Modelación Estadística

Charlas de Acercamiento ICMAT

Departamento de Matemática
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

Estructura de la presentación

0. Presentación del grupo de Estadística.
1. Esquema de modelamiento estadístico.
2. Cursos del área de estadística.
3. Posibles tópicos de investigación.
4. Alumnos y ex-alumnos asociados a la línea.
5. Publicaciones científicas.

Modelación Estadística: Académicos

(a) Alfredo Alegría

(b) Francisco Cuevas

(c) Felipe Osorio

(d) Ronny Vallejos

Modelación Estadística: Alfredo Alegría[1]

- Ph.D. Mathematics, PUCV-UTFSM-UV, Chile (supervisor: Emilio Porcu).
- Áreas de investigación: Multivariate spatial statistics, Geostatistics for large datasets, Non-gaussian random fields.

- Red de Colaboración: Emilio Porcu, Peter Diggle (Reino Unido), Jorge Mateu (España), Reinhard Furrer (Suiza), Stefano Castruccio (USA), Moreno Bevilacqua, Xavier Emery (Chile).
- Trabajos relevantes: Computational Statistics & Data Analysis, Electronic Journal of Statistics, Environmentrics, International Statistical Review, Journal of Multivariate Analysis, SIAM Journal on Scientific Computing, Spatial Statistics, Statistics and Computing.
- Proyectos FONDECYT.
- Postdoctorado en la Universidad de Newcastle, UK.

[1] Página web: https://sites.google.com/site/alfredoalegriajimenez/

Modelación Estadística: Francisco Cuevas[2]

- Ph.D. Mathematics and Physics, Ålborg University Denmark, and PUCV-UTFSM-UV, (supervisor: Jesper Møller).
- Áreas de investigación: Multivariate spatial statistics, Point patterns, Functional data.

- Red de Colaboración: Peter Diggle (Reino Unido), Marie-Helene Descary (Canadá), Jean-François Coeurjolly (Francia), Jesper Møller, Rasmus Waagepetersen, Christophe Biscio (Dinamarca), Moreno Bevilacqua (Chile).
- Trabajos relevantes: Biometrika, Environmentrics, Journal of Nonparametric Statistics, Spatial Statistics, Statistics and Computing.
- Proyecto ANID Postroctorado.
- Postdoctorado en Université du Québec à Montréal, Montreal, Canadá.
- Miembro del centro AC3E.

[2] Página web: https://fcocuevas87.github.io

Modelación Estadística: Felipe Osorio[3]

- D. Sc. Statistics, Universidade de São Paulo, Brasil.
  (supervisor: Gilberto A. Paula).
- Áreas de investigación:
  Modelos para datos longitudinales,
  Diagnóstico de influencia,
  Funciones de inferencia.

- Red de Colaboración: Gilberto A. Paula, Cibele Russo (Brasil), Manuel Galea (Chile), Federico Crudu (Italia).
- Trabajos relevantes: Annals of the Institute of Statistical Mathematics, Biometrical Journal, Computational Statistics & Data Analysis, Economics Letters, Signal Image and Video Processing, Spatial Statistics, Statistical Papers, Statistics and Computing.
- Proyectos FONDECYT, de cooperación internacional (PROSUL, CNPq).
- Creador de paquetes contribuídos a R (fastmatrix, heavy, L1pack, SpatialPack).
- Editor de la revista *Chilean Journal of Statistics*.

[3] Página web: http://fosorios.mat.utfsm.cl

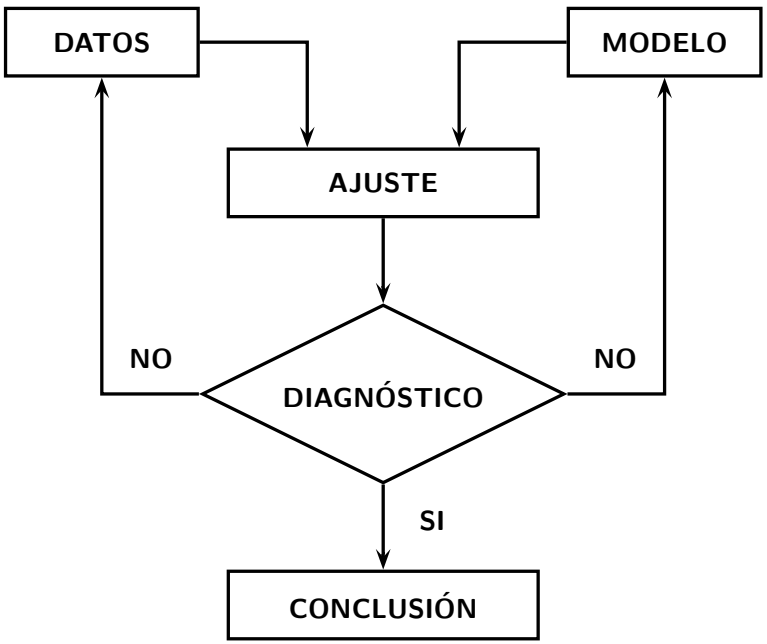
Modelación Estadística: Ronny Vallejos[4]

- Ph.D. Statistics, University of Maryland Baltimore County, USA. (supervisor: Andrew L. Rukhin).
- Áreas de investigación: Spatial statistics, Robust modelling, Statistical image modelling, Time series.

- Red de Colaboración: Daniel Griffith, Aaron Ellison (USA), Silvia Ojeda, Oscar Bustos (Argentina), Hannah Buckley, Bradley Case (New Zealand).
- Trabajos relevantes: Chance, Electronic Journal of Statistics, Journal of Mathematical Imaging and Vision, Journal of Statistical Planning and Inference, Natural Resource Modeling, Spatial Statistics, Stochastic Environmental Research and Risk Assessment.
- Proyectos FONDECYT, de cooperación internacional (CECYT, Math-AmSud), PIA.
- Ex Editor en Jefe de la revista *Chilean Journal of Statistics*, miembro de los centros AM2V y AC3E.

[4] Página web: http://rvallejos.mat.utfsm.cl

Esquema de Modelación Estadística



Recolección de datos: Muestreo.

Análisis exploratorio de datos.

Análisis Multivariado.

Técnicas de Regresión.

Series de Tiempo, entre (muchas) otras.

Inferencia Estadística.

Bondad de ajuste, técnicas gráficas.

Análisis de Sensibilidad.

Comunique sus resultados!

Cursos de especialidad: Modelación Estadística

Obligatorios:

- MAT-041: Probabilidad y Estadística
- MAT-263: Teoría de Probabilidades y Procesos Estocásticos.
- MAT-206: Inferencia Estadística.

Especialidad:

- MAT-266: Análisis de regresión.
- MAT-267: Series de tiempo.
- MAT-269: Análisis multivariado.

Complementarios:

- Minería de datos (ELO)
- Proc. imágenes digitales (ELO)
- Teoría de información (ELO)
- Base de datos (INF)
- Física computacional (FIS)
- Inteligencia artificial (INF)

Objetivo:

Estudiar una variable de respuesta, y [asuminda continua] como función de algunas variables explicativas o regresores, $x_1, x_2, \ldots$ [pueden ser discretas y/o continuas].

En ocasiones la relación funcional es conocida salvo algunos coeficientes (parámetros).

Es decir, la relación puede estar gobernada por un proceso físico o por leyes bien aceptadas, tal que:

$$Y \approx f(x_1, \ldots, x_p; \boldsymbol{\beta}),$$

en cuyo caso, el interés recae en estimar el vector de parámetros $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$.

Asumiremos variables aleatorias independientes $Y_1, \ldots, Y_n$, tal que

$$Y_i = \mu_i + \epsilon_i, \qquad \mathsf{E}(\epsilon_i) = 0, \qquad i = 1, \ldots, n,$$

esto es,

$$\text{respuesta} = \text{parte sistemática} + \text{error aleatorio}$$

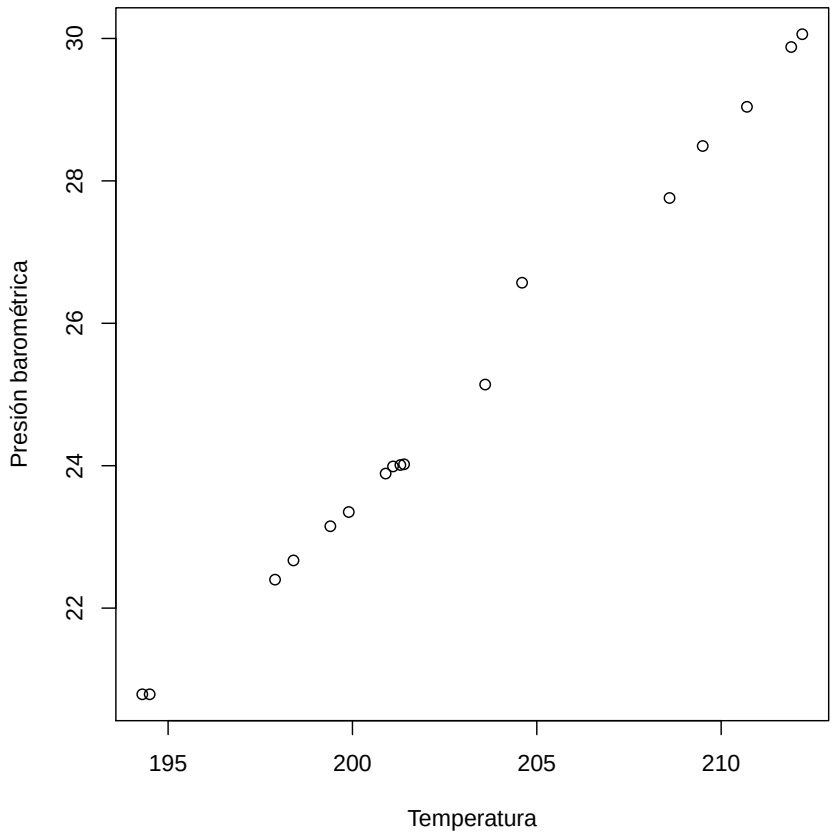
Idea:

Se desea **"estructurar"** la función de media como

$$\mu_i = \mu_i(\boldsymbol{\beta}), \qquad i = 1, \ldots, n,$$

con $\boldsymbol{\beta} \in \mathbb{R}^p$ y $n \gg p$.

Presión barométrica en pulgadas de mercurio y temperatura de ebullición del agua en grados Fahrenheit para 17 diferentes altitudes (Forbes, 1857)[5].



[5] Transactions of the Royal Society of Edinburgh **21**, 235-243.

Para describir la relación entre la temperatura y la media de la presión barométrica, podemos considerar:

$$\mu = \beta_0 + \beta_1 x,$$

note que

$$\mu = \boldsymbol{x}^\top \boldsymbol{\beta}, \qquad \boldsymbol{x} = (1, x)^\top, \qquad \boldsymbol{\beta} = (\beta_0, \beta_1)^\top,$$

y $\mu = \boldsymbol{x}^\top \boldsymbol{\beta}$ se denomina predictor lineal.

El conjunto de datos consiste del vector de respuestas.[6]

$$\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top,$$

y una matriz de diseño $n \times 2$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}.$$

[6] Todos los vectores siempre serán columna.

Para los datos de Forbes podemos considerar un modelo de regresión lineal definido como

$$Y_i \overset{\text{ind}}{\sim} \mathsf{N}_1(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2), \qquad i = 1, \ldots, n,$$

o bien, escribir en forma compacta,[7]

$$\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n),$$

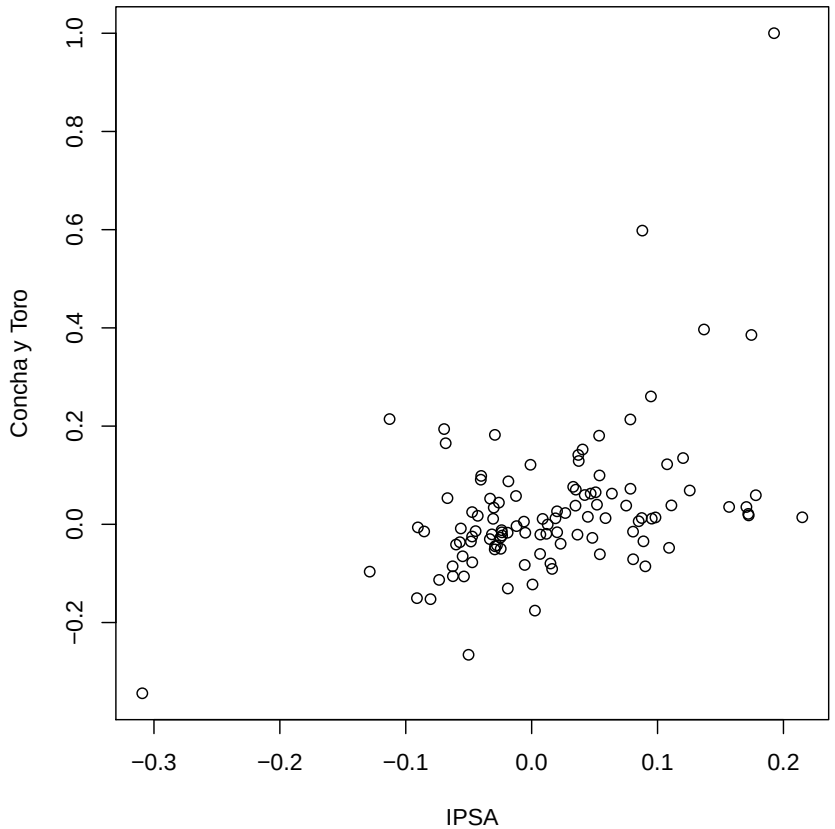
lo que lleva a,

$$\mathsf{E}(Y_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \qquad \mathsf{var}(Y_i) = \sigma^2, \qquad \mathsf{Cov}(Y_i, Y_j) = 0,$$

para $i, j = 1, \ldots, n$.

[7] Evidentemente también podemos escribir, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

Rentabilidades mensuales de Concha y Toro vs. IPSA, ajustados por bonos de interés del Banco Central entre marzo/1990 a abril/1999 (Osorio y Galea, 2006)[8].



[8] Statistical Papers **47**, 31-38

Modelo CAPM (Valoración de Activos de Capital), (Sharpe, 1964)[9]

$$\mathsf{E}(r) = r_f + \beta(\mathsf{E}(r_m) - r_f),$$

usando datos observados, podemos escribir

$$R_t = \alpha + \beta\, \mathsf{IPSA}_t + \epsilon_t, \qquad t = 1, \ldots, T.$$

Características del problema:

- Relación lineal entre las variables.
- Posibles periodos de alta volatilidad.

Hipótesis de interés:

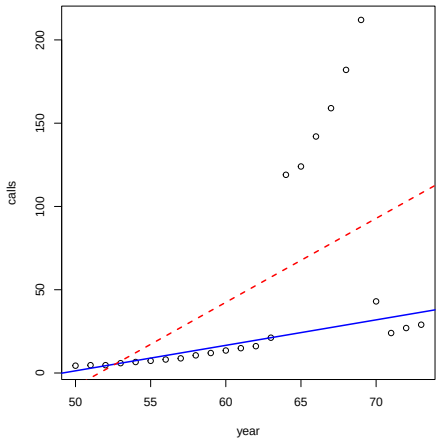
- $H_0 : \beta > 1$ (Amante del riesgo).
- $H_0 : \beta = 1$ (Neutral al riesgo).
- $H_0 : \beta < 1$ (Averso al riesgo).

[9] Journal of Finance **19**, 425-442

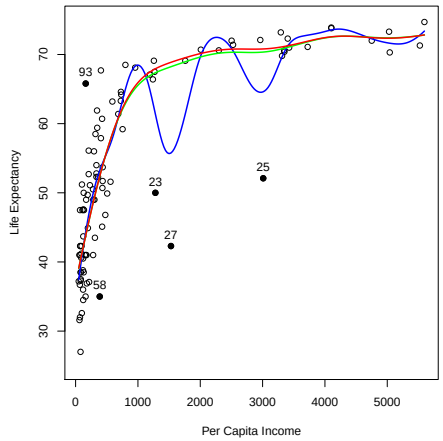
- El requisito formal de la asignatura es MAT-041: Probabilidad y Estadística
- El curso se enfoca en estudiar:
  - Preliminares distribucionales.
  - Inferencia en el modelo de regresión lineal.
  - Análisis de los supuestos del modelo.
  - Identificación del mejor conjunto de regresores.
  - Alternativas a mínimos cuadrados.
  - Tópicos adicionales.

Modelos lineales son los *bloques de construcción* para metodologías más complejas, tales como:

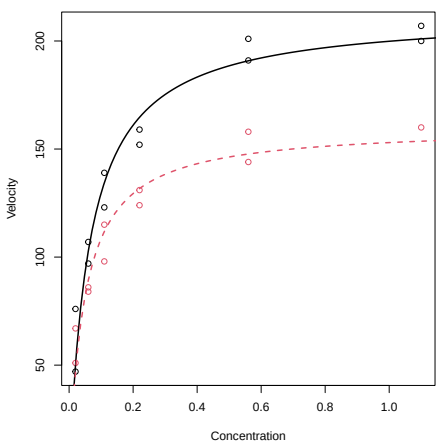
- Modelos lineales generalizados.
- Modelos no lineales.
- Modelos de regresión espacial.
- Regresión multivariada.
- Datos longitudinales, GMANOVA.
- Regresión semiparamétrica.
- Modelos con efectos mixtos.



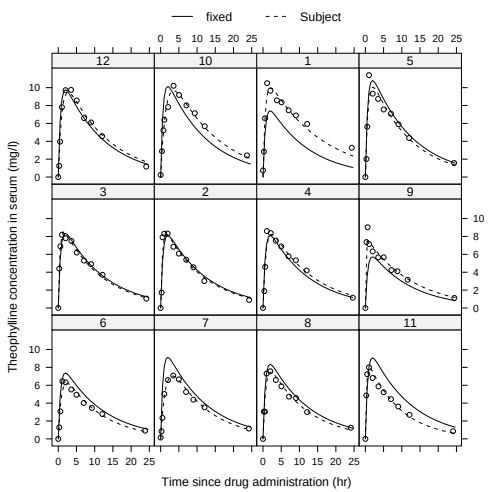
(a) regresión LAD



(b) splines penalizados



(c) regresión no-lineal



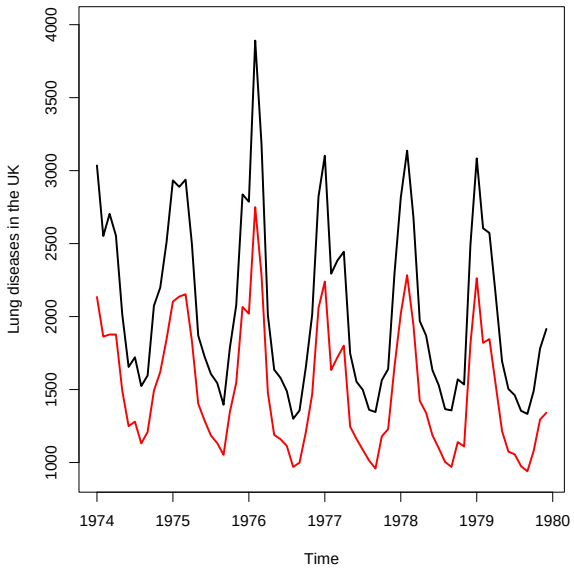
(d) modelos mixtos

Objetivo:

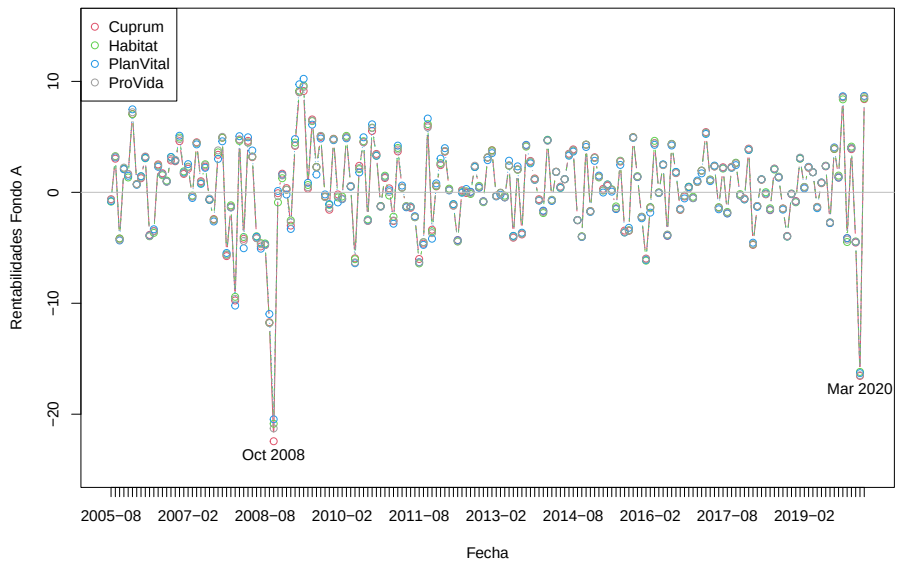
Se presentan procedimientos para modelar y hacer pronósticos en series cronológicas.

Se describe métodos basados en suavizamiento exponencial, modelos ARMA y ARIMA así como modelos dinámicos y el filtro de Kalman univariado.

Se aborda la identificación, estimación y validación de modelos que consideran la presencia de correlación serial, para posteriormente aplicarl la metodología en problemas de predicción o control.



(a) muertes en UK



(b) datos de AFPs

Algunos modelos que se describen en la asignatura son:

- Modelo AR(p):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \epsilon_t,$$

- Modelo MA(q):

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}.$$

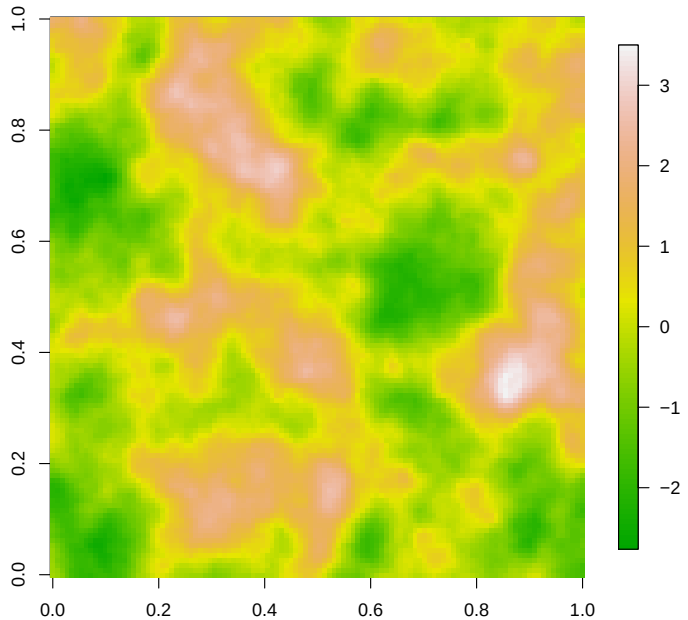
- Modelo ARMA(p, q):

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$

- Modelo de espacio de estados:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{A}_t \boldsymbol{z}_t + \boldsymbol{u}_t, \\ \boldsymbol{z}_t &= \boldsymbol{\Phi}_{t-1} \boldsymbol{z}_{t-1} + \boldsymbol{v}_{t-1}. \end{aligned}$$

- El requisito de la asignatura es MAT-266: Análisis de Regresión
- El curso se enfoca en estudiar:
  - Métodos de suavizamiento exponencial.
  - Procesos estocásticos estacionarios.
  - Modelos ARMA, ARIMA y ARIMA estacional.
  - Análisis espectral.
  - Variables de estado y filtro de Kalman.
  - Tópicos adicionales.



(a) Kriging

(b) Imagen SAR

Datos multivariados:

Tenemos una muestra aleatoria $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ donde para cada observación se ha medido $p \geq 2$ variables (o características) de interés.[10]

Podemos disponer la información en una matriz de datos:

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}.$$

Observación:

Por simplicidad asumiremos que $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ son variables aleatorias independientes desde $F_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (con F_p común).

[10] Es decir, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ es vector p-dimensional.

Desde 1988 el SIMCE evalua los resultados de aprendizaje de los estudiantes del sistema de educación chileno.

Objetivos:

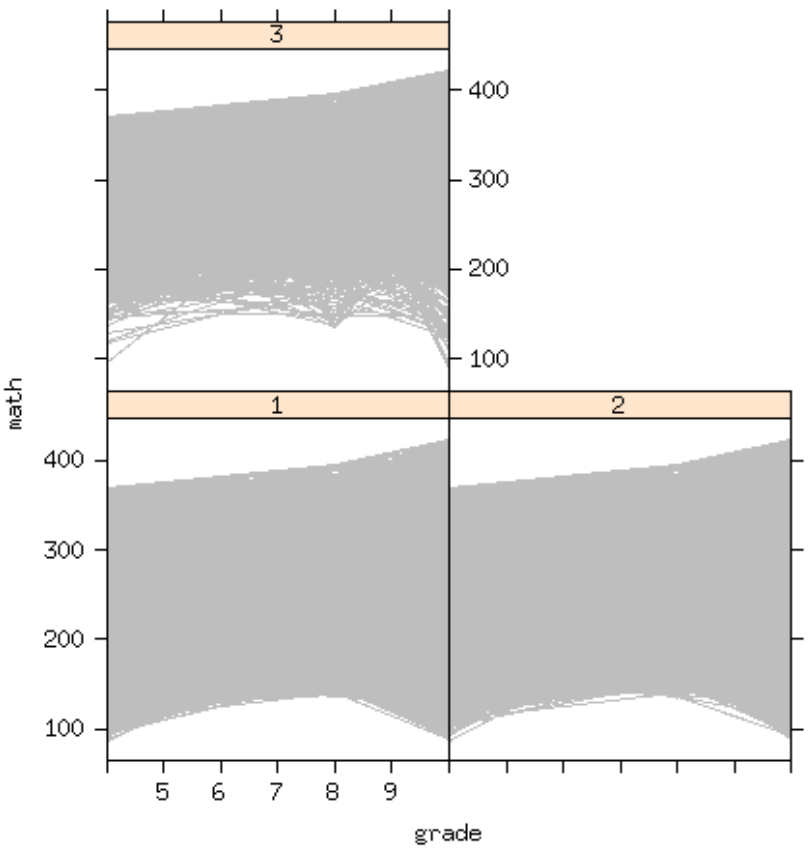
- Describir el comportamiento del aprendizaje de los estudiantes.
- Determinar si existe diferencias significativas entre el tipo de dependencia (municipal, subvencionado, particular).

Características del problema:

- Mediciones de un mismo individuo (estudiante) a través del tiempo (4º y 8º básico, 2º medio).[11]
- Datos disponibles para los años 2007, 2011 y 2013, pruebas de Lenguaje y Matemáticas.

[11] Conocido como: datos con estructura longitudinal.

Perfiles individuales de los puntajes del SIMCE en matemáticas, organizados por tipo de dependencia.



Características del problema:

- Aproximadamente 132K estudiantes para ser analizados (base de datos de mediano porte).
- Crecimiento lineal (cuadrático?) a través del tiempo.
- Igual número de mediciones por individuo (datos balanceados).

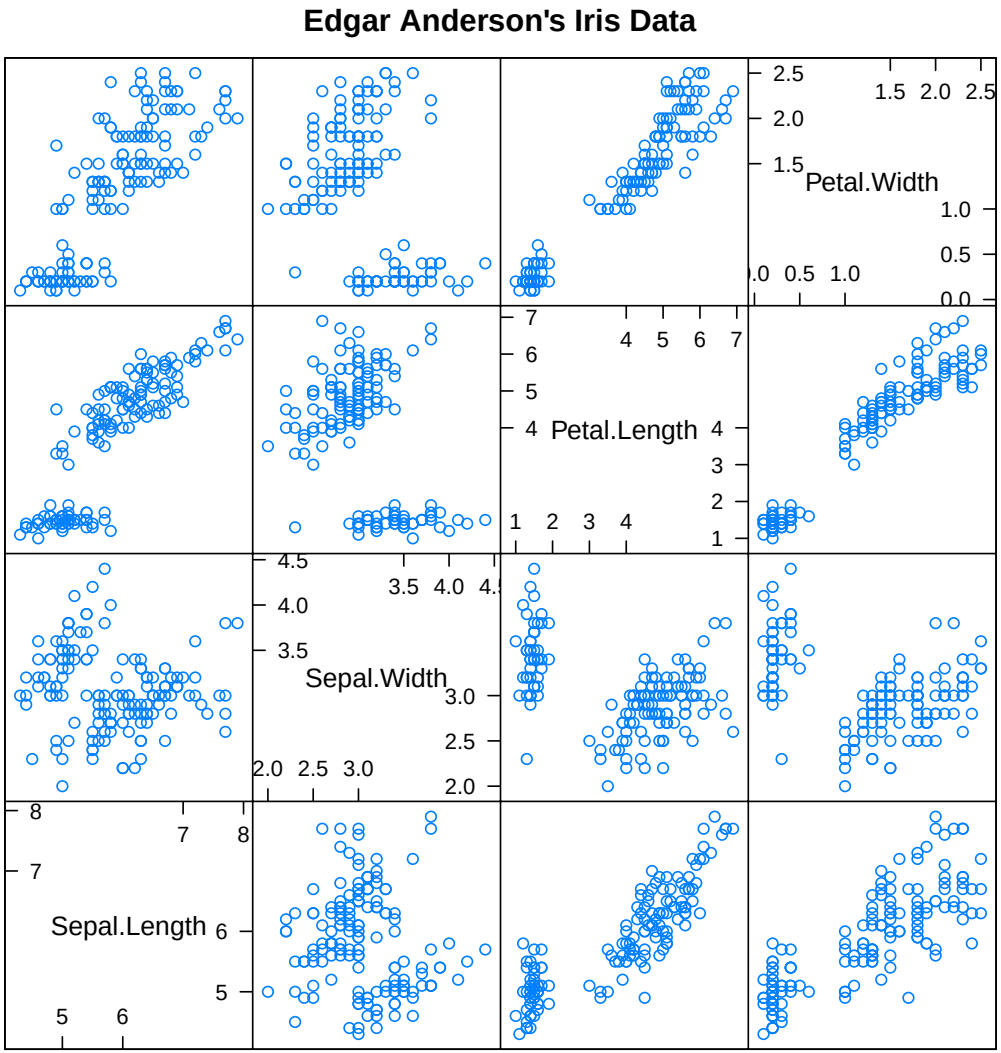
Alternativas para análisis:

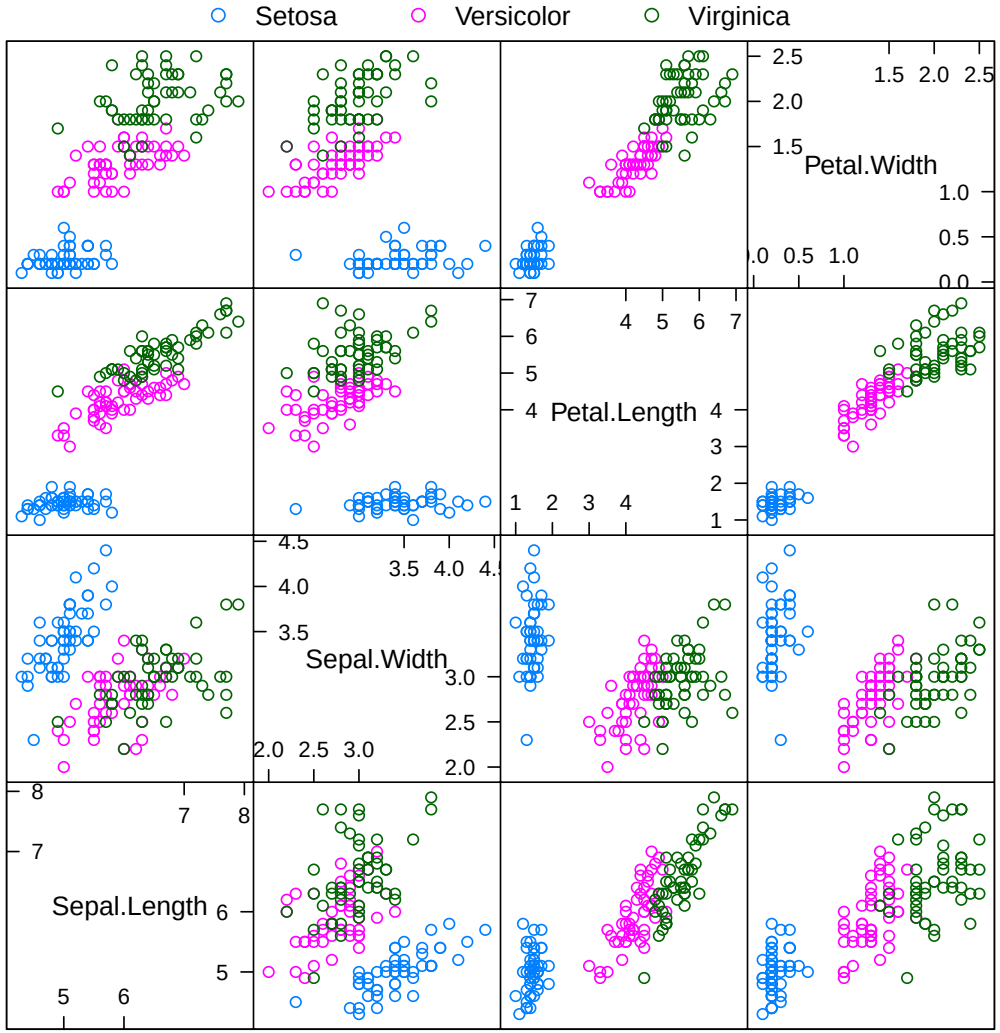
- Modelos con efectos-mixtos.
- Modelos multi-nivel.
- Modelo de curvas de crecimiento (GMANOVA).

Iris setosa

Iris versicolor

Iris virginica


Edgar Anderson's Iris Data
Petal.Width
Petal.Length
Sepal.Width
Sepal.Length
Scatter Plot Matrix



Scatter Plot Matrix

Datos observados:

Mediciones (cm) del largo y ancho de los sépalos y el largo y ancho de pétalos para 50 flores desde 3 especies de Iris (setosa, virginica y versicolor).

Objetivo:

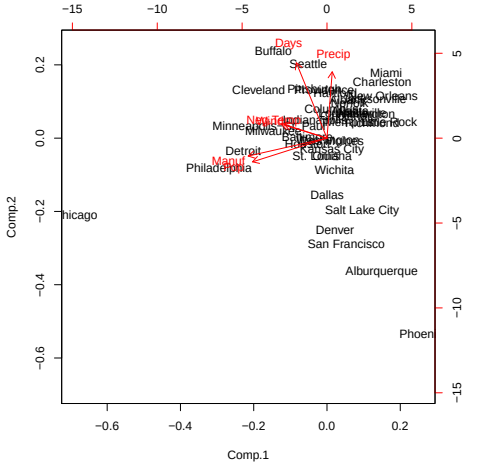
- Obtener una función que permita discriminar entre especies.
- Usando las medidas de una flor, clasificarla apropiadamente.

Características del problema:

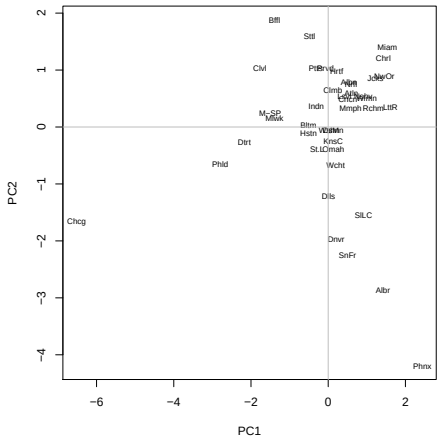
- El análisis exploratorio revela una separación evidente en 2 grupos.
- Técnicas más refinadas permiten identificar las 3 especies, p.ej.:
  - Análisis discriminante,
  - Técnicas de clasificación (Reconocimiento de patrones),
  - Aprendizaje de máquina (Máquinas de soporte vectorial, Data mining).

MAT-269: Análisis Estadístico Multivariado

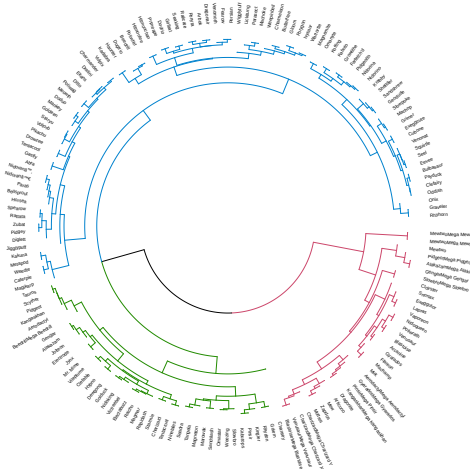
- El requisito de la asignatura es MAT-266: Análisis de Regresión.
- Inferencia en análisis multivariado.
  - Estimación y test de hipótesis para una muestra aleatoria desde $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Técnicas multivariadas.
  - Regresión multivariada y GMANOVA.
  - Análisis de componentes principales.
  - Análisis factorial.
  - Métodos de clasificación y agrupamiento.
- Tópicos adicionales.



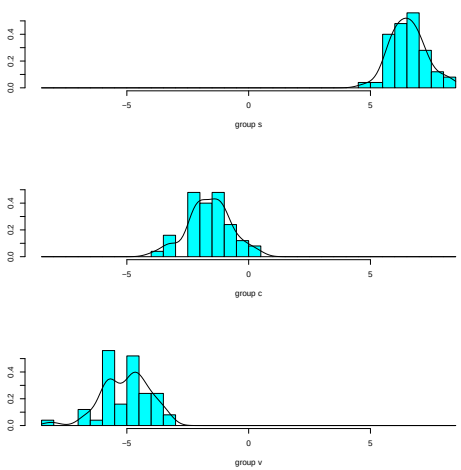
(a) biplot



(b) PCA



(c) cluster



(d) LDA

Cursos de postgrado: Modelación Estadística

Magíster en Matemática:

- MAT-431: Teoría de Probabilidades (Corresponde a MAT-263).
- MAT-460: Inferencia Estadística (Corresponde a MAT-206).
- MAT-420: Procesos Estocásticos.
- MAT-417: Series de Tiempo. (Corresponde a MAT-267).
- MAT-467: Modelos Espacio-Temporales.
- MAT-466: Modelos Lineales Generalizados.
- MAT-468: Simulación Estocástica.

Reportes preliminares sobre el COVID-19



Profesores:

- Pedro Gajardo (Optimización).
- Rodrigo Lecaros (Problemas Inversos).
- Felipe Osorio (Estadística).

Estudiantes:

- Claudia Álvarez.
- José Fuentealba.
- Alonso Ogueda.
- Fabián Ramírez.
- Bernardo Recabarren.

Reportes preliminares sobre el COVID-19

- El grupo ha realizado **8** reportes preliminares sobre el COVID-19 en Chile.
- Se ha participado en **2** Workshops sobre el modelamiento del brote de COVID-19.
- Colaboración con:
  - Centro de Modelamiento Matemático (CMM) de la Universidad de Chile.
  - Centro de Epidemiología y Políticas de Salud (CEPS) de la Universidad del Desarrollo.
- Objetivos de los reportes:
  - Estimación de la demanda de camas UCI (usando un modelo compartimentado).
  - Efectividad de las cuarentenas dinámicas.
  - Evaluación de la reapertura de escuelas (para algunas regiones).
  - Regresión segmentada para estimar los tiempos de duplicación del número de casos de COVID-19.

Método de regresión segmentada (Muggeo, 2003)[12]

Considere un GLM, donde $\mu_i = \mathsf{E}(Y_i)$, con

$$g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \delta_0 z_i + \delta_1 (z_i - \psi)_+,$$

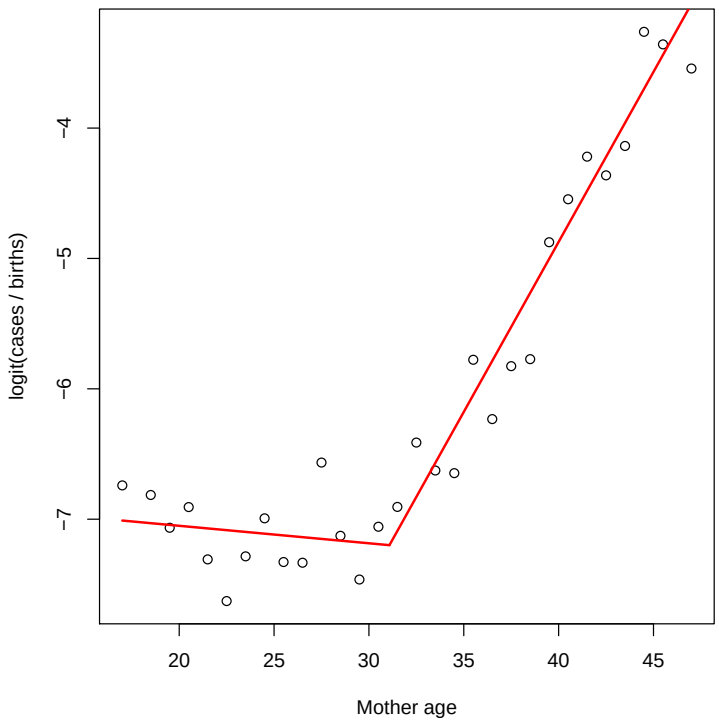
donde $\boldsymbol{x}_i$ denota covariables, $\boldsymbol{\beta}$ son coeficientes de regresión, δ_0, δ_1 denotan la pendiente y el incremento de cada segmento, mientras que ψ indica el punto de cambio.

El procedimiento puede ser extendido para estimar $K + 1$ regímenes, como:

$$g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \delta_0 z_i + \delta_1 (z_i - \psi_1)_+ + \cdots + \delta_K (z_i - \psi_K)_+.$$

[12] Statistics in Medicine **22**, 3055-3071.

Regresión segmentada para evaluar la efectividad de la cuarentena



- El método de estimación es basado en el uso de una aproximación lineal.
- La propuesta es bastante más eficiente que su contraparte basada en búsqueda de grilla.
- Podemos probar la existencia de un punto de cambio usando test de hipótesis.

$$H_0 : \delta_j(\psi_j) = 0, \qquad j = 1, \ldots, K.$$

- Método implementado en el paquete R: `segmented`.

Regresión segmentada para datos de COVID-19 (Muggeo et al., 2020)[13]

Sea Y_t el número acumulado de casos infectados en el día $t = 1, 2, \ldots$. El modelo de crecimiento exponencial, lleva al modelo de regresión Poisson para $\mu_t = \mathrm{E}(Y_t)$, como:

$$\log \mu_t = \beta_0 + \beta_1 t + \delta_1 (t - \psi_1)_+ + \cdots + \delta_K (t - \psi_K)_+.$$

Es decir, tenemos $K + 1$ regímenes, y

$$\beta_1, \quad \beta_2 = \beta_1 + \delta_1, \quad \ldots, \quad \beta_{K+1} = \beta_1 + \sum_{k=1}^{K} \delta_k,$$

denota las pendientes para cada uno de los segmentos.

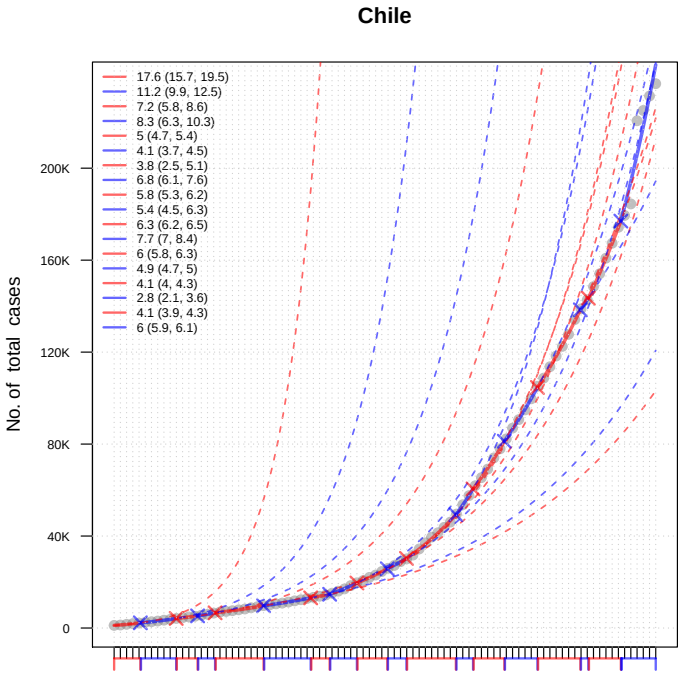
Observación:

Basado en lo anterior podemos calcular las tasas de crecimiento y los tiempos de duplicación, como:

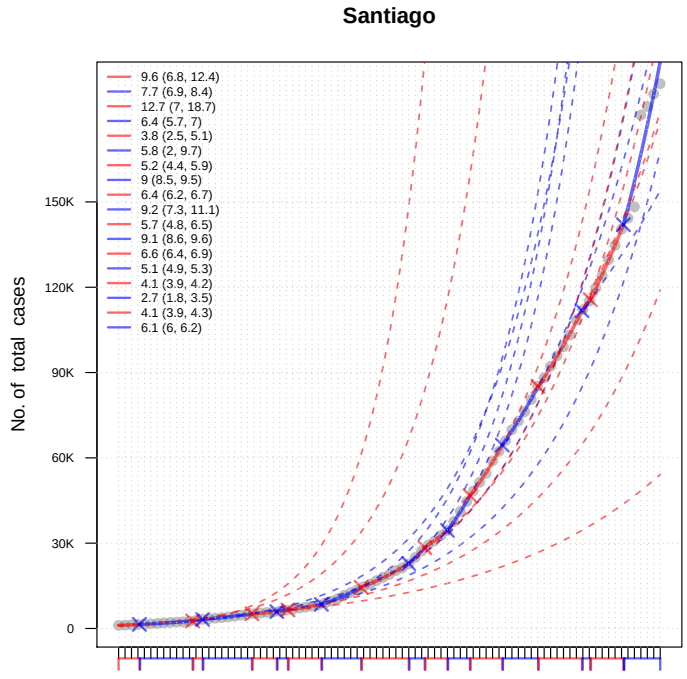
$$r_k = \exp(\beta_k) - 1, \qquad d_k = \log(2)/\beta_k, \qquad k = 1, \ldots, K + 1.$$

[13] doi: 10.13140/RG.2.2.32798.28485

Regresión segmentada y puntos de cambio: Chile y RM

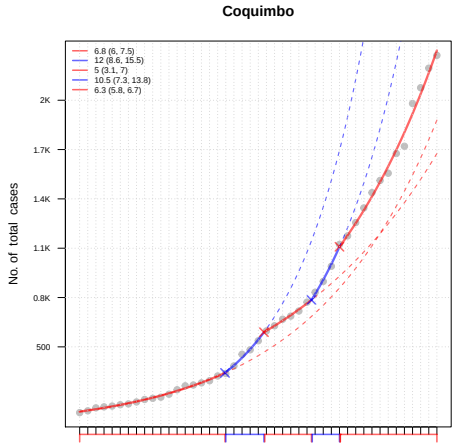


(a)

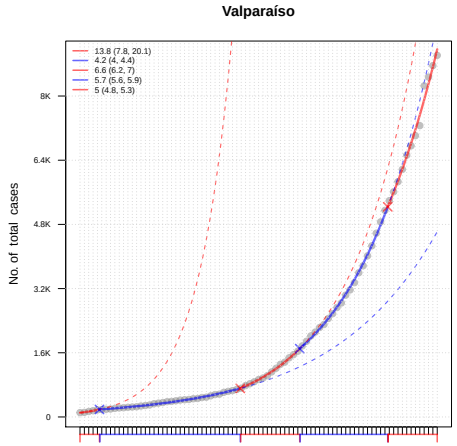


(b)

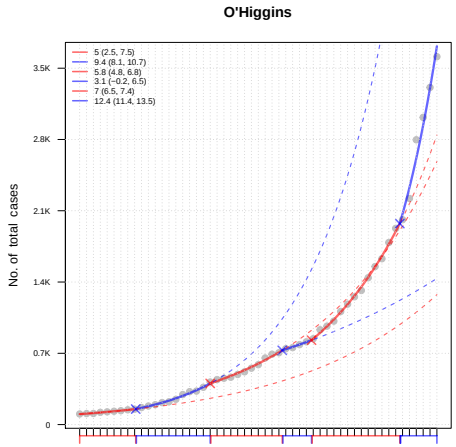
Regresión segmentada y puntos de cambio: Coquimbo al Maule



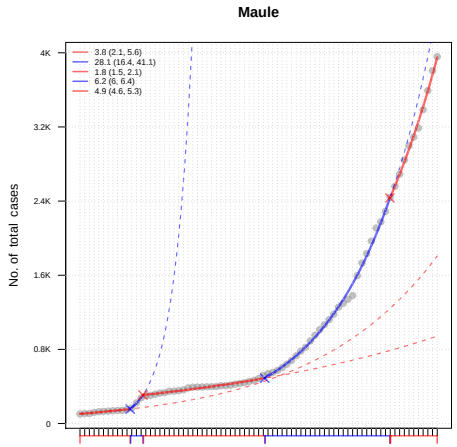
(a)



(b)



(c)



(d)

Resumen de estimación[14]

Región	casos totales	número de segmentos	Tasa de crecimiento		Tiempos de duplicación	
Arica	1 325	3	1.7	3.8	40.10	18.50
Tarapacá	5 143	6	6.5	3.8	11.03	18.38
Antofagasta	6 215	4	3.6	6.5	19.79	11.05
Atacama	650	2	2.7	6.0	25.64	11.88
Coquimbo	2 273	5	10.5	6.3	6.93	11.38
Valparaíso	9 014	5	5.7	5.0	12.46	14.08
Metropolitana	191 577	18	4.1	6.1	17.31	11.66
O'Higgins	3 613	6	7.0	12.4	10.26	5.92
Maule	3 958	5	6.2	4.9	11.51	14.40
Ñuble	2 135	5	2.2	2.6	31.82	26.53
Biobío	4 658	8	3.9	6.7	18.23	10.76
La Araucania	2 920	6	1.8	1.0	37.87	71.63
Los Ríos	589	3	1.2	2.9	58.20	24.64
Los Lagos	1 335	3	1.6	2.1	43.37	32.97
Aysén	28	3	0.3	3.9	198.90	18.12
Magallanes	1 315	3	0.6	2.4	120.42	28.85
Chile	236 748	18	4.1	6.0	17.23	11.95

[14] Descargados el día Sábado 20 de Junio 2020,
URL: https://github.com/MinCiencia/Datos-COVID19/tree/master/output/producto3

Estimación del SSIM

Trabajo:

Osorio, F., **Vallejos, R.**, **Barraza, W.**, Ojeda, S., Landi, M.A. (2022). Statistical estimation of the structural similarity index for image quality assessment. *Signal, Image and Video Processing* **16**, 1035-1042.

Considere dos imágenes $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_{+}^{N}$, el coeficiente de similaridad estructural (SSIM) es dado por:

$$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma},$$

donde

$$l(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\overline{x}\,\overline{y} + c_1}{\overline{x}^2 + \overline{y}^2 + c_1}, \qquad c(\boldsymbol{x}, \boldsymbol{y}) = \frac{2s_x\, s_y + c_2}{s_x^2 + s_y^2 + c_2},$$

$$s(\boldsymbol{x}, \boldsymbol{y}) = \frac{s_{xy} + c_3}{s_x\, s_y + c_3}.$$

Objetivo:

Basado en imágenes observadas $\boldsymbol{x}$ e $\boldsymbol{y}$, estimar $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^{\top}$ y probar la hipótesis:

$$H_0 : \alpha = \beta = \gamma = 1.$$

Estimación del SSIM

Se consideró un modelo no-lineal heteroscedástico[15] bajo el supuesto:

$$Z_i \sim \mathrm{N}(\phi f_i(\boldsymbol{\theta}), f_i^2(\boldsymbol{\theta}) g^2(\phi)), \qquad i = 1, \ldots, n,$$

donde

$$f_i(\boldsymbol{\theta}) = \mathrm{SSIM}(\boldsymbol{x}_i, \boldsymbol{y}_i; \boldsymbol{\theta}), \qquad g^2(\phi) = \phi^2(\phi^2 - 1),$$

corresponden a funciones de media y de varianza y $Z_i = 1/\,\mathrm{RMSE}(\boldsymbol{x}_i, \boldsymbol{y}_i)$.

Resultados:

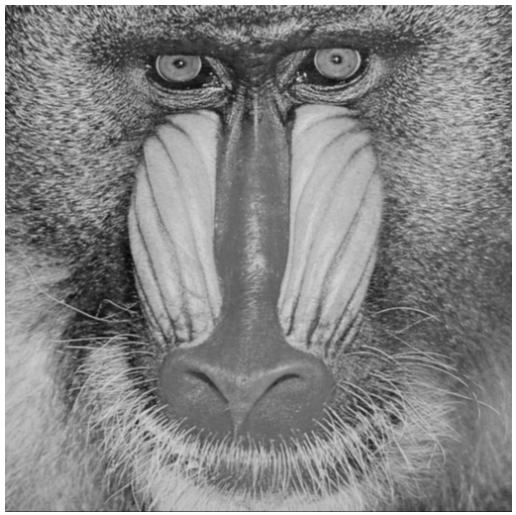
- Algoritmo de estimación: Híbrido entre método secante multivariado (BFGS) con pseudo-verosimilitud (método de Brent).[16]
- Test de hipótesis usando el estadístico gradiente (Terrell, 2000).
- Matriz de información de Fisher y método eficiente para evaluar función score.
- Experimento numérico con datos sintéticos (desde base de datos USC-SIPI) y datos desde la constelación de satélites ICEYE SAR.

[15] Inspirado en el contexto de funciones de producción (Econometría).

[16] Código C y R disponible en `github.com/faosorios/SSIM`.

Estimación del SSIM: estudio de simulación Monte Carlo[19]

(a) texmos2.S512



(b) Baboon

(c) Lenna

Imágenes de referencia[17] ($\boldsymbol{x} \in \mathbb{R}_+^N$) fueron contaminadas con ruido multiplicativo[18] usando una distribución Gamma(L, L), es decir,

$$\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{w}, \qquad w_t \sim \text{Gamma}(L, L), \qquad t = 1, \ldots, N,$$

para $L = 1, 2, 4, 8, 16$ y 32 looks. Para cada look 1,000 imágenes fueron construídas.

[17] Extraídas desde la base de imágenes USC-SIPI, URL: http://sipi.usc.edu/database

[18] Disponible en la función `imnoise` desde el paquete **R** `SpatialPack`.

[19] Se ajustaron 144 000 modelos (tiempo: 34 hrs, 40 min, 16 seg).

Estimación del SSIM: contaminación de Lenna

(a) $L = 1$

(b) $L = 2$

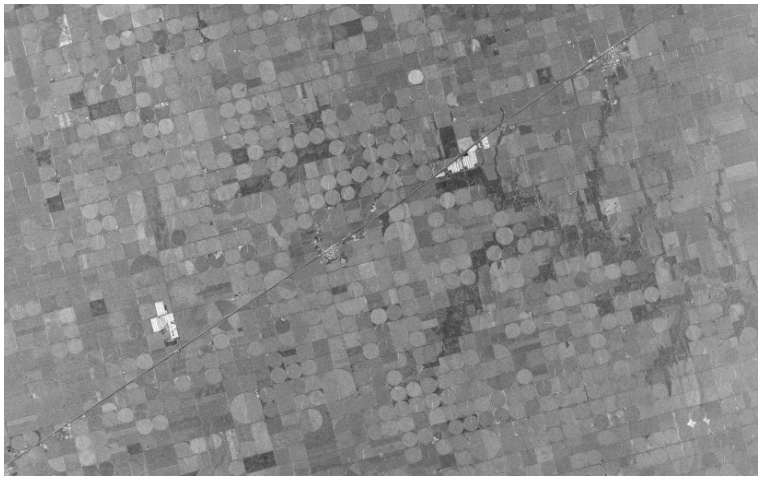
(c) $L = 4$

(d) $L = 8$

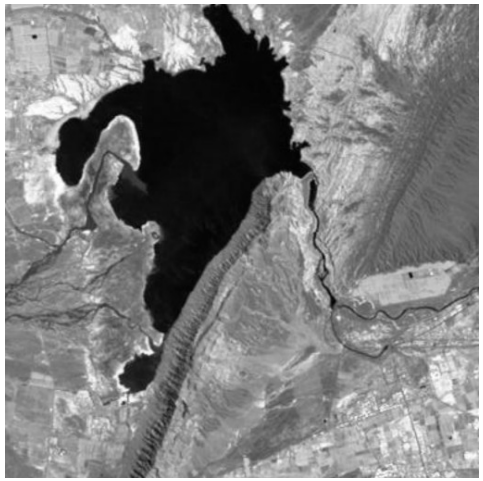
(e) $L = 16$

(f) $L = 32$

Estimación del SSIM: imágenes de Radar de Apertura Sintética (SAR)[20]

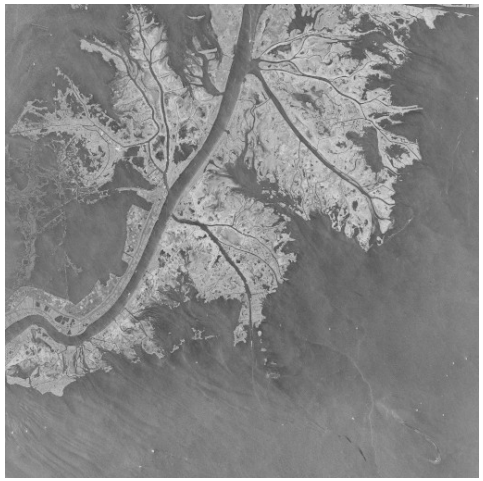


(a) Copeland



(b) Dam

(c) Corpus Christi



(d) Mississippi

[20] Constelación de satélites ICEYE SAR: https://www.iceye.com/downloads/datasets

Estimación del SSIM: resultados con imágenes SAR

Copeland (4096×2560)				Corpus Christi (4096×2560)		
Filter	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
Lee	1.891	1.932	3.136	1.610	1.639	2.705
Enhanced Lee	1.691	1.722	2.987	1.600	1.626	2.512
Kuan	1.691	1.722	2.987	1.600	1.626	2.512
Dam (1200×1200)				Mississippi (4096×4096)		
Filter	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
Lee	1.000	1.000	1.000	1.591	1.623	2.670
Enhanced Lee	1.000	1.000	1.000	1.468	1.491	2.553
Kuan	1.000	1.000	1.000	1.468	1.491	2.553

Estimación del SSIM: resultados con imágenes SAR[21]

Copeland (4096×2560)			Corpus Christi (4096×2560)	
Filter	SSIM		SSIM	
	Under H_0	Under H_1	Under H_0	Under H_1
Lee	0.680	0.326	0.828	0.614
Enhanced Lee	0.652	0.306	0.818	0.615
Kuan	0.652	0.306	0.818	0.615
Dam (1200×1200)			**Mississippi** (4096×4096)	
Filter	SSIM		SSIM	
	Under H_0	Under H_1	Under H_0	Under H_1
Lee	0.999	0.999	0.890	0.740
Enhanced Lee	0.999	0.999	0.884	0.737
Kuan	0.999	0.999	0.884	0.737

[21] Recuerde que $H_0 : \alpha = \beta = \gamma = 1$.

Temas tentativos para tesis: Alfredo Alegría[22]

- Campos aleatorios para modelar datos espacio-temporales sobre grandes porciones del planeta.
- Funciones de covarianza cruzada flexibles y su aplicación en el análisis de datos espacio-temporales multivariados.
- Métodos de estimación y predicción para campos aleatorios no-gaussianos y/o para grandes conjuntos de datos.

[22] E-mail: alfredo.alegria@usm.cl

Temas tentativos para tesis: Francisco Cuevas[23]

- Alineamiento y registro de funciones.
- Aproximaciones Bayesianas computacionales para campos aleatorios.
- Regresión penalizada para patrones puntuales.

[23] E-mail: francisco.cuevas@usm.cl

Temas tentativos para tesis: Felipe Osorio[24]

- Diagnóstico de influencia en el contexto de estimación máximo Lq-verosímil.
- Probabilidad de concordancia entre dos sistemas de medición usando P-splines robustos.
- Probabilidad de concordancia para varios instrumentos de medición.

[24] E-mail: felipe.osorios@usm.cl (no olvidar la 's' !)

Temas tentativos para tesis: Ronny Vallejos[25]

- Desarrollo de medidas de concordancia para datos espacio-temporales.
- Extensión de la probabilidad de concordancia para datos espaciales.
- Tamaño muestral efectivo para datos espacio-temporales.

[25] E-mail: ronny.vallejos@usm.cl

Modelación Estadística: Estudiantes de pregrado

- Nicolás Alfaro.
- Manuel Jara.
- Marcela Miranda.
- Eric Muñoz.
- Fabián Rubilar.
- Matías Sasso.
- Bastián Sepúlveda.
- Edgard González (2022).
- Sebastián Vera (2022).
- Pablo Huenchulao (2021).
- Gabriel Vidal (2021).
- Gabriel Molina (2020).
- Alberto Rubio (2019).
- Alexis Tapia (2019).
- Alonso Ogueda (2018).
- Francisco Alfaro (2017).
- Wilson Barraza (2017).
- Javier Pérez (2017).
- Ignacio Vásquez (2017).
- Ángelo Gárate (2016).
- Carlos Schwarzenberg (2016).
- Jean Paul Maidana (2015).
- Agustín Uribe (2015).
- Alfredo Alegría (2014).
- Claudio Henríquez (2014).
- Consuelo Moreno (2014).
- Jonathan Acosta (2013).
- Francisco Cuevas (2011).
- Danilo Pezo (2011).
- Jorge Littin (2006).

Modelación Estadística: Estudiantes de postgrado

- John Gómez.
- Fabian Ramírez (2022).
- Alonso Ogueda (2021).
- Carlos Schwarzenberg (2021).
- Francisco Alfaro (2019).
- Javier Perez (2019).
- Sebastián Torres (2019).
- Ángelo Garate (2018).
- Jonathan Acosta (2017).
- Diego Mancilla (2014).
- Francisco Cuevas (2013).
- Danilo Pezo (2013).
- Paola Carvajal (2006).
- Ronny Vallejos (1999).

Algunos trabajos recientes

Cuevas, F., Coeurjolly, J.F., Descary, M.H. (2022+).
Fast estimation of a convolution type model for the intensity of spatial point processes.
Spatial Statistics (to appear).

Osorio, F., **Vallejos, R.**, Barraza, W., Ojeda, S., Landi, M. (2022).
Statistical estimation of the structural similarity index for image quality assessment.
Signal, Image and Video Processing **16**, 1035-1042.

Vidal, G., Yuz, J., **Vallejos, R.**, **Osorio, F.** (2022).
Point-process modeling and divergence measures applied to the characterization of passenger flow patterns of a metro system.
IEEE Access **10**, 26529-26540.

Emery, X., **Alegría, A.** (2022).
The Gauss hypergeometric covariance kernel for modeling second-order stationary random fields in Euclidean spaces: Its compact support, properties and spectral representation.
Stochastic Environmental Research and Risk Assessment **36**, 2819-2834.

Acosta, J., **Alegría, A.**, **Osorio, F.**, **Vallejos, R.** (2021).
Assessing the effective sample size for large spatial datasets: A block likelihood approach.
Computational Statistics & Data Analysis **162**, 107282.

Alegría, A., Bissiri, P.G., Cleanthous, G., Porcu, E., White, P. (2021).
Multivariate isotropic random fields on spheres: Nonparametric bayesian modeling and L^p fast approximations.
Electronic Journal of Statistics **1**, 2360-2392.

Algunos trabajos recientes

Alegría, A., **Cuevas, F.**, Diggle, P., Porcu, E. (2021).
The $\mathcal{F}$-family of covariance functions: A Matérn analogue for modeling random fields on spheres.
Spatial Statistics **43**, 100512.

Alegría, A., Emery, X., Porcu, E. (2021).
Bivariate Matérn covariances with cross-dimple for modeling coregionalized variables.
Spatial Statistics **41**, 100491.

Emery, X., **Alegría, A.**, Arroyo, D. (2021).
Covariance models and simulation algorithm for stationary vector random fields on spheres crossed with Euclidean spaces.
SIAM Journal on Scientific Computing **43**, A3114-A3134.

Papaterra, M., Ojeda, S., Landi, M., **Vallejos, R.** (2021).
Strategy for selecting a quality index for images.
International Journal of Computer Information Systems and Industrial Management Applications **13**, 348-363.

Vallejos, R., Acosta, J. (2021).
Effective sample size for multivariate spatial processes with an application to soil contamination.
Natural Resource Modeling **34**, e12322.

Alegría, A. (2020).
Cross-dimple in the cross-covariance functions of bivariate isotropic random fields on spheres.
Stat **9**, e301.

Algunos trabajos recientes

Alegría, A., **Cuevas, F.** (2020).
Karhunen-Loève expansions for axially symmetric gaussian processes: Modeling strategies and L^2 approximations.
Stochastic Environmental Research and Risk Assessment **34**, 1953-1965.

Alegría, A., Emery, X., Lantuéjoul, C. (2020).
The turning arcs: A computationally efficient algorithm to simulate isotropic vector-valued gaussian random fields on the d-sphere.
Statistics and Computing **30**, 1403-1418.

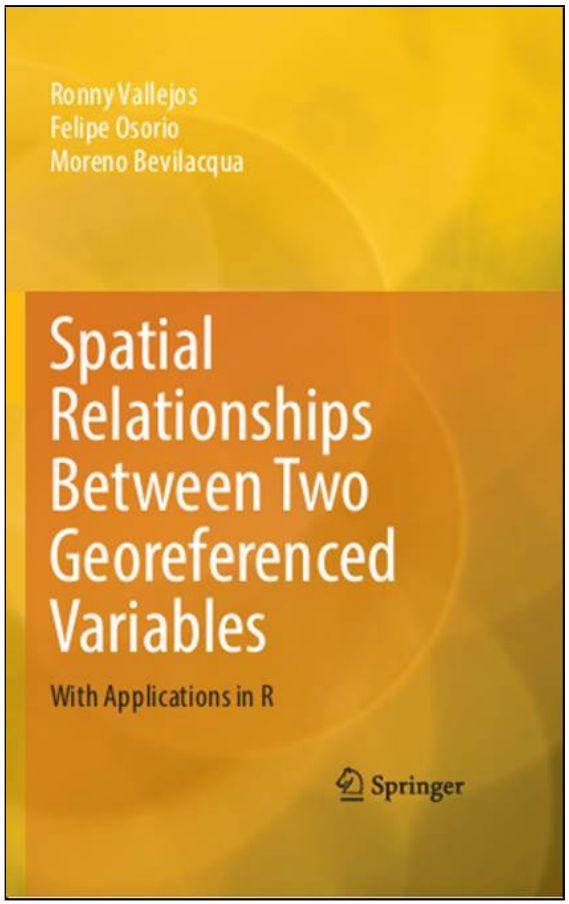
Choiruddin, A., Coeurjolly, J.-F. **Cuevas, F.**, Waagepetersen, R. (2020).
Regularized estimation for highly multivariate log Gaussian Cox processes.
Statistics and Computing **30**, 649-662.

Crudu, F., **Osorio, F.** (2020).
Bilinear form test statistics for extremum estimation.
Economics Letters **187**, 108885.

Emery, X., **Alegría, A.** (2020).
A spectral algorithm to simulate nonstationary random fields on spheres and multifractal star-shaped random sets.
Stochastic Environmental Research and Risk Assessment **34**, 2301-2311.

Vallejos, R., Pérez, J., Ellison, A., Richardson, A. (2020).
A spatial concordance correlation coefficient with an application to image analysis.
Spatial Statistics **40**, 100405.

Asociación espacial entre dos procesos estocásticos (libro Springer)



- Ronny Vallejos, Felipe Osorio (USM) y Moreno Bevilacqua (UAI).
- Asociación entre dos procesos espaciales:
  - procedimientos de test de hipótesis.
  - coeficientes de asociación/codispersión.
  - asociación entre imágenes.
- Paquetes en R: `SpatialPack` y `GeoModels`.

Vallejos, R., **Osorio, F.**, Bevilacqua, M. (2020). *Spatial Relationships Between Two Georeferenced Variables: With Applications in R*. Springer, Cham.

Comentarios/dudas?