



Universidad de Santiago de Chile
Departamento de Ingeniería Informática
Análisis de Datos

Profesor: Dr. Max Chacón Pacheco
Ayudante: Gustavo Hurtado A.

Laboratorio 1 - Análisis estadístico

1. 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.

2. 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*

3. Informe

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

- Ortografía, redacción y formato **(10 %)**.
- Introducción (máximo 1 plana) **(15 %)**.
- Descripción del problema: En este contexto debe quedar explícito el significado de cada una de las variables involucradas, como también las instancias de cada una de ellas. Se debe mencionar las siguientes secciones (máximo 8 páginas) **(20 %)**:
 - Descripción de la base de datos.
 - Descripción de clases y variables.
- 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 %)**

4. Observaciones

- Todas las consultas deben ser realizadas al correo gustavo.hurtado@usach.cl
- La entrega debe ser subida al sitio Campus Virtual hasta las 23.55 hrs el 22 de Abril del 2022.
- La información de las bases de dato 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 punto menos por cada día extra de retraso.
- La detección de copia será evaluada con la nota mínima