



Universidad Nacional de Córdoba
Facultad de Matemática, Astronomía,
Física y Computación

Jornada de Matemática

**Grupo de Probabilidad
y Estadística**

María Laura Nores

Córdoba, 2 de octubre de 2017

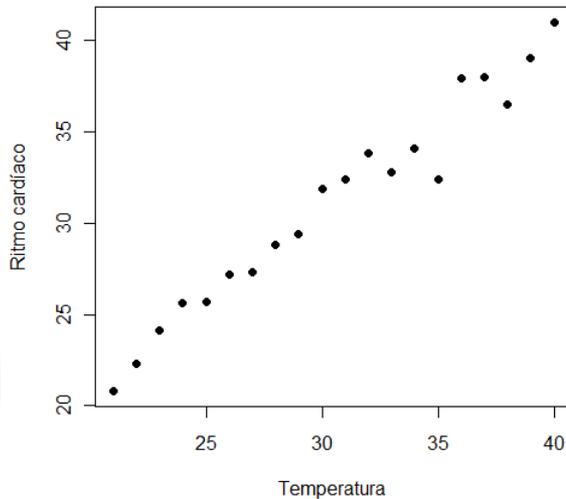
Integrantes del Grupo de Probabilidad y Estadística

- ✓ **Dr. Jorge G. Adrover** (Responsable del Grupo)
- ✓ **Mag. Patricia I. Bertolotto**
- ✓ **Lic. Grisel Britos**
- ✓ **Dr. Oscar H. Bustos**
- ✓ **Dra. Ana Georgina Flesia**
- ✓ **Dra. Aldana González Montoro**
- ✓ **Dra. María Laura Nores**
- ✓ **Dra. Silvia M. Ojeda** (Subresponsable del Grupo)
- ✓ **Dr. Jorge A. Sánchez**

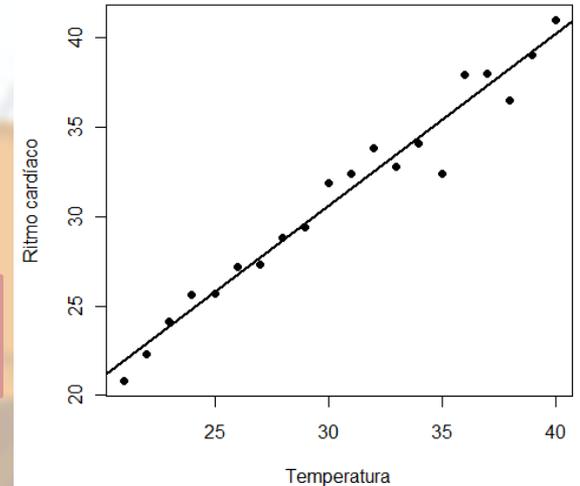
Modelo estadístico

Es una forma simplificada de representar la realidad, que usa la **probabilidad** e incluye un conjunto de **suposiciones** sobre la generación de las observaciones extraídas de una población.

Ejemplo: Modelo de Regresión Lineal Simple



$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$



$\{\varepsilon_i\}$ son variables aleatorias independientes, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

❖ Posibles extensiones:

- ✓ Más variables predictoras (x)
- ✓ Más variables de respuesta (Y)
- ✓ Observaciones repetidas en el tiempo
- ✓ Otra relación entre Y y las variables x
- ✓ Otras distribuciones

Algunas posibilidades en el campo de la Estadística:

- ❖ **Aplicación** de modelos y técnicas estadísticas a conjuntos de datos reales
- ❖ **Desarrollo** de nuevas metodologías
 - ✓ **Formalización** de la propuesta
 - ✓ Estudio de sus **propiedades teóricas**
 - ✓ **Implementación computacional**
 - ✓ Evaluación de su desempeño mediante **simulaciones**

Estadística Robusta

- ✓ Tiene como objetivo producir procedimientos estadísticos que puedan funcionar y cumplir sus objetivos predictivos aún cuando los supuestos del modelo estadístico que se usa para explicar el fenómeno se cumplen aproximadamente o sólo una proporción de los datos los cumplen.
- ❖ **Problemas de regresión** con alta cantidad de variables explicativas y escaso número de observaciones
 - ✓ Mejorar la **capacidad predictiva** considerando que las variables se acomodan en grupos con distintos niveles de influencia.
- ❖ **Análisis canónico**, que es una técnica de **análisis multivariado** que a partir de dos conjuntos de vectores, construye combinaciones lineales de cada conjunto que aumenten la **correlación** entre ambos conjuntos para poder explicar uno en términos del otro \longrightarrow **S-estimación**

- ❖ Series de Tiempo
- ❖ Estadística Espacial
- ❖ Modelado Estadístico de Imágenes
- ❖ Procesos ARMA Espaciales
- ❖ Métodos Robustos en Imágenes
- ❖ Similaridad de Imágenes



50	38	44	70	110	74	52	66	162	200	206
66	66	70	88	110	74	58	84	178	206	206
86	94	84	86	74	56	68	122	198	210	206
80	88	74	70	50	66	110	176	212	212	210
60	60	62	66	74	108	160	200	212	206	210
64	58	76	92	124	156	194	198	204	196	206
100	98	112	136	164	192	194	200	202	210	208
138	136	150	162	176	192	194	198	196	202	200



✓ **Proyecto de Investigación Actual:**
"Medidas de similaridad e índices de calidad para imágenes digitales. Teoría y aplicaciones".

Aprendizaje automático (*Machine learning*)

- ✓ Tiene como objetivo desarrollar técnicas que permitan a las computadoras “aprender”.
- ❖ Aplicado a problemas **de visión por computadoras y análisis de imágenes**.
 - ✓ **Clasificación semántica de imágenes** (¿qué hay?, ¿dónde?, ¿cuál es su contorno?)
 - ✓ **Aprendizaje supervisado**: Se cuenta con un banco de imágenes con etiquetas semánticas.
- ❖ Empezando a trabajar en **problemas multimodales** (mezcla de imágenes, texto, audio, etc.)

¿Qué se utiliza? \longrightarrow Inferencia estadística, optimización numérica, matemática computacional

- ❖ Desarrollar temas de **Estadística Matemática** que traten los aspectos formales de técnicas usadas en el **análisis estadístico de imágenes**.
 - ✓ Análisis estadístico y colaboración en la **interpretación de imágenes satelitales** (Instituto Gullich, CONAE-UNC).
 - ✓ Trabajo Final de Lic. en Computación de Francisco Trucco: Definir e implementar en forma computacionalmente eficiente un método de **clasificación en imágenes** que tiene en cuenta los píxeles "vecinos" de cada píxel (**ICM** o **medias condicionadas**).
 - ✓ Curso de posgrado en Geometría y Estadística (junto a Eduardo Hulett). Capacitación en "**Estereología en Estadística**". Se trata de los aspectos formales de Matemática y Estadística de técnicas tales como tomografía.

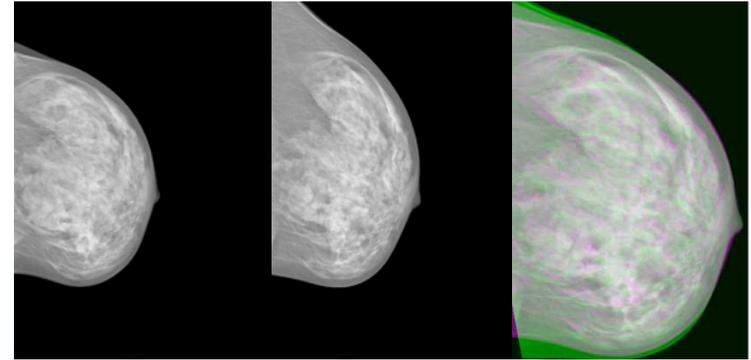
- ❖ Análisis estadístico de **Big Data** provenientes de **imágenes, filmaciones y secuencias**
- ❖ **Clasificación, etiquetado y determinación de patrones**
- ❖ Uso del **software R, Python y Matlab** en estadística para Big Data
- ❖ **Modelos estocásticos y algoritmos** para cálculo de estadísticos

Últimos trabajos de investigación publicados:

- ✓ The fractal organization of ultradian rhythms in avian behavior.
- ✓ A novel non-parametric method for uncertainty evaluation of correlation-based molecular signatures: Its application on PAM50 algorithm.
- ✓ Network dynamics: quantitative analysis of complex behavior in metabolism, organelles, and cells, from experiments to models and back.
- ✓ Fisher vector for PolSAR image Classification.

❖ **Procesamiento de imágenes médicas**

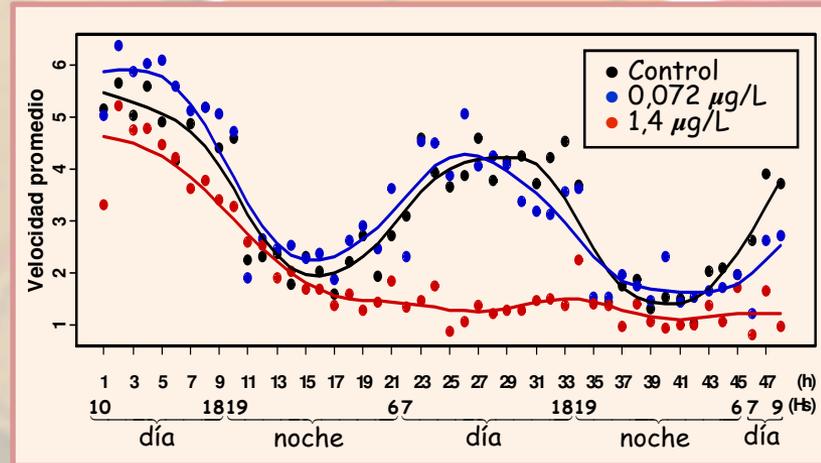
- ✓ **Co-registro** de dos imágenes de un mismo individuo tomadas en momentos diferentes
- ✓ **Detección de cambios**



❖ Construcción de **tests de hipótesis** para comparar curvas de tratamientos en **estudios longitudinales**

Modelos Aditivos Generalizados
Bootstrap

- ✓ Evaluar **cambios en el comportamiento** de peces expuestos a diferentes dosis de tóxicos.

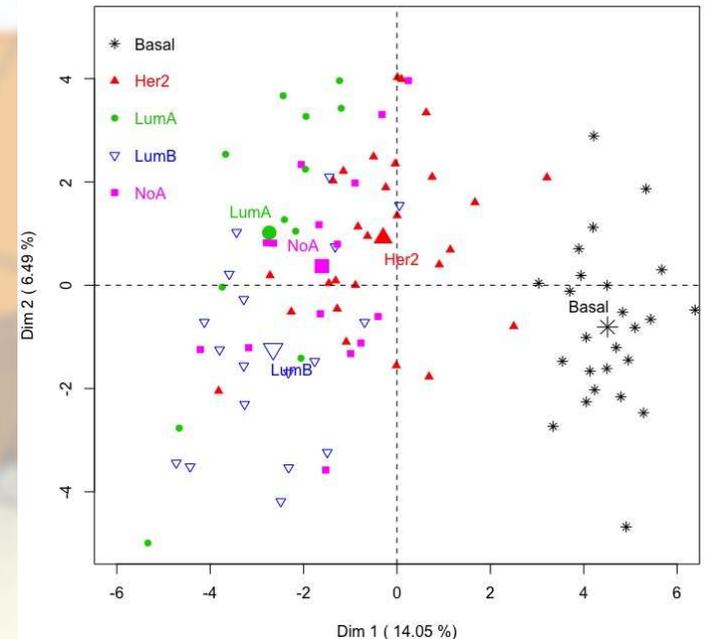


- ✓ Comparar **tasas de mortalidad por cáncer** en la prov. de Córdoba.

❖ Análisis de **datos transcriptómicos** y **proteómicos** de cáncer de mama mediante **métodos multivariados**

- ✓ **Cuatro tipos** de cáncer de mama (Basal, Luminal A, Luminal B, Her2).
- ✓ **Variables:** expresión de genes (medida con dos tecnologías distintas: microArrays y RNASeq) y de expresión de proteínas (iTRAQ).
- ✓ Más de **10000 genes** en cada base de datos.

- ✓ Se utiliza un método de **análisis multivariado multibloque** llamado **Multiple Factor Analysis (MFA)** que permite analizar simultáneamente bloques de datos de distinta naturaleza e interpretar gráficamente cómo se aglomeran los individuos, las variables y los bloques para encontrar **patrones y relaciones**.



- ✓ Se busca **caracterizar** a las pacientes de cáncer de mama (en general, a una población) en función de varias fuentes de información, para poder **clasificar** nuevas pacientes según su tipo de tumor.

Centro de transferencia

“Asesoramiento en Aplicaciones en Probabilidad y Estadística”

FAMAF, UNC

dirigido por la Dra. Silvia Ojeda

- ✓ **Asesoramiento técnico estadístico en trabajos de investigación.**
- ✓ **Apoyo técnico a alumnos de distintas Unidades Académicas para el desarrollo de tesinas de grado y tesis de posgrado.**
- ✓ **Asistencia y consultoría a empresas.**
- ✓ **Dictado de cursos para organismos oficiales o privados en diferentes áreas de la Probabilidad y la Estadística, manejo de paquetes y software estadístico, etc.**



❖ **Asesoramientos a empresas:**

- ✓ Deepvision
- ✓ Infoxel
- ✓ STI
- ✓ URG-Urgencias
- ✓ INVAP

❖ **Capacitación en manejo de software:**

- ✓ URG-Urgencias

❖ **Convenios entre Maestría en Estadística Aplicada- UNC y Ministerio de Gobierno de la Prov. de Córdoba:**

- ✓ **Diseño y aplicación de herramientas para el análisis estadístico periódico de las bases de datos del Observatorio de Estudios sobre Convivencia y Seguridad Ciudadana.**
- ✓ **Análisis estadístico de la Encuesta Córdoba de Victimización y Percepción Social del Delito diseñada e implementada por el Observatorio de Estudios sobre Convivencia y Seguridad Ciudadana.**

❖ **Docencia en Maestrías y Especializaciones:**

- ✓ **Maestría en Estadística Aplicada (FCE, FCA y FAMAFA-UNC)**
- ✓ **Maestría en Análisis y Procesamiento de Imágenes (FAMAFA y FCEFyN-UNC)**
- ✓ **Especialización en Criminalística y Actividades Periciales (FAMAFA, FCEFyN, FCA y FCQ - UNC en conjunto con la Dirección General de Policía Judicial del Ministerio Público Fiscal de la Prov. de Córdoba)**
- ✓ **Maestría en Aplicaciones de Información Espacial (Instituto Mario Gullich, CONAE-UNC y FAMAFA-UNC)**
- ✓ **Maestría en Tecnología Satelital (CONAE y Facultad Regional Córdoba, UTN)**
- ✓ **Maestría en Instrumentos Satelitales (CONAE y Facultad Regional Mendoza, UTN)**
- ✓ **En proceso de creación: Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones (FAMAFA, UNC)**

❖ Vínculos Académico-Científicos:

- ✓ Distintas unidades académicas de la UNC
- ✓ Instituto Mario Gullich, CONAE-UNC
- ✓ Servicio de Diagnóstico por Imágenes, Sanatorio Allende, Córdoba
- ✓ Servicio de Nefrología, Hospital Privado, Córdoba
- ✓ Universidad Católica de Córdoba
- ✓ Universidad Nacional de Mar del Plata
- ✓ Universidad Nacional del Sur (Bahía Blanca)
- ✓ Universidad Técnica Federico Santa María, Chile
- ✓ Pontificia Universidad Católica de Chile
- ✓ Universidad de Toulon, Francia
- ✓ Laboratorios de ex-Xerox (NAVER Labs)

❖ Últimos trabajos publicados

- ❖ Redolfi J, Sánchez J, Flesia G. **Fisher vectors for PolSAR classification.** *Geoscience and Remote Sensing Letters*. En prensa.
- ❖ Lassen O, Tabares S, Ojeda S, Dotto G, Bertolotto P, Sembaj A. **Genetic polymorphisms of manganese-dependent superoxide dismutase in chagas disease.** *Infectious Diseases in Clinical Practice*. En prensa.
- ❖ González-Montoro A, Prato L, Casares F, Balzarini M, Fernández EA, 2017. **Appropriate sample size for standardization parameters estimation reduces misdiagnoses of molecular-based risk predictors in breast cancer.** *Med Sci Tech* 58: 111-118.
- ❖ Guzmán DA, Flesia AG, Aon MA, Pellegrini S, Marin RH, Kembro JM. 2017. **The fractal organization of ultradian rhythms in avian behavior.** *Scientific Reports* 7(1):684.
- ❖ Fresno C, González GA, Merino GA, Flesia AG, Podhajcer OL, Llera AS, Fernández EA. 2017. **A novel non-parametric method for uncertainty evaluation of correlation-based molecular signatures: Its application on PAM50 algorithm.** *Bioinformatics* 33(5):693-700.
- ❖ Kurz FT, Kembro JM, Flesia AG, Armoundas AA, Cortassa S, Aon MA, Lloyd D. 2017. **Network dynamics: quantitative analysis of complex behavior in metabolism, organelles, and cells, from experiments to models and back.** *Wiley Interdiscip Rev Syst Biol Med*. 9(1) e1352.
- ❖ Britos G, Ojeda S. 2017. **Sensitivity study of estimation methods of the two-dimensional autoregressive model.** *Revista de Matemática Aplicada, Computacional e Industrial* 6: 573-576.
- ❖ Britos G, Ojeda S. 2017. **Robust BMM estimator in two-dimensional autoregressive models.** *Trabajos de Matemática, Serie B, FAMA-UNC, N° 119.*
- ❖ Sánchez J, Caverzasi A. 2017. **Leveraging off-the-shelf models for entry-level tag prediction and ranking.** *Workshop on Closing the Loop Between Vision and Language, International Conference on Computer Vision.*

- ❖ Pistonesi S, Martínez J, Ojeda S, Vallejos R. **Structural similarity metrics for quality image fusion assessment: Algorithms**. IPOL Image Processing on Line. Preprint.
- ❖ Landi MA, Di Bella C, Ojeda S, Salvatierra P, Argañaraz J, Bellis ML, 2017. **Selecting control sites for post-fire ecological studies using biological criteria and MODIS time series data**. Fire Ecology 13(2): 1-17.
- ❖ Bustos OH, Rulloni V, Martínez J. 2017. **Distribuciones de Gibbs: Volumen 1**. Trabajos de Matemática, Serie B, FAMAF-UNC, N° 65.
- ❖ Nores ML, Díaz MP. 2016. **Bootstrap hypothesis testing in generalized additive models for comparing curves of treatments in longitudinal studies**. Journal of Applied Statistics 43: 810-826.
- ❖ Pistonesi S, Martínez J, Ojeda S, 2016. **A sensibility study of the autobinomial model estimation methods based on a feature similarity index**. Computational Statistics 31(4): 1327-1357.
- ❖ Lassen O, Herrera J, Dotto G, Ojeda S, Garutti A, Bertolotto P, Tabares S, Sembaj A. 2016. **Plasmatic biochemical variables associated with polymorphisms in the endothelin-1 and endothelin-1 receptor A genes in hypertensive patients: Pilot study**. British Journal of Medicine and Medical Research 11(7): 1-8.
- ❖ Gagliano AE, Foglino EV, Lezama J, Valente M, Nores ML, Rulloni V. **Evaluation of change detection techniques in simulated mammography**. 2016. 2015 XVI Workshop on Information Processing and Control, IEEE Xplore.
- ❖ Sánchez J, Pury P, Marconi V, 2016. **Un algoritmo modular para el seguimiento de partículas en videos de microscopía**. XXII Congreso sobre métodos numéricos y sus aplicaciones. ENIEF, Córdoba.
- ❖ Sánchez J, Pury P, Marconi V. 2016. **Algoritmo modular de trackeo**. Reunión de la Asociación Física Argentina. Tucumán.

¡Muchas gracias por su atención!

The background features a collage of business-related graphics. In the center is a 3D pie chart with four segments in shades of orange, brown, and yellow. To the right is a bar chart with four bars in red, green, and blue. The background also includes a grid pattern and some faint, illegible text and mathematical symbols like a plus sign and a function notation.