



Universidad de Santiago de Chile  
Facultad de Ingeniería  
Departamento de Ingeniería Informática

**Asignatura** : Análisis de Datos e Inteligencia de Negocios  
**Programa** : Ingeniería Civil Informática  
Ingeniería de Ejecución en computación e informática  
**Profesor** : Dr. Felipe-Andrés Bello Robles  
**Ayudante** : NA

**Fecha Entrega** : 2 de mayo de 2019.

## TALLER 1: Planteamiento del problema y análisis estadístico

### Objetivos:

- Definir el problema a resolver mediante el conjunto de datos seleccionado
- 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 instancias, 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 indicada en la parte de observaciones y así lograr el estudio correcto del problema., donde **deben** consultar los diferentes **artículos** 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.
- Para el estudio de los atributos de su problema, se solicita además de la descripción detallada de éstos, aplicar técnicas de estadística descriptiva e inferencial según corresponda (naturaleza del atributo binario, real, etc...), 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 (ANOVA), 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 o inferencia estadística.
- 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: generalmente archivo \*.data, archivo \*.names

**Informe:**

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

<b>Estructura del Trabajo</b>	<b>Puntos a evaluar</b>	<b>Porcentaje</b>
	Presentación, ortografía y redacción, Referencias en formato APA 6	10%
	Introducción.	10%
	Descripción del problema: <ul style="list-style-type: none"><li>• Debe contener dos secciones:<ul style="list-style-type: none"><li>○ “Descripción de la base de datos”: Problema</li><li>○ “Descripción de clases y variables”: debe quedar explícito el significado de cada una de las variables involucradas, como también las instancias de cada una de ellas.</li></ul></li></ul>	20%
	Análisis Estadístico: <ul style="list-style-type: none"><li>• Estadística Descriptiva</li><li>• Estadística Inferencial</li><li>• Conclusiones sobre la muestra en base al análisis estadístico y en relación a la resolución del problema.</li></ul>	20%
	Conclusiones Respecto a : <ul style="list-style-type: none"><li>• Análisis estadístico de los datos</li><li>• Análisis estadístico y la resolución del problema</li><li>• Comparación con la literatura</li></ul> ¿Se resuelve el problema?	30%
	Código Fuente explicado o con comentarios.	10%

**Observaciones:**

Consultas al mail [felipe.bello@usach.cl](mailto:felipe.bello@usach.cl)

El trabajo debe ser presentado de forma escrita (informe) en horario de clases el día 2 de mayo de 2019.

La información de las bases de datos se encuentra en la página:

<http://archive.ics.uci.edu/ml/>

Para aprobar el laboratorio es obligación realizar **TODAS** las experiencias