



Análisis de datos

Laboratorio 1: Análisis estadístico

Profesor Max Chacón

Ayudante Adolfo Guzmán Leiva

Objetivos

Estudiar e interpretar los datos correspondientes a cada base de datos. Para ello es necesario explicar de forma detallada el significado de clases, atributos y sus valores, lo que permitirá obtener el correcto análisis del problema planteado.

Aspectos importantes a considerar

Para obtener los resultados y cumplir los objetivos del laboratorio, se debe tener en cuenta los siguientes puntos:

- Buscar la descripción de los datos en la página dada en la parte de observaciones y así lograr el estudio correcto del problema. Además, deben consultar los diferentes papers relacionados, los cuales han sido seleccionados porque utilizan dicha base de datos. De ésta manera obtendrá una interiorización apta para poder desarrollar los sucesivos laboratorios, pues utilizará la misma base de datos durante todo el semestre.
- Para el estudio de los atributos de su problema, se recomienda, además de la descripción detallada de éstos, deben aplicar técnicas de estadística descriptiva e inferencial según corresponda, tales como: Medidas de centralización (media, moda y mediana), distribución de probabilidades y Medidas de dispersión (rango y varianza), tests de hipótesis, análisis de varianza, etc... Para ello, es necesario utilizar “R” <http://www.r-project.org/> y así obtener un conocimiento un poco más profundo de la base de datos, por ejemplo las relaciones que existen entre las variables del problema mediante visualización. Finalmente, si es que es posible, se recomienda incluir información relevante entregada por un experto en el área de su problema, que ayude determinar la importancia de los atributos y descripción del problema.
- Archivos de datos: archivo *.data, archivo *.names

El informe se debe regir por el reglamento de titulación v 1.3, apéndice C, apartado C.3 y contener los siguientes puntos:

Informe	Puntos a evaluar	Porcentaje
	Presentación, ortografía y redacción.	10%
	Introducción. (Máximo 1 página)	15%
	Descripción del problema. Debe contener dos secciones: “Descripción de la base de datos” y “Descripción de clases y variables”. En este contexto debe quedar explícito el significado de cada una de las variables involucradas, como también las instancias de cada una de ellas. (Máximo 8 páginas)	20%
	Análisis Estadístico e Inferencial según corresponda. (Máximo 12 páginas)	25%
	Conclusiones respecto al problema y respecto al análisis del conjunto de datos. (Máximo 2 páginas)	25%
	Referencias: usar formato APA 6	5%

Observaciones

- Todas las consultas deben ser realizadas al mail adolfo.guzman@usach.cl
- La entrega debe ser subida al sitio Web Usach-Virtual hasta las 23.55 hrs del día Sábado 2 de Septiembre.
- La información de las bases de datos se encuentra en la página <http://archive.ics.uci.edu/ml/>
- Es necesario realizar TODAS las experiencias para aprobar el laboratorio
- Cualquier página más allá del máximo permitido no será revisada.
- Una entrega atrasada será evaluada con dos puntos menos y un puntos menos por cada día extra de retraso.