chatalov**

**Resumo:** Em todo o mundo, a gestão eficaz de algumas pescarias e stocks explorados comercialmente é prejudicada pela falta de estimativas confiáveis do status do stock. Vários fatores contribuem para a indisponibilidade das informações necessárias para realizar uma avaliação quantitativa do *stock*, para estimar o *status* do *stock* e a dinâmica histórica da pescaria. A procura por um sistema de gestão baseado na ciência tem aumentado, nomeadamente *stocks* com informação escassa e/ou dados limitados.

Esta necessidade é reforçada pelas mudanças climáticas que estão ocorrendo principalmente nos ecossistemas costeiros e que estão afetando diretamente o setor pesqueiro. Pretende utilizar-se modelação por simulação, incorporando conhecimento biológico prévio e avaliação MSY para avaliar regras de captura empíricas alternativas no procedimento de gestão de pesca, visando o rendimento máximo sustentável. Adicionalmente, um algoritmo de busca estocástica será utilizado para otimizar o procedimento de gestão de raias em águas portuguesas.

12:00 – 12:30

### **Análise de sobrevivência de espécies de raias (Rajiformes)**

**Marta Pereira**

**Resumo:** As espécies de raias (Rajiformes) são um grupo de espécies alvo de capturas acessórias da pesca comercial e conhecidas por terem uma capacidade de sobrevivência relativamente elevada, inclusive nas primeiras fases do seu ciclo de vida e depois de terem sido capturadas por artes de pesca.

A obrigação de desembarque segundo as diretrizes da União Europeia levanta questões sobre como as estimativas de sobrevivência são incorporadas no processo de avaliação de stocks, sendo recomendada a investigação neste domínio.

Desta forma, o presente projeto tem por objetivo investigar métodos estatísticos, como o desenvolvimento de modelos de sobrevivência adequados aos casos-estudo e a estimação das respetivas taxas de sobrevivência, aplicados a este grupo de espécies, considerando os seguintes casos-estudo: 1) sobrevivência de embriões de *Raja undulata* (raia curva) em cativeiro; 2) sobrevivência à pesca de espécies de raias (Rajiformes) capturadas por redes de tresmalho.

12:30 – 13:00

### **Tempo até ao diagnóstico da tuberculose: uma análise de dados com truncatura à direita**

**Inês Gambutas**

**Resumo:** A tuberculose é uma doença infecciosa transmitida maioritariamente por via aérea através da inalação de gotículas, expelidas pela tosse, fala ou espirros, que afeta geralmente os pulmões. Um tempo longo desde o início dos sintomas até ao diagnóstico é um problema importante de saúde pública, visto que o contágio ocorre até a pessoa infetada começar o tratamento.

Os dados analisados neste trabalho são provenientes da base de dados do Sistema de Vigilância da Tuberculose (SVIG-TB), e incluem as datas de início de sintomas e de diagnóstico, assim como características clínicas e sociodemográficas dos indivíduos.

É um estudo observacional retrospectivo, que decorreu entre 01/01/2008 e 31/12/2017. Neste estudo existe truncatura à direita, visto que apenas são incluídos os indivíduos para os quais o acontecimento de interesse (diagnóstico) tenha ocorrido até 31/12/2017.

O objetivo deste projeto consiste na identificação dos fatores que têm influência no tempo até ao diagnóstico, ajustando um modelo de regressão de Cox aos dados com truncatura à direita.