

Tópicos de Estadística para Ciencia Política

January 16, 2018

Información del curso

Instructor: Javier Márquez Peña

Horario: Martes, 16:00-19:00 hrs

Lugar: Por definir

1 Objetivos del curso

El principal objetivo del curso es proporcionar a los alumnos las herramientas necesarias para entender y aplicar tres grandes familias de métodos estadísticos:

- Modelos Lineales Generalizados (GLM)
- Modelos mixtos/jerárquicos/multinivel
- Introducción al aprendizaje automático o aprendizaje de máquinas (*machine learning*)

El programa pretende que los alumnos dominen dichos métodos a partir fundamentos sólidos de teoría estadística y programación.

2 Laboratorio

El curso requiere el uso intensivo de software estadístico (R) y nociones de programación. Por lo anterior, todas las sesiones se llevarán a cabo en una sala de cómputo. Los laboratorios no son presenciales: consistirán en completar diversos módulos de la plataforma DataCamp (datacamp.com). En clase se darán instrucciones a los alumnos para ingresar a nuestra sesión privada. Durante el curso se asignarán algunos de los siguientes temas (por determinar):

- Introduction to R (4 hrs)

- Importing Data in R (Part 1) (3 hrs)
- Importing Data in R (Part 2) (3 hrs)
- Cleaning data in R (4 hrs)
- Importing and cleaning data in R: Case studies (4hrs)
- Data Visualization in R (4 hrs)
- Exploratory Data Analysis (4 hrs)
- Exploratory Data Analysis in R: Case Study (4 hrs)
- Exploring Pitch Data with R (4 hrs)
- Intermediate R (6 hrs)
- Intermediate R - Practice (4 hrs)
- Foundations of Probability in R (4 hrs)
- Foundations of Inference (4 hrs)
- Statistical Modeling in R (Part 1) (4 hrs)
- Statistical Modeling in R (Part 2) (4 hrs)
- Supervised Learning in R (4 hrs)
- Unsupervised Learning in R (4 hrs)

El sitio Kaggle (kaggle.com) hospedará una competencia sobre predicción para el proyecto final de clase.

3 Evaluación

	%
Asistencia	0%*
Participación y lecturas	10
Examen Parcial 1	10
Examen Parcial 2	20
Examen Parcial 3	20
Proyecto (Kaggle)	10
Tareas	30

La asistencia no otorgará puntos, pero los alumnos requieren asistir al 90% de las sesiones (incluyendo laboratorios) para tener derecho a calificación final.

En cada sesión, un alumno elegido al azar expondrá la lectura asignada.

La fecha de los exámenes será pactada con los alumnos dependiendo de la carga de trabajo de las demás materias del semestre.

4 Calendario y temas a cubrir

El calendario del curso puede sufrir cambios dependiendo del progreso de los alumnos. En dichos casos, las lecturas serán reasignadas.

4.1 Breve repaso de inferencia estadística

Sesgos estadísticos

- [Tversky and Kahneman \(1974\)](#). Heurísticos y sesgos cognitivos (10 págs).
- [Wheelan \(2013, capítulo 3\)](#). Juicio en descripción de datos (22 págs).

Introducción a muestreo y simulación estadística.

- [Salsburg \(2001, capítulo 17\)](#). Historia del muestreo (14 págs).
- [Wheelan \(2013, capítulo 8\)](#). Teorema del Límite Central (9 págs).

Estimación.

- [Wheelan \(2013, capítulo 10\)](#). Introducción (14 págs).
- [Agresti and Finlay \(1999, capítulo 5\)](#). Intervalos de confianza (25 págs).

Pruebas de hipótesis.

- [Wheelan \(2013, capítulo 9\)](#). Introducción (20 págs).
 - [Agresti and Finlay \(1999, capítulo 6\)](#). Significancia estadística (30 págs).
-

4.2 Modelo de regresión lineal

Modelo de Regresión Lineal. Repaso.

- [Wheelan \(2013, capítulo 11\)](#). Introducción al modelo de regresión (23 págs).
- [Gelman and Hill \(2006, capítulo 3\)](#). Introducción a modelo de regresión (18 págs).

Regresión múltiple. Repaso.

- [Agresti and Finlay \(1999, capítulo 10\)](#). Asociación y causalidad (13 págs).
- [Gelman and Hill \(2006, capítulo 4\)](#). Otros temas de estimación (21 págs).

Repasso y abusos de modelos de regresión

- [Wheelan \(2013, capítulo 12\)](#). Supuestos del modelo de regresión (13 págs).
 - [Schrodt \(2014\)](#). Supuestos del modelo de regresión (14 págs).
-

4.3 Modelos Lineales Generalizados

Regresión Logística

- [Gelman and Hill \(2006, capítulo 5\)](#)

Modelos Lineales Generalizados

- [Gelman and Hill \(2006, capítulo 6\)](#)

Modelos Lineales Generalizados, cont.

- [Gelman and Hill \(2006, capítulo 6\)](#)

Simulación estadística

- [Gelman and Hill \(2006, capítulo 7\)](#)
-

4.4 Modelos Multinivel

Introducción a estructuras multinivel

- Gelman and Hill (2006, capitulo 11)

Introducción a estructuras multinivel

- Gelman and Hill (2006, capitulo 11)

Modelo lineal multinivel

- Gelman and Hill (2006, capitulo 12)

Modelo lineal multinivel, cont.

- Gelman and Hill (2006, capitulo 13)

Modelo logístico multinivel.

- Gelman and Hill (2006, capitulo 14)
-

4.5 Introducción a machine Learning

Aprendizaje supervisado: Modelos de árbol

- James et al. (2013, capitulo 8)

Aprendizaje supervisado: Support Vector Machines

- James et al. (2013, capitulo 9)

Aprendizaje no-supervisado

- James et al. (2013, capitulo 10)

Aprendizaje no-supervisado, cont.

- James et al. (2013, capitulo 10)
-

Competencia Kaggle

References

- Agresti, A. and B. Finlay. 1999. *Statistical Methods for the Social Sciences*. Prentice Hall.
- URL:** https://books.google.com.mx/books?id=_WIPAQAAQAAJ
- Gelman, Andrew and Jennifer Hill. 2006. *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press.
- James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani. 2013. *An introduction to statistical learning*. Vol. 112 Springer.
- Salsburg, David. 2001. *The lady tasting tea: How statistics revolutionized science in the twentieth century*. Macmillan.
- Schrodt, Philip A. 2014. “Seven deadly sins of contemporary quantitative political analysis.” *Journal of Peace Research* 51(2):287–300.
- Tversky, Amos and Daniel Kahneman. 1974. “Judgment under uncertainty: Heuristics and biases.” *science* 185(4157):1124–1131.
- Wheelan, Charles. 2013. *Naked statistics: stripping the dread from the data*. WW Norton & Company.