9 Estadística y Estocasticidad

1. **Expositor:** Dr. Guillermo Ferreira
   Afiliación: Departamento de Estadística, Universidad de Concepción, Concepción, Chile
   Título: Bootstrapping Regression Models with Locally Stationary Disturbances
   Resumen: A linear regression model with errors following a time-varying process is considered. In this class of models the smoothness condition both in the trend function and in the correlation structure of the error term ensures that these models can be locally approximated by stationary processes, leading to a general class of linear regression models with locally stationary errors. We focus here on the bootstrap approximation to the distribution of the least squares estimator (LSE) for such class of regression models. We compare and discuss the results on both the classical and bootstrap confidence intervals through an intensive simulation study. The trend is also discussed through a real data analysis on time series of monthly inflation in US with locally stationary errors.

2. **Expositor:** Jorge Figueroa
   Afiliación: Departamento de Estadística, Universidad de Concepción, Concepción, Chile
   Título: Trapezoidal Distributions and Trapezoidal beta Distribution
   Resumen: Beta Models have shown to be a powerful tool in the analysis and modeling of limited-range continuous variables. However, in many cases in practice, the data are bounded and tail-area events occur at either end or at both ends independently. To model this scenario, in this work we propose the trapezoidal beta (TB) model and we extend the construction to trapezoidal distributions (TD) using any bounded distribution. This work is centered on the development of the TB model, which is a generalization of both, the beta model and the rectangular beta model. We study its probability density function and we derive some fundamental properties such as the moments, moment generating function and characteristic function, then we extend this properties to TD. We show that, the TB model can be rewritten conveniently as a mixture of specifics beta distributions and that any bounded distribution can acquire the trapezoidal properties if this is mixtured with two specifics four-parameters beta distributions. In any case, we will add only two new parameters at the original distribution with a very intuitive interpretation. In addition, we show that the parameters in any TD can be easily estimated using the EM algorithm. For the TB distribution, we report both results of a simulation study, and an application to a real data set. Model fitting comparisons with several alternative models indicates that the model proposed presents the best fit and so it can be quite useful in real applications.

3. **Expositor:** Inés M. Varas
   Afiliación: Departamento de Estadística, Pontificia Universidad Católica

de Chile, Santiago, Chile
Título: Equipercentile equating under the NEAT design: the latent equating case

Resumen: Equating methods are a family of statistical methods used to adjust scores on different test forms so that scores can be comparable and used interchangeably [1]. These methods lie on functions to transform scores on two or more versions of a test. More specifically, the equating transformation [2] must be defined. The most popular one is the equipercentile function defined as $\varphi(x) = F_Y(F_X^{-1}(x))$ where F_X and F_Y are the cumulative distribution function (cdf) of the test scores X and Y , respectively. In the equating literature, the proposed approaches for the estimation of the equipercentile function are based on continuous approximations of the test scores distributions (procedure named as continuization step) because most of the time test scores are defined on discrete scale scores. However, there are some drawbacks with equated scores obtained under continuization: they could be not only out of the range of the scale score they are defined but also they are no longer integer equated scores. Considering scores as ordinal random variables [4] and [3] have proposed the Latent equating method to tackle the discreteness of scores in test equating procedures. Authors propose to perform the equipercentile transformation considering a Bayesian nonparametric model for the latent representation of the scores distributions and define a procedure to obtain discrete equated scores.

Before to obtain comparable scores from the estimated equating transformation, it is necessary to control for test takers ability differences. Several data collection designs have been described in the equating literature to this purpose [5]. These equating designs differ in that either common persons or common items are used to perform the score transformation. The non equivalent groups with anchor test (NEAT) design is widely used in test equating. Under this design, two groups of test takers are administered separate test forms with each test form containing a common subset of items. Authors in [4] and [3] proposed the Latent equating method only for independent groups of test takers (equivalent group design). In this project we extend the Latent equating method to other equating designs with emphasis on the results of the NEAT design. Several methods to evaluate the performance of the method are discussed and applied in both a simulation study and in a real dataset.

References

[1] Kolen, M. J., Brennan, R. L. *Test equating, scaling, and linking: Methods and practices.* (3rd edition). New York: Springer, (2014).

[2] Gonzalez, J., Wiberg, M., *Applying test equating methods: Using R.* Springer, 2017.

[3] Varas, I. M., González, J., Quintana, F. A., *A Bayesian non-parametric latent approach for score distributions in test equating*, Journal of Educational and Behavioural Statistics (2020).

[4] Varas, I., González, J., Quintana, F. A., *A new equating method through latent variables*, In M. Wiberg, S. A. Culpepper, R. Janssen, J. González, D. Molenaar (Eds.), Quantitative Psychology, 343-353. Cham: Springer. (2019).

[5] von Davier, A. A., Holland, P., Thayer, D., *The kernel method of test equating*, NewYork: Springer, (2004).

4. **Expositor:** Eduardo Alarcón-Bustamante
Afiliación: Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile
Título: Measuring the effect of an explanatory random variable in partitioned populations
Resumen: We define the concept of Global Marginal Effect as the marginal effect observed through the conditional expectation of a partitioned population. The related mathematical expression is characterised by the law of total probability and is shown to be a function of the independent variable. We discuss on the interpretation of the new measure of global marginal effect and illustrate its use with a real data set involving scores form a university entrance test in Chile.

5. **Expositor:** Milan Stehlík
Afiliación: Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile and Department of Applied Statistics & Linz Institute of Technology, Johannes Kepler University in Linz, Linz, Austria.
Título: Data transformations and transfer functions for biological and ecological complexity
Resumen: Biological and ecological systems offer a lot of complexity which should be well understood before we can make valuable regulations. We will address both complex and extreme measurements from given systems. There is a necessity to classify appropriate learning mechanisms and define transfer functions and statistics. I will address learning mechanisms of data transformation and aggregation. In particular, I will introduce SPOCU transfer function [1] and provide some of its unique properties for processing of complex data. Statistical learning will be discussed and tuning of parameters of SPOCU based neural networks will be given. Attractive applications to biological systems e.g. statistical assessment of nutrition level of bacteria [2], or mass balance in the ecosystem of glaciers in Patagonia [3], or methane emissions from wetlands [4] will be addressed.

6. **Expositor:** Javiera Arias Gutiérrez
   Afiliación:
   Título: Identificando brechas de género en la academia: un enfoque analítico con Data Science
   Resumen: En este análisis, se examina una base de datos de proyectos de la Universidad de Concepción para identificar variables relacionadas con las brechas de género. El proceso incluye la limpieza e imputación de datos, visualizaciones bivariadas, y la construcción e interpretación de un modelo de regresión logística múltiple que predice si un proyecto sería liderado por un hombre o una mujer, seleccionando previamente las variables más influyentes en la clasificación. Este estudio aporta evidencia sólida sobre la efectividad de técnicas de Machine Learning para abordar problemas sobre brechas de género. Los resultados indican que los hombres lideran proyectos en áreas de estudio relacionadas con ciencias físicas, matemáticas y químicas, y ocupan la mayoría de los cargos de alta jerarquía. Además, se observó una presencia más destacada de los hombres en proyectos en la cima de sus edades y que tienen asignados más proyectos con mayor financiamiento en comparación con las mujeres

7. **Expositor:** Ignacio Vidal García
   Afiliación: Instituto de Matemáticas, Universidad de Talca, Talca, Chile
   Título: Errores no forzados en el problema de concordancias
   Resumen: La solución al problema de concordancias de Montmort (Montmort's matching problem) puede verse como la función de probabilidad del número de coincidencias en un experimento para evaluar la correcta clasificación de variables nominales. Adicionalmente al número de coincidencias, Vidal and de Castro (2021) presentaron una generalización de este problema que considera el orden cronológico en que se realizan las asignaciones y el conteo del número de errores no forzados. Estos autores llevaron a cabo un análisis Bayesiano del problema, pero en esta oportunidad encontramos una solución a este nuevo enfoque desde un punto de vista frecuentista. Encontramos la función de probabilidad bivariada del número de coincidencias y el número de errores no forzados. Además, encontramos la distribución marginal del número de errores no forzados y la expresamos en términos de los números de Stirling de segunda especie. Estudiamos propiedades fundamentales de las distribuciones encontradas como los momentos, simetría y la distribución marginal del número de errores forzados, entre otras. Como consecuencia práctica, proponemos un nuevo experimento y una nueva forma de recopilar y analizar los datos para evaluar la correcta clasi cación de variables nominales. Analizamos dos conjuntos de datos reales utilizando la metodología propuesta, donde se evalúa el aprendizaje de un niño con Trastorno del Espectro Autista.

References

[1] Vidal, I. and de Castro, M., *A Bayesian analysis of the matching problem*, Journal of Statistical Planning and Inference, **212**:194–200, (2021).