

Workshop

Statistical Modeling and Applications for Environmental Data

📍 **Venue:** Department of Economics and Business, University of Catania, Corso Italia 55, 95129 Catania, Room 5 (Aula 5, Palazzo delle Scienze)

📅 Dates & Schedule

Monday, September 22. Sessions begin at 9:30 AM (speakers underlined).

From 9:30 a.m. to 10:30 a.m.

- *Statistics on Manifolds: An Introduction to Directional Models*

Authors: Priyanka Nagar and Andriette Bekker

Abstract: There are a large number of applications that require the analysis of data not realized in the Euclidean space, but rather on some manifold (circles, spheres, hyperspheres, cylinders, tori, etc.). Directional statistics constitutes a specialised branch within statistical analysis that focuses on angular observations. An inherent challenge in handling directional data lies in the non-linearity of the sample space, typically represented as a circle or circular manifold. This unique feature has garnered increased attention over the past two decades, primarily driven by the rise in data and the corresponding demand for customized statistical methodologies. However, in many studies, the cyclicity of a directional variable is often neglected and mistreated as a linear variable. This presentation provides an overview of the world of directional statistics, emphasising the need for this adapted statistical technique. The fascinating background and recent developments within this field will be explored. The usefulness of these techniques will be illustrated with some examples based in various application domains.

From 10:30 a.m. to 11:00 a.m.

Clustering Incomplete Emissions Data via Finite Mixtures of Scale Mixtures of Skew-Normal Distributions

Authors: Jason Pillay, Cristina Tortora, Antonio Punzo, and Andriette Bekker

Abstract: This presentation investigates the use of cluster analysis on incomplete CO2 emissions data, with an emphasis on uncovering patterns in greenhouse gas and air pollution emissions across different countries and among different sectors. This presentation utilizes model-based clustering with finite mixture models to categorize countries based on their emissions profiles by extending finite mixtures of the scale mixture of skew normal (SMSN) distributions to handle incomplete data under a missing at random (MAR) mechanism. The skew-normal distribution, as parameterised by Azzalini and Beaver, offers several advantages when used in model-based clustering since it can model skewed clusters while encompassing the symmetric normal distribution as a special case. The SMSN distributions account for skewness and excess kurtosis and, with the new extension, also include data with missing values. We found homogeneous subclasses across countries, accounting for the skewed and leptokurtic nature inherent to emissions data.

From 11:00 a.m. to 11:30 a.m.

- *Modeling Bounded Count Environmental Data Using a Contaminated Beta-Binomial Regression Model*

Authors: Arnoldus F. Otto, Antonio Punzo, Andriëtte Bekker, Johannes T. Ferreira, Salvatore D. Tomarchio, and Cristina Tortora

Abstract: Climate change is a crucial aspect of environmental challenges, with profound implications for human well-being, affecting vital ecosystem services such as clean water, food production, and pollination. It also causes species displacement, habitat loss, and increased extinction risks. We investigate two environmental applications related to climate change, where observations consist of bounded counts. The first application examines the effect of winter malnutrition on mule deer (*Odocoileus hemionus*) fawn mortality. Animals are crucial in ecosystem services, including seed dissemination, pollination, and pest control. Since climate change affects species' survival, it is essential to study these impacts further. The second application investigates the public perception of climate change, which is vital for shaping effective policies and environmental strategies. The Eurobarometer 95.1 survey, conducted in March–April 2021, included a key question assessing the perceived severity of climate change on a scale from 1 to 10. Analyzing responses to this question is crucial for gaining deeper insights into public sentiment regarding climate change. The binomial and beta-binomial models are commonly used for bounded count data, with the beta-binomial model offering the advantage of accounting for potential overdispersion. However, extreme observations in real-world applications may hinder the performance of the beta-binomial model and lead to misleading inferences. To address this issue, we propose the contaminated beta-binomial distribution, which provides the necessary flexibility to accommodate extreme observations. The contaminated beta-binomial model accounts for overdispersion and extreme values while maintaining the mean and variance properties of the beta-binomial distribution. The availability of covariates that improve inference about the mean of the bounded count variable motivates the further proposal of the contaminated beta-binomial regression model. Different versions of the contaminated beta-binomial regression model – where none, some, or all of the contaminated beta-binomial parameters are regressed on available covariates – are fitted to the datasets. The effectiveness of our model is also demonstrated through a sensitivity analysis to assess the impact of extreme values on parameter estimation.

From 11:30 a.m. to 12:00 a.m.

- *Model-Based Strategies for Handling Cluster-Wise Skewness*

Authors: Cristina Tortora

Abstract: In model-based clustering, the population is assumed to be a combination of sub-populations. Typically, each sub-population is modeled by a mixture model component, distributed according to a known probability distribution. Each component is considered a cluster. Two primary approaches have been used in the literature when clusters are skewed: 1) transforming the data within each cluster and applying a mixture of symmetric distributions to the transformed data, and 2) directly modeling each cluster using a skewed distribution. The generalized hyperbolic distribution is notably flexible among skewed distributions and includes many other known distributions as special or limiting cases.

This talk will introduce and compare those two approaches to handle skewness in model-based clustering.

💡 This event is organized within the **PRIN 2022 Project: “THE SMILE PROJECT: Statistical Modelling and Inference to Live the Environment.”**