

# Práctica 1

Jorge Humberto Sierra Florido  
2123065656@cua.uam.mx  
UAM Cuajimalpa  
Ingeniería en Computación  
Ciudad de México, México

María de Jesús Sánchez Zepeda  
2153068423@cua.uam.mx  
UAM Cuajimalpa  
Ingeniería en Computación  
Ciudad de México, México

## RESUMEN

Aquí va el abstract de la práctica...

### ACM Reference Format:

Jorge Humberto Sierra Florido and María de Jesús Sánchez Zepeda. 2021. Práctica 1. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1. INTRODUCCIÓN

En el análisis de datos es importante la limpieza de los datos previos a realizar modelos predictivos con ellos. Con limpieza nos referimos al tratamiento de los datos faltantes con el fin de no inducir un sesgo en el modelo.

Un método para encontrar relaciones entre varias variables de un conjunto de datos es la regresión lineal.

Se tienen dos conjuntos de datos de diferentes tamaños a los cuales se les desea encontrar un modelo una ecuación que modele la relación entre variables determinadas.

Para ello se limpiarán los datos de cada conjunto, se realizará una regresión lineal para posteriormente evaluar dicho modelo. Texto introductorio al tema en que se enfoca la práctica y lo que se desarrollará en ella. Se debe escribir un texto que introduzca el tema de la práctica, definición del problema, los objetivos, motivación, y resultados esperados.

## 2. CONCEPTOS PREVIOS

## 3. METODOLOGÍA

Se inicio la practica revisando los datasets proporcionados, estos corresponden a los datos de temperatura y salinidad medidas, así como las características de algunos vehículos.

El primer paso fue cargar los datos en R a fin de poder analizar los datasets. para esta carga se usaron las siguientes instrucciones:

### Listing 1: lectura de los datos

```
data = read.csv("water.csv", header=TRUE)
data2 = read.table("mtcars.txt", header=TRUE, sep = " ")
```

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference'17, July 2017, Washington, DC, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

TdegC		Salnty	
Min.	1.44	Min.	28.43
1st Qu.	7.68	1st Qu.	33.49
Median	10.06	Median	33.86
Mean	10.80	Mean	33.84
3rd Qu.	13.88	3rd Qu.	34.20
Max.	31.14	Max.	37.03
NA's	10963	NA's	47354

Cuadro 1: Resumen de el dataset water.csv

Una vez cargada la información se procedió a obtener el resumen de la misma con la siguiente instrucción(1)

```
summary(data)
```

En la cual podemos observar que existen varios registros vacíos, 10963 en el campo de temperatura y 47354 en la salinidad

## 4. RESULTADOS

Describir con detalle todos los resultados. Hacer una discusión de lo obtenido. Se debe mostrar un enfoque analítico sobre los resultados generados. Mostrar absolutamente todos los productos obtenidos debido a los pasos mostrados en la metodología.

## 5. CONCLUSIONES Y REFLEXIONES

Conclusiones generales de la práctica. Añadir una reflexión analítica por cada miembro del equipo.

## REFERENCIAS