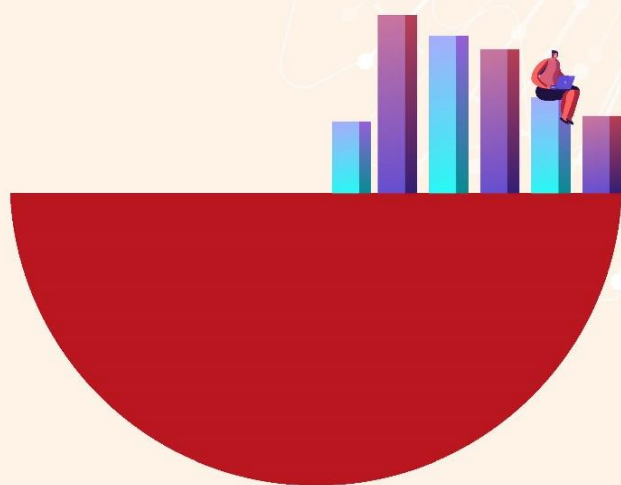# 4η

ΠΑΓΚΥΠΡΙΑ
ΗΜΕΡΙΔΑ
ΣΤΑΤΙΣΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ

STATISTICS

Πανεπιστήμιο Λευκωσίας

Παρασκευή, 9 Ιουνίου 2023
9:00 - 18:00

Αμφιθέατρο Jean Monnet
Αμφιθέατρα Da Vinci και Newton

Συνδιοργανωτές:

ΚΥΠΡΙΑΚΗ
ΣΤΑΤΙΣΤΙΚΗ
ΕΤΑΙΡΕΙΑ
CYPRUS STATISTICAL SOCIETY

UNIVERSITY of NICOSIA

Χορηγοί:

ClinBAY
Statistical solutions to drug development

pulse
market research

Πρόγραμμα - 4ᵗ Παγκύπρια Ημερίδα Στατιστικής, 9 Ιουνίου 2023
Πανεπιστήμιο Λευκωσίας – Κυπριακή Στατιστική Εταιρεία

Program - 4th Pancyprian Conference in Statistics, 9 June 2023
University of Nicosia – Cyprus Statistical Society

9:00 – 10:00	Αμφιθέατρο Jean Monnet (M-203) Εγγραφές - Registrations	
10:00 – 11:00	**Αμφιθέατρο Jean Monnet (M-203)** **Προεδρεύουσα – Χαριτίνη Τσαγκάρη** **Chair – Haritini Tsangari** • **Χαιρετισμοί – Welcoming messages** • **Προσκεκλημένη ομιλήτρια - Ξένια Μισκουρίδου** **Keynote Speaker – Xenia Miscouridou** "Inferential Artificial Intelligence (iAI): Statistics and Machine Learning for real-world problems"	
11:00 – 11:30	**Αμφιθέατρα στο Κτήριο Europa** **Διάλειμμα – Καφές** **Coffee Break**	
11:30 – 12:50	*Αμφιθέατρο Newton* **Επιστήμη Δεδομένων – Data Science** **Προεδρεύων – Δημήτρης Τριχινάς** 1. **Ανδρέας Νεοκλέους:** Ψηφιοποίηση και ανάλυση δεδομένων σχετικά με τις τάσεις και πρότυπα τιμών βασικών καταναλωτικών αγαθών στην Κύπρο 2. **Charalambia Varnava:** SMART: A new MCMC code for studying galaxy evolution 3. **Constantia Malekkou:** Service point placement by customer Geolocation Clustering 4. **Θεόδωρος Μωυσιάδης:** Καινοτόμος δυναμικός προβλεπτικός αλγόριθμος για την εξατομίκευση της πρόβλεψης του χρόνου πρώτης θεραπείας στη χρόνια λεμφοκυτταρική λευχαιμία	*Αμφιθέατρο Da Vinci* **Στατιστική και Υγεία – Statistics and Health** **Προεδρεύων – Κώστας Χριστοφή** 1. **Theo Economou:** Hierarchical Distributed Lag Models 2. **Maria Zanti:** A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants 3. **Kristia Yiangou:** Polygenic risk score differences among European ancestry populations; implications for breast cancer risk prediction

12:50 – 14:00	Γεύμα – Lunch Break (Block Restaurant)	
14:00 – 15:20	*Αμφιθέατρο Newton* **Φύλο και Στατιστική –** **Gender and Statistics** **Προεδρεύουσα – Πετρούλα Μαυρικίου** 1. **Μαριλένα Παρασκευά:** Συμμετοχή Γυναικών σε Εθνικά Προγράμματα Έρευνας και Καινοτομίας 2. **Margarita Zachariou:** Evaluation of gender parity in the context of gender representation and collaboration in the Cyprus Institute of Neurology and Genetics 3. **Petroula Mavrikiou:** Practices to Achieve Gender Balance in Academia 4. **Haritini Tsangari:** Mind the *gender* gap!	*Αμφιθέατρο Da Vinci* **Κοινωνική Στατιστική –** **Social Statistics** **Προεδρεύων – Σέργιος Αγαπίου** 1. **Konstantinos Giannakou**: The impact of the COVID-19 pandemic on perceived fear of future pandemics: A cross-sectional study 2. **Christos Ntais:** Medical cannabis: establishing its roots in pharmaceutical marketing 3. **Michailina Siakalli:** The role of conspiracy theories, perceived risk, and trust in science on COVID-19 vaccination decisiveness: evidence from Cyprus.
15:30 – 16:30	*Αμφιθέατρο Newton* **Εκπαιδευτική Στατιστική –** **Educational Statistics** **Προεδρεύουσα – Ιουλία Τελεβάντου** 1. **Elena Papanastasiou:** Persistence in test-taking. New insights from examining timing data 2. **Ioulia Televantou:** Mathematics self-concept: Investigating the Immigrant Paradox with TIMSS 2019 data from Germany 3. **Danxia Chen:** E-learning adoption and Academic Self-Efficacy: An Application of the Technology Acceptance Model	*Αμφιθέατρο Da Vinci* **Στατιστική Θεωρία και Μεθοδολογία – Statistical Theory and Methodology** **Προεδρεύων – Ανδρέας Αναστασίου** 1. **Sophia Loizidou:** A data-adaptive algorithm for multiple change-point detection 2. **Eleni Manoli:** Assessing the effectiveness of indirect questioning techniques by detecting liars 3. **Κώστας Διαμαντίδης:** Διαδικασία αξιολόγησης της εφαρμογής του Κώδικα Ορθής Πρακτικής για τις ευρωπαϊκές στατιστικές
16:30 – 16:50	Διάλειμμα - Καφές	
16:50 – 17:50	*Αμφιθέατρο Newton* **Ετήσια Γενική Συνέλευση της ΚΣΕ** **Annual General Assembly of CSS**	

<u>Σημείωση:</u> Στο πρόγραμμα αναφέρονται τα ονόματα των ομιλητών/τριών. Το βιβλίο των περιλήψεων περιλαμβάνει τα ονόματα όλων των συν-συγγραφέων (όπου υπάρχουν).

<u>Note:</u> The program includes the names of the speakers. The book of abstracts includes the names of all co-authors (if applicable).

ΧΟΡΗΓΟΙ:

4 η Παγκύπρια Ημερίδα Στατιστικής
4 th Pancyprian Conference in Statistics
09/06/2023

https://www.unic.ac.cy/el/stats2023/

ΠΕΡΙΛΗΨΕΙΣ ΕΡΓΑΣΙΩΝ

BOOK OF ABSTRACTS

ΠΡΟΣΚΕΚΛΗΜΕΝΗ ΟΜΙΛΙΑ - KEYNOTE SPEECH

"Inferential Artificial Intelligence (iAI): Statistics and Machine Learning for real-world problems"

Xenia Miscouridou
Imperial College London, Department of Mathematics and the
I-X center for Artificial Intelligence

Abstract

Statistics and Machine Learning have undoubtedly proved to be useful in tackling problems in a variety of application areas including social science, public health, environmental science, media, or transportation. While these two scientific disciplines share a great number of tools, they have some fundamental differences. Nevertheless, the synergies between them provide a promising avenue in delivering truly useful, explainable and efficient methodologies that can be used in impactful problems.

In this talk I aim to show some examples related to my research, of how Statistics and Machine Learning have been used to answer real-life strategic questions. The first concerns the recent pandemic - one of the biggest challenges we had to face as humanity. Using Covid-19 case and death data we developed methods to characterise the evolution of the pandemic in Brazil, influencing public policy measures. However, the quality of the results relies on the quality of the data. This is another example I will focus on as it is another problem that concerned us during the pandemic: how to correct for delays in data reporting.

Secondly, I will talk about approaches for modelling data arising in a social science context with particular focus on crime data. This is a synergy example between Statistics and Machine Learning delivering a flexible model that can explain and predict how events cluster in space and time.

1. ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ – DATA SCIENCE

Τίτλος: Ψηφιοποίηση και Ανάλυση Δεδομένων Σχετικά με τις Τάσεις και Πρότυπα Τιμών Βασικών Καταναλωτικών Αγαθών στην Κύπρο

Ανδρέας Νεοκλέους

University of Nicosia

Email: a.neocleous12@hotmail.com

This project focuses on exploring the trends and patterns in the prices of basic consumer goods in Cyprus. It displays the methodology used to digitize data previously saved in PDF format and employs exploratory data analysis techniques to analyze the price fluctuations of various products. The data visualization and analysis are performed in Python. For improved and interactive visuals, a dashboard was created using Google Data Studio, a web-based data visualization tool.

Motivation behind this study stems from the frustration of consumers constantly feeling overcharged for basic goods and the desire to stay informed about fluctuating prices. Navigating the world of deals and price changes can be overwhelming, and this project seeks to alleviate that burden by providing a reliable source of price information.

The objectives of this project are to digitize the data on basic consumer goods, conduct data analysis, create a user-friendly dashboard, and make price predictions. These objectives form the core of the project, aiming to develop a framework or toolkit that supports consumers in their purchasing decisions. By achieving these objectives and contributing to the development of a comprehensive platform, this project seeks to understand and communicate the basic consumer goods market behavior. The platform created for this project, constitutes the first basic consumer goods platform in Cyprus.

Title: SMART: A new MCMC code for studying galaxy evolution

Charalambia Varnava[*] and Andreas Efstathiou

European University Cyprus

[*]Speaker's email: Ch.Varnava@external.euc.ac.cy

Probabilistic data analysis, including Bayesian inference, has transformed scientific research over the last three decades. Many of the most significant gains have come from numerical methods for approximate inference, especially Markov chain Monte Carlo (MCMC). For example, many problems in cosmology and astrophysics have directly benefited from MCMC, because the models are often expensive to compute, there are many free parameters and the observations are usually low in signal-to-noise. Bayesian approaches and especially MCMC-based spectral energy distribution (SED) fitting codes are relatively new, but have been growing in popularity, due to their significance for studies of galaxy evolution. The complexity of the models increases, as a more accurate description of the underlying physics is needed. In this way the number of unknown parameters increases as well, including the unavoidable existence of degeneracies and correlations among them. A good statistical approach for dealing with parameter degeneracies are Bayesian methods, because they allow the user to obtain reliable confidence ranges for parameter estimates. The importance of a Bayesian study for general SED fitting has prompted the development of several algorithms in the last years. MCMC-based algorithms sample the space to infer galaxy parameters, which have been derived from the models. The spectra of galaxies are usually decomposed into a number of components. A number of radiative transfer models for the components of emission in galaxies, as well as methods of fitting them to SED data, are currently available. We have developed a new method for fitting radiative transfer models to data, using an MCMC code: Spectral energy distributions Markov chain Analysis with Radiative Transfer models (SMART). We currently test SMART with a large sample of galaxies with excellent photometry and infrared spectrophotometry. The novelty of SMART is that, although it utilizes exclusively radiative transfer models, it takes comparable time to popular energy balance methods, such as CIGALE and MAGPHYS. In particular, we utilize four libraries of models as input of SMART for fitting the SEDs: these are the starburst (Efstathiou et al. 2000, Efstathiou & Siebenmorgen 2009), active galactic nucleus (AGN) torus (Efstathiou & Rowan-Robinson 1995), spheroidal (Efstathiou et al. 2021) and disc (Efstathiou, in preparation) models. We have also added a component of polar dust in the fitting, which has been recognized as an important component in AGN. We test SMART with a large sample of galaxies of various types and at a range of redshifts with Spitzer/IRS spectroscopy and Herschel photometry. We fit the data with our method and the results of the fitting are then post-processed to derive physical quantities, such as stellar mass, star formation rate, starburst timescale, AGN fraction, dust masses etc., which are essential for studying galaxy evolution. In this contribution we will discuss these results and how SMART can be used to study galaxies and obscured supermassive black holes at any redshift. Our method promises to be very useful for analysing multiwavelength SEDs of galaxies and obscured AGN from a number of ground-based and space-born facilities, such as Spitzer, Herschel and JWST.

Title: Service point placement by customer Geolocation Clustering

Constantia Malekkou[*] and Demetris Trihinas

University of Nicosia

[*]Speaker's email: malekkou.c@live.unic.ac.cy

Determining the location of a physical service point (i.e., store branch) is of outmost importance for the success of a retailer company that relies on a distributed branch network to promote its goods and services. The distance to the nearest service point is an important factor to consider, as it is a measure of convenience for its customers. The initial set-up costs required for the establishment of a new service point should be carefully weighed against the expected benefits, as future relocations or new establishments translate to additional, unnecessary costs. In this study, we extract data from the Cyprus National Open Data portal in reference to Cyprus-based legal entities to determine both the location and size that the service points of a retailer company should be placed to serve the population. To achieve this, we model the challenge as a maximal covering location problem and use unsupervised clustering to develop our solution. The results of our solution are the optimal number and geographical positions of service points for a retailer, based on the actual point geolocations of the retailer's customers. We statistically evaluate our results by comparing against the service point placement of two competitor companies. We show that the service points proposed by the developed model minimize the distance-to-be-travelled by the customers to the nearest service point, while suggesting a reduced number of service points, resulting to better customer service with reduced operating costs.

Τίτλος: Καινοτόμος δυναμικός προβλεπτικός αλγόριθμος για την εξατομίκευση της πρόβλεψης του χρόνου πρώτης θεραπείας στη χρόνια λεμφοκυτταρική λευχαιμία

Θ. Μωυσιάδης[1]*, Δ. Κοπαράνης[1], Κ. Λιάπης[1], Μ. Γανοπούλου[2], Γ. Βραχιώλιας[1], Ι. Κατάκης[3], Χ. Μωυσιάδης[4], Ι. Βιζιριανάκης[5], Λ. Αγγελής[2], Κ. Φωκιανός[6], Ι. Κοτσιανίδης[1]

[1]Αιματολογική κλινική, Πανεπιστημιακό Γενικό Νοσοκομείο Αλεξανδρούπολης, Δημοκρίτειο Πανεπιστήμιο Θράκης, 68100, Αλεξανδρούπολη
[2]Τμήμα Πληροφορικής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 54124 Θεσσαλονίκη
[3]Τμήμα Πληροφορικής, Πανεπιστήμιο Λευκωσίας, Λευκωσία, 2417
[4]Τμήμα Μαθηματικών, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 54124 Θεσσαλονίκη
[5]Τμήμα Φαρμακευτικής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 54124 Θεσσαλονίκη
[6]Τμήμα Μαθηματικών & Στατιστικής, Πανεπιστήμιο Κύπρου, 2109 Λευκωσία

*Email ομιλητή: moysiadis.theodoros@gmail.com

Η εξατομικευμένη πρόβλεψη αποτελεί ιδεατό στόχο στη χρόνια λεμφοκυτταρική λευχαιμία (ΧΛΛ). Έχει καταβληθεί συστηματική προσπάθεια για την ανάπτυξη εξελιγμένων μοντέλων κατηγοριοποίησης των ασθενών σε ομάδες κινδύνου. Ωστόσο, τα μοντέλα αυτά χρησιμοποιούν κυρίως πληροφορία κατά τη διάγνωση ή/και προ έναρξης θεραπείας, χωρίς να ενσωματώνουν τις χρονοεξαρτώμενες μεταβολές κατά την παρακολούθηση του ασθενούς, με αποτέλεσμα να παρατηρούνται μεγάλες αποκλίσεις στην πρόβλεψη ακόμα και σε ασθενείς της ίδιας ομάδας κινδύνου. Η μελέτη αυτή στοχεύει να συνεισφέρει στην εξατομικευμένη πρόβλεψη στη ΧΛΛ, προτείνοντας έναν δυναμικό προβλεπτικό αλγόριθμο του χρόνου πρώτης θεραπείας (ΧΠΘ) για τον μεμονωμένο ασθενή.

Ο αλγόριθμος συμπεριλαμβάνει τον υπολογισμό ενός Σκορ, που υπολογίζεται για κάθε ασθενή σε κάθε χρονικό στιγμιότυπο (επίσκεψη στο νοσοκομείο), και αξιοποιεί το σύνολο των χρονοεξαρτώμενων καταγραφών του ασθενούς (αιματολογικές/βιοχημικές), σχηματίζοντας μία συνεχή χρονοσειρά Σκορ (ΣΧΣ). Οι υψηλότερες τιμές του Σκορ αντιπροσωπεύουν μεγαλύτερη επιδείνωση νόσου σε σχέση με τις χαμηλότερες τιμές. Ο αλγόριθμος περιλαμβάνει τα ακόλουθα κύρια βήματα: (α) Ανάπτυξη μίας βάσης αναφοράς ΧΛΛ-ασθενών που αναλύονται αναδρομικά. Σε ασθενείς με θεραπεία, υπολογίζεται η ΣΧΣ σε όλα τα χρονικά στιγμιότυπα από τη διάγνωση έως το ΧΠΘ, και εκτιμάται ένα κατώφλι που μετατρέπει τη ΣΧΣ σε δίτιμη με τιμές 0/1 για χαμηλότερο/υψηλότερο Σκορ, αντίστοιχα. Σε ασθενείς χωρίς θεραπεία (μάρτυρες), υπολογίζεται η ΣΧΣ από τη διάγνωση έως την τελευταία καταγραφή. (β) Για έναν «νέο» ασθενή που αναλύεται προοπτικά και δεν έχει λάβει θεραπεία στο στιγμιότυπο n (εδώ εκκινεί η διαδικασία πρόβλεψης), υπολογίζεται η ΣΧΣ από τη διάγνωση έως το n, και εντοπίζονται οι «πλησιέστεροι» ΧΛΛ-ασθενείς εντός της βάσης αναφοράς. Στο βήμα αυτό, μπορεί να αξιοποιηθεί και επιπρόσθετη βιολογική πληροφορία που αφορά τον «νέο» ασθενή, περιορίζοντας την ομάδα των «πλησιέστερων» ασθενών σε ασθενείς με παρόμοια βιολογικά χαρακτηριστικά. (γ) Εφαρμογή μοντέλων δίτιμων χρονοσειρών στους «πλησιέστερους» ασθενείς, και αξιοποίηση των αποτελεσμάτων για πρόβλεψη του ΧΠΘ του «νέου» ασθενούς, εφόσον πληρούνται οι απαραίτητες προϋποθέσεις.

Ο αλγόριθμος εφαρμόστηκε σε 20 ασθενείς (14 με θεραπεία, 6 μάρτυρες). Ο μέσος χρόνος μεταξύ στιγμιότυπων ήταν τρεις μήνες (υποτέθηκαν ισαπέχοντα). Καθένας από τους 14 ασθενείς με θεραπεία, αντιμετωπίστηκε ως «νέος» ασθενής και συγκρίθηκε με τους υπόλοιπους 19. Το n τέθηκε ίσο με 13. Η πρόβλεψη πρώτης θεραπείας πραγματοποιήθηκε σε 8/14 περιπτώσεις με 7/8 προβλέψεις να απέχουν λιγότερο από 4 χρονικά στιγμιότυπα σε σύγκριση με τον πραγματικό χρόνο. Στη συνέχεια, καθένας από τους 6 μάρτυρες, αντιμετωπίστηκε ως «νέος» ασθενής και συγκρίθηκε με τους υπόλοιπους 19. Βρέθηκε ότι, ορθώς, σε καμία από τις 6 αυτές περιπτώσεις ο αλγόριθμος δεν πραγματοποίησε πρόβλεψη του ΧΠΘ.

Συνοψίζοντας, ο προτεινόμενος αλγόριθμος μπορεί να αξιοποιήσει το σύνολο της πληροφορίας του ασθενούς (αιματολογικές/βιοχημικές εξετάσεις, βιολογικό υπόβαθρο), και να χρησιμοποιηθεί για την πρόβλεψη του χρόνου πρώτης θεραπείας υπό τις κατάλληλες προϋποθέσεις. Αντλώντας με δυναμικό τρόπο πληροφορία, και αξιοποιώντας τη διαθέσιμη γνώση, μπορεί να βελτιώσει την πρόβλεψη για τον μεμονωμένο ασθενή, προάγοντας την εξατομικευμένη πρόβλεψη στη χρόνια λεμφοκυτταρική λευχαιμία.

2. ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΥΓΕΙΑ – STATISTICS AND HEALTH

Title: Hierarchical Distributed Lag Models

Theo Economou

The Cyprus Institute

Email: t.economou@cyi.ac.cy

Distributed lag models (DLMs) are an extremely useful statistical modelling class for characterising the effects of a covariate and its temporally lagged values on the mean of a response. In essence, DLMs are regression models, albeit ones that involve constraints in the values the coefficients can take to avoid instability. In this work we argue that using smooth but penalised functions is possibly the closest to an objective way of constraining DLMs. Moreover, we show that when implemented as Generalized Additive Models (GAMs), DLMs can be extended to include the lagged effects of covariates as well as their interaction. In addition, we illustrate how to extend DLMs to allow hierarchical structures, in order to achieve pooling across discrete spatial units. The results are illustrated in quantifying the effect of heat stress (temperature and relative humidity) on mortality, utilising data from Cyprus and the city of Thessaloniki in Greece.

Title: **A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants**

Maria Zanti[1][*], Denise G. O'Mahony[1], Michael T. Parsons[2], Amanda B. Spurdle[2], Douglas F. Easton[3,4], David E. Goldgar[5], Kyriaki Michailidou[1,3]

[1] **Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus**
[2] **Population Health Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia**
[3] **Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK**
[4] **Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK**
[5] **Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA**

[*]**Speaker's email: mariaz@cing.ac.cy**

Clinical genetic testing for genetic syndromes identifies either variants with a pathogenic effect on gene function, neutral (benign) variants, or variants with uncertain effect on gene function. Following the recommendations of the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP), the frequency in case-control datasets (PS4 criterion), can inform the interpretation of these variants of uncertain significance (VUS). The specific criterion states that a Relative Risk (RR) or an Odds Ratio (OR) > 5.0, with nominal significance, can provide strong evidence in favor of pathogenicity. We present a novel case-control likelihood ratio-based method that following a survival analysis framework, incorporates gene-specific age-related penetrance to compare the likelihood of the distribution of the variant of interest among cases and controls, under the hypothesis that the variant is associated with similar risks of the disease in question, as the "average" pathogenic variant, compared to the likelihood under the hypothesis that it is a benign variant not associated with increased risk. We demonstrate the utility of this method in the analysis of simulated and real datasets. In the analyses of simulated data, the likelihood ratio method was more powerful compared to other methods. Likelihood ratios were calculated for a case-control dataset (75,657 breast cancer cases and 52,987 healthy controls) of BRCA1 and BRCA2 variants from the Breast Cancer Association Consortium (BCAC), and compared with logistic regression results. A larger number of variants reached evidence in favor of pathogenicity, and a substantial number of variants had evidence against pathogenicity - findings that would not have been reached using other case-control analysis methods. Our novel method provides greater power to classify rare variants compared to classical case-control methods. We provide user-friendly scripts and pre-formatted Excel calculators for the implementation of the method for rare variants in BRCA1, BRCA2, and other high-risk genes with known penetrance.

Title: Polygenic risk score differences among European ancestry populations; implications for breast cancer risk prediction

Kristia Yiangou[1*], Nasim Mavaddat[2], Joe Dennis[2], Andreas Hadjisavvas[3], Maria A. Loizidou[3], Jacques Simard[4], Antonis C. Antoniou[2], Douglas F. Easton[2,5] and Kyriaki Michailidou[1] on behalf of the Breast Cancer Association Consortium.

[1]Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
[2] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[3] Department of Cancer Genetics, Therapeutics and Ultrastructural Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
[4] Genomics Center, Centre Hospitalier Universitaire de Québec—Université Laval Research Center, Québec City, Canada
[5] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

[*]**Speaker's email: kristiay@cing.ac.cy**

Breast cancer is the most common cancer among females worldwide. Disease risk differs between individuals and is influenced by environmental and genetic risk factors. To-date, multiple genetic variants have been discovered to contribute to breast cancer susceptibility, including common, low-risk variants identified via genome-wide association studies (GWAS). Polygenic risk scores (PRSs), combine additively the weighted effects of multiple common low-risk variants, and can be used for the prediction of breast cancer risk at an individual's level. Thus, PRSs can be potentially used in targeted preventative strategies and population screening. Recently a 313 variants PRS (PRS313) has been constructed for the classification of women of European ancestry based on their breast cancer risk. However, specific evaluation of the distribution of the mean PRS313 across the different European populations has not been evaluated. Potential differences in this distribution if are ignored when defining risk categories based on the PRS, could influence the classification of some individuals and thus affect their clinical management.

In this study, we aimed to explore the distribution of the mean PRS313 across 21 countries using data from 111,814 female breast cancer cases and 94,718 female controls from 84 studies participating in the Breast Cancer Association Consortium (BCAC). All samples were previously genotyped using the iCOGS or the OncoArray platforms. PRS313 was initially calculated in each individual and then the mean PRS313 was calculated by country separately in cases and controls. To explore for potential sources of the distribution variability across the countries, we recalculated the mean PRS by country excluding CHEK2 correlated variants, those with the most variable frequency in controls, and finally, by adjusting for the ancestry informative principal components (PCs). In addition, an empirical Bayes approach was used as an alternative method to derive country-specific estimates.

The mean PRS313 differed markedly across the countries of Europe with heterogeneity I2=80% (p-value0.05). Alternatively, when an empirical Bayes approach was used, the mean PRS313 for countries with small available sample size had greater shrinkage towards the overall estimate.

These results, indicated that the distribution of the mean PRS313 differs even within Europe leading to an overestimation of the risk in east-south Europe and underestimation of risk in western Europe if not appropriately accounted for. Therefore, the implementation of the PRS in risk prediction will require country-specific evaluation. Principal components seem to correct the distribution differences in the controls. The empirical Bayes approach can be alternatively used, as it considers both the true variability in the mean PRS across the countries and also the uncertainty due to the small available sample size.

3. ΦΥΛΟ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗ – GENDER AND STATISTICS

Τίτλος: Συμμετοχή Γυναικών σε Εθνικά Προγράμματα Έρευνας και Καινοτομίας

Μαριλένα Παρασκευά

Ίδρυμα Έρευνας και Καινοτομίας (ΙδΕΚ)

Email: marilena@research.org

Μέσα από την σύντομη αυτή παρουσίαση οι παρευρισκόμενοι/ες θα έχουν την ευκαιρία να ενημερωθούν για την συμμετοχή των γυναικών σε προγράμματα Έρευνας και Καινοτομίας κατά την τελευταία επταετία με ιδιαίτερη αναφορά στις θεματικές περιοχές των Προγραμμάτων RESTART 2016-2020. Επιπρόσθετα θα παρουσιαστούν ενέργειες που έχουν υλοποιηθεί από το ΙδΕΚ με σκοπό να βοηθήσουν στην μεγαλύτερη εμπλοκή των γυναικών στα Προγράμματα Έρευνας και Καινοτομίας του ΙδΕΚ.

Title: Evaluation of Gender Parity in the Context of Gender Representation and Collaboration in the Cyprus Institute of Neurology and Genetics

Margarita Zachariou [*], Stavroulla Xenophontos, Pavlos Polycarpou, Elena Ioannidou, Vera Kazandjian, Maria Lagou, Anna Michaelidou, George M. Spyrou, Marios A. Cariolou and Leonidas Phylactou

Cyprus Institute of Neurology and Genetics

[*]Speaker's email: margaritaz@cing.ac.cy

Females are underrepresented in the science, technology, engineering, mathematics and medicine (STEMM) disciplines globally and although progress has been made, the gender gap persists. Our aim was to explore gender parity in the context of gender representation and internal collaboration at the Cyprus Institute of Neurology and Genetics (CING), a leading national biomedical organisation accredited as an equal opportunity employer. Towards this aim we (1) explored trends in gender parity within the different departments, positions and qualifications and in student representation in the CING's postgraduate school and, (2) investigated the degree of collaboration between male and female researchers within the Institute and the degree of influence within its co-authorship network. We recorded an over-representation of females both in the CING employees and the postgraduate students. The observed female over-representation in pooled CING employees was consistent with a similar over-representation in less senior positions and was contrasted with an observed male over-representation in only one middle rank and culminated in gender equality in the top rank in employee hierarchy. In terms of collaboration, both males and females tended to collaborate with each other without any significant preference to either inter-group or intra-group collaboration. Further comparison of the two groups with respect to their influence in the network in terms of occupying the positions of highest centrality scores, indicated that both gender and seniority level (head vs non-head) were significant in shaping the authors' influence, with no significant difference in those belonging in the same seniority level with respect to their gender. To conclude, our study has validated the formal recognition of the CING's policies and procedures pertinent to its egalitarian culture through the majority of the metrics of gender equality assessed in this study and has provided an extendable paradigm for evaluating gender parity in academic organisations.

Title: Practices to Achieve Gender Balance in Academia

Petroula Mavrikiou

Frederick University

Email: p.mavrikiou@frederick.ac.cy

Even though gender balance awareness is rising, women are still underrepresented in academia. This is even more visible in male-dominated fields such as Computer Science or, more generally in STEM (science, technology, engineering, and mathematics) areas. Women in STEM often face career obstacles that are hard to balance. On the other hand, research shows that diverse groups are better at solving problems, an essential factor in the research.

The primary purpose of this work is to present the preliminary results of the data gathered in the online "Survey on best practices in gender balance in Informatics". The survey was addressed to faculty and staff of Higher Education Institutions (HEI) and Research Organisations (RO) in Europe and seeks to collect best practices regarding recruiting, promoting, and retaining women in academia. The survey was distributed in 2022 and received 57 valid responses from representatives from different institutions. The results show that the institutions more often use recruiting practices compared to promoting and retaining women at their institutes.

<u>**Title**</u>**: Mind the *gender* gap!**

Haritini Tsangari

University of Nicosia

Email: tsangari.h@unic.ac.cy

The presentation aims to illustrate the multidimensionality of gender inequality, which still exists today, worldwide. Three dimensions of the gender gap are considered.

First, the gender gap in education is presented, focusing on areas of under- and over-representation of women and the changes that the COVID-19 pandemic has brought to conventional and online learning in different countries.

Second, the gender pay gap is discussed. Despite the legislations on equal pay, women are still paid less than men, even after adjusting for education and qualifications. Related indexes and statistics from various regions are presented.

Third, special attention is given to the gender gap in leadership. The number of women in managerial positions has increased significantly over the last few, however the rate of progress is slow and uneven. Evidence from international indexes and official statistics is provided. In addition, two studies conducted in Cyprus are presented. More specifically, our results showed the difficulty of women to break the glass ceiling, the barriers that they face in attaining managerial positions, the challenges that they confront if they are in higher ranks in their organizations and the stereotypes that need to be overcome. A comparison between male and female employees' opinions is also given.

The road to closing the gender gap is still long and more directed efforts are necessary.

4. ΚΟΙΝΩΝΙΚΗ ΣΤΑΤΙΣΤΙΚΗ – SOCIAL STATISTICS

**Title: The impact of the COVID-19 pandemic on perceived fear of future pandemics:
A cross-sectional study**

**Romina Alexandrou[1], Maria Kyprianidou[1], Galatia Photiou[2], Angelos P. Kassianos[3] and
Konstantinos Giannakou[1] ***

**[1] Department of Health Sciences, School of Sciences, European University Cyprus, Nicosia, Cyprus
[2] Cyprus International Institute for Environmental and Public Health, School of Health Sciences,
Cyprus University of Technology, Limassol, Cyprus
[3] Department of Nursing, School of Health Sciences, Cyprus University of Technology, Limassol,
Cyprus**

***Speaker's email: K.giannakou@euc.ac.cy**

Background: The COVID-19 pandemic has not only affected global health, economy, and society but has also led to the development of fear due to various factors. The negative impacts of prolonged and intense fear on mental and physical health highlight the importance of understanding the reasons behind it during pandemics. This study aims to examine the impact of COVID-19 pandemic on the development of fear of future pandemics among the general population of Cyprus.

Methods: A cross-sectional study was conducted using a proportionate quota sampling method according to five government-controlled municipalities, sex, and age. The study collected information on sociodemographic and health-related characteristics, health literacy, trust in vaccines, public health authorities, medicine, and science, COVID-19 vaccination information, and fear of future epidemics or pandemics using a self-administered questionnaire.

Results: The survey included 1075 participants, with 53.7% of them reporting fear of future pandemics. Women (OR=2.37, 95% CI: 1.78-3.16) and those who were vaccinated against COVID-19 (OR=1.57, 95% CI: 1.02-2.43) were significantly more likely to be afraid of future pandemics. Participants with higher levels of trust (OR=1.04, 95% CI: 1.02-1.06) and lower health literacy (OR=1.05, 95% CI: 1.03-1.08) were also significantly more likely to fear future pandemics. Conversely, being unemployed (OR=0.30, 95% CI: 0.13-0.65) and having a postgraduate education decreased the likelihood of fearing future pandemics (OR=0.56, 95% CI: 0.34-0.90).

Conclusions: The COVID-19 pandemic has led to the development of fear of future pandemics among a significant proportion of the population of Cyprus. Public health messaging should be targeted to address the specific concerns of different groups of individuals, with a particular focus on improving health literacy to effectively manage fear during pandemics.

Title: Medical cannabis: establishing its roots in pharmaceutical marketing

Christos Ntais[1][*], Jean Suvatjis[2] and Yioula Melanthiou[3]

[1]Doctoral Researcher, School of Business, University of Nicosia, Cyprus
[2] International Brand Marketing Consultant, Athens, Greece
[3]Department of Public Communication, Faculty of Communication and Media Studies,
Cyprus University of Technology, Limassol, Cyprus

[*]Speaker's email: chrisntais@yahoo.gr

This research investigates whether cannabis and its extracts could be considered a medicinal product for people with specific diseases and how marketing and branding could contribute to such product's acceptance by consumers as an alternative medical solution.

Our methodology involved a set of research hypotheses covering a range of medical cannabis-related issues, such as adequacy of scientific evidence in support of its medicinal use, social taboos, segmentation-targeting-positioning strategies, branding and brand architecture considerations, and social marketing approaches. The sample size of this quantitative study included 385 respondents (285 Individuals and 100 Business Managers). The 285 Individuals were selected randomly during one month in Athens city center, Greece, using skip interval - every 120th individual, using systematic probability sampling-population numbering. The 100 Business Managers were selected from the official designated list of a professional database of pharmaceutical companies headquartered in Athens. The survey sample size, according to MIL-STD-105 D / ISO 2859, satisfied the statistical sample size requirement. The ten-point Likert-type rating scale served as the primary research instrument. The analysis focused on the level of statistical significance of the factors involved in the attempt to conclude whether the respondents in each factor group agree or disagree (and to what degree) with each statement of the research instrument. Special "normality tests" were applied.

Furthermore, "non-parametric statistics" were applied to a field that is not bound by the validity conditions of the Normal Distribution. The appropriate technique for this case was the Sign Test for Median which was applied to the response data of the 285 Individuals and the response data of the 100 Business Managers. Finally, parametric statistics, such as the analysis of variance, which allows the simultaneous analysis of many factors, was used to accept or reject the hypotheses.

Our results indicate that cannabis has the potential to be considered as a medical alternative for specific health conditions; however, more research is needed to confirm its medicinal benefits. In addition, social issues related to medical cannabis, such as negative stereotyping, taboos, and stigmatization, need to be overcome, changed, and transformed to motivate thorough research of the use of cannabis for potential medicinal purposes; branding and marketing should be used as tools for the differentiation of cannabis and its extracts within the highly competitive pharmaceutical industry.

In conclusion, our research revealed that the destigmatization of medical cannabis is critical for its transformation into an alternative medical solution. Therefore, a tri-stage process is considered for destigmatizing medical cannabis. The contribution of social marketing, acting as an ambassadorial agent to attain a consensus and to acquit medical cannabis, unfolding its real potential in medicine, is also emphasized. Finally, a strategic marketing approach and an appropriate brand architecture are suggested in support of medical cannabis' candidacy to become a pharmaceutical brand.

Title: The role of conspiracy theories, perceived risk, and trust in science on COVID-19 vaccination decisiveness: evidence from Cyprus.

Marilena Mousoulidou, Andri Christodoulou, Michailina Siakalli* and Marios Argyrides

Neapolis University Pafos

*Speaker's email: michailina27@gmail.com

COVID-19 reminded us of the importance of vaccinating for successfully overcoming health-related crises. Yet, vaccine hesitancy is still present. The study that will be presented examined the impacts of conspiracy theories, perceived risk, and trust in science on COVID-19 vaccination decisiveness. The study was conducted at the end of the third wave of the pandemic, in July 2021, in Cyprus. Data were collected via an online self-administered anonymous survey using convenience and snowball sampling methods. Participants were 363 adults who completed a set of questionnaires that examined their believability in ten vaccine-related conspiracy theories, their perceived dangerousness of COVID19, and their level of trust in science and scientists. The results suggest that (a) participants with a high conspiracy theory belief are less likely to be vaccinated, (b) participants who perceive COVID-19 as a dangerous disease are more likely to be vaccinated, and (c) participants with high trust in science are more likely to be vaccinated. The implications of the findings are discussed and can be used by public health officials in their campaigns.

5. ΕΚΠΑΙΔΕΥΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ – EDUCATIONAL STATISTICS

Title: Persistence in test-taking. New insights from examining timing data

Elena Papanastasiou* and Evi Konstantinidou

University of Nicosia

***Speaker's email: papanastasiou.e@unic.ac.cy**

The digital transition in large-scale assessments generated a plethora of new variables that can be used in relation to test scores use and interpretation. One such category of variables is based on timing data. The study aims to present two indicators of examinee test-taking persistence (i.e., successful persistence and unsuccessful persistence) that are based on timing data, and examine their relationship with student background variables, as well as with examinee performance in mathematics and science. Large-scale, representative data from e-TIMSS (Trends in International Mathematics and Science Study) 2019 and from 4th graders in the USA who participated in both mathematics and science assessments were used. Based on the preliminary results of the study, the unsuccessful persistence indicator was more highly correlated with achievement, indicating that students with lower levels of achievement ended up spending more time than average on items that were eventually answered incorrectly; whereas higher achieving students had lower amounts of unsuccessful persistence in incorrect answers. Our study departs from examining only whether a student answered correctly or incorrectly and sheds light on the process of answering an item, including the degree of persistence an examinee demonstrates.

Title: Mathematics self-concept: Investigating the Immigrant Paradox with TIMSS 2019 data from Germany

Ioulia Televantou[1]*, Rolf Strietholt[2] and Ingrid Gogolin[3]

[1]European University Cyprus
[2]Research & Analysis Unit, IEA, Hamburg
[3]University of Hamburg

*Speaker's email: i.televantou@euc.ac.cy

Recent (first-generation) immigrants have better motivational outcomes than native and later (second-generation) immigrants, despite the considerable barriers they face when moving to another country (e.g., speaking a different language than that of instruction, having access to lower economic resources). This phenomenon is referred to in the literature as the "immigrant paradox." The scope of the present study is to investigate the prevalence of the paradox with TIMSS 2019 data from Germany and identify factors explaining the gap in mathematics self-concept between students with immigration backgrounds and native students in Germany.

Academic self-concept (ASC) refers to the specific component of self-concept that denotes how individuals perceive their academic abilities and competencies in a specific subject. ASC has been valued as an educational outcome in its own right, but it has also been shown to be positively related to other educational outcomes. For example, higher ASC has been linked to higher academic achievement (Reciprocal Effects Model; REM). However, until now, very few studies have examined the immigrant paradox concerning mathematics self-concept, despite this phenomenon being investigated concerning other motivational outcomes (e.g., educational aspirations and expectations).

Our sample consisted of 3437 grade four primary school students from Germany, 6.3% of whom are first-generation immigrants and 11.3% second-generation immigrants. For the operational definition of ASC, four items were selected from the mathematics attitudes scale (Chen & Hastedt, 2022; Marsh et al., 2014; 2015); Cronbach's alpha was higher than .8, denoting adequately high reliability for all measures. In our models, we controlled for the students' socioeconomic status (SES), defined as the highest parental educational level in years. Multilevel analysis was used, with level 1 being the student and level 2 being the school.

Our findings suggest that the immigrant paradox is prevalent with Mathematics self-concept data from Germany: Given the same mathematics achievement, first-generation immigrants have, on average higher mathematics self-concept than their native peers. Even though SES does predict mathematics self-concept in a positive and statistically significant way, it does not seem to influence the estimated effect of immigration status on mathematics self-concept. Higher SES predicts higher mathematics self-concept for all students (native & students with immigration backgrounds). Contrary to conventional wisdom, attending a school with higher mean achievement or higher mean socioeconomic status does not positively affect students' mathematics self-concept (controlling for student mathematics achievement and socioeconomic status). Controlling for the proportion of students with an immigration

background in the school led to the disappearance of the immigrant paradox. Hence the proportion of students with an immigration background in the school could be a mechanism leading to the occurrence of the paradox, given theoretical justification. Other factors potentially explaining the immigrant paradox should be explored.

Title: E-learning adoption and Academic Self-Efficacy: An Application of the Technology Acceptance Model

Danxia Chen[1*], Ioulia Televantou[1] and Elena Papanastasiou[2]

[1]European University Cyprus
[2]University of Nicosia

***Speaker's email: dc222448@students.euc.ac.cy**

The present study uses cross-sectional survey data from a sample of 430 first-year postgraduate students at a Distance Learning program of a private university in Cyprus, evaluating the Technology Acceptance Model (TAM, Davies et al., 1989) in the context of e-learning adoption in higher education. According to TAM, students' attitudes toward a particular e-learning system depend on two key factors: perceived usefulness (PU) and perceived ease of use (PEOU). PU is the degree to which a person believes using a specific system can facilitate their learning. In contrast, PEOU of service is the degree to which an individual acknowledges that using a particular technology does not imply extra effort. Our study validates the TAM using Structural Equation Modelling (SEM) and, Mplus statistical package.

Furthermore, it shows that PU and PEOU are negatively predicted by students' computer anxiety and positively predicted by their perceived quality of technical support. Crucially, our study demonstrates the importance of the student's attitudes towards the learning management system in predicting their Academic Self-Efficacy (ASE): failing to adjust to the e-learning system (e.g., due to high computer anxiety) negatively impacts students' ASE. This might, in turn, hinder students' academic achievements in the course studied – in our study, Quantitative Research Methods. Given the broad literature (Balloo, 2019; Murtonen & Lehtinen, 2003) on the multiple difficulties that higher education students experience in their learning of quantitative methodology, investigating how such difficulties may be manifested in an online learning environment is highly relevant to exist research and practice.

6. <u>ΣΤΑΤΙΣΤΙΚΗ ΘΕΩΡΙΑ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ – STATISTICAL THEORY AND METHODOLOGY</u>

<u>**Title**</u>**: A data-adaptive algorithm for multiple change-point detection**

Sophia Loizidou[1][*] and Andreas Anastasiou[2]

[1]University of Luxembourg
[2]University of Cyprus

[*]Speaker's email: sophia.loizidou@uni.lu

In this talk, a new data-adaptive method, called DAIS (Data Adaptive ISolation), is introduced for the estimation of the number and the location of change-points in a given data sequence. The proposed method can detect changes in various different signal structures; we focus on the examples of piecewise-constant and continuous, piecewise-linear signals. We highlight, however, that our algorithm can be extended to other frameworks, such as piecewise-quadratic signals. The data-adaptivity of our methodology lies in the fact that, at each step, and for the data under consideration, we search for the most prominent change-point in a targeted neighbourhood of the data sequence that contains this change-point with high probability. Using a suitably chosen contrast function, the change-point will then get detected after being isolated in an interval. The isolation feature enhances estimation accuracy, while the data-adaptive nature of DAIS is advantageous regarding, mainly, computational complexity and accuracy. The simulation results presented indicate that DAIS is at least as accurate as state-of-the-art competitors.

Title: Assessing the effectiveness of indirect questioning techniques by detecting liars

Pier Francesco Perri[1], Eleni Manoli[2] and Tasos C. Christofides[2]

[1]University of Calabria
[2]University of Cyprus

*Speaker's email: eleni_manoli55@yahoo.com

In many fields of applied research, mostly in sociological, economic, demographic and medical studies, misreporting due to untruthful responding represents a nonsampling error that frequently occurs especially when survey participants are presented with direct questions about sensitive, highly personal or embarrassing issues. Untruthful responses are likely to affect the overall quality of the collected data and flaw subsequent analyses, including the estimation of salient characteristics of the population under study such as the prevalence of people possessing a sensitive attribute. The problem may be mitigated by adopting indirect questioning techniques which guarantee privacy protection and enhance respondent cooperation. In this paper, making use of direct and indirect questions, we propose a procedure to detect the presence of liars in sensitive surveys which allows researchers to evaluate the impact of untruthful responses on the estimation of the prevalence of a sensitive attribute. We first introduce the theoretical framework, then apply the proposal to the Warner randomized response method, the unrelated question model, the item count technique, the crosswise model and the triangular model. To assess the effectiveness of the procedure, a simulation study is carried out. Finally, the presence and the amount of liars is discussed in two real studies concerning racism and workplace mobbing.

Τίτλος: **Διαδικασία αξιολόγησης της εφαρμογής του Κώδικα Ορθής Πρακτικής για τις ευρωπαϊκές στατιστικές**

Κώστας Διαμαντίδης

Στατιστική Υπηρεσία

Email: cdiamantides@cystat.mof.gov.cy

Ο Κώδικας Ορθής Πρακτικής για τις ευρωπαϊκές στατιστικές αποτελεί το πλαίσιο που διασφαλίζει την ποιότητα των στατιστικών που παράγονται από το Ευρωπαϊκό Στατιστικό Σύστημα (ΕΣΣ). Με τις Επιθεωρήσεις Ομοτίμων επιτυγχάνεται η αξιολόγηση της συμμόρφωσης του κάθε μέλους του ΕΣΣ με τον Κώδικα. Το αντικείμενο της παρουσίασης αφορά την περιγραφή του Κώδικα Ορθής Πρακτικής και τις εμπειρίες που αποκτήθηκαν από τη διαδικασία αξιολόγησης της εφαρμογής του Κώδικα στην Κύπρο.

4th

PANCYPRIAN
CONFERENCE
IN STATISTICS
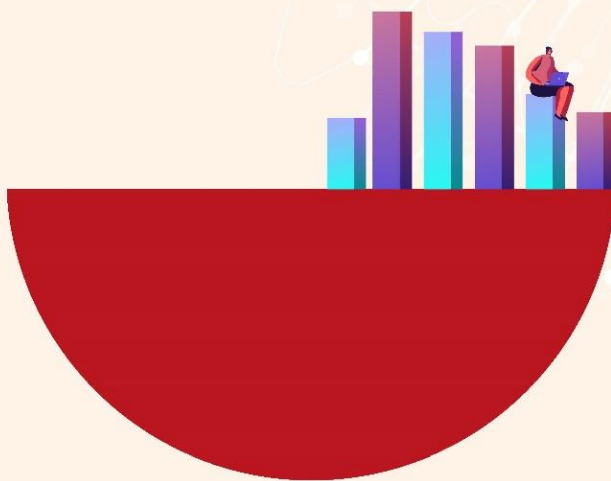
University of Nicosia

Friday, June 9, 2023
9:00 - 18:00

Jean Monnet Amphitheater
Da Vinci and Newton Amphitheaters

Co-organizers:

ΚΥΠΡΙΑΚΗ ΣΤΑΤΙΣΤΙΚΗ ΕΤΑΙΡΕΙΑ

UNIVERSITY of NICOSIA

Sponsors:

ClinBAY
Statistical solutions to drug development

pulse
market research