

## PROGRAMA

# DIPLOMADO EN ANÁLISIS DE DATOS CON R PARA LA ACUICULTURA

*2da versión*

**1 OCTUBRE - 31 NOVIEMBRE 2021**

### DESCRIPCIÓN DEL DIPLOMADO

El **Diplomado** tiene como propósito que los estudiantes desarrollen habilidades para realizar análisis de datos en acuicultura usando el lenguaje de programación R y Rstudio-cloud. Durante el diplomado se entregan los fundamentos teóricos de la ciencia moderna del análisis de datos, así como de las competencias prácticas claves para el almacenamiento, lectura, procesamiento y presentación de resultados. Bajo el paradigma de la investigación reproducible el diplomado desarrolla en los estudiantes habilidades para trabajar con Rmarkdown y GitHub. Finalmente, durante el curso los alumnos aprenderán a realizar análisis exploratorio de datos, inferencia estadística, diferentes modelos de regresión lineal simple, múltiple y logística, así como algunos análisis multivariantes como los análisis de cluster y los análisis de componentes principales. Los contenidos del curso se explican usando ejemplos de datos y casos de estudio en diferentes ámbitos de la producción en acuicultura.

### PERFIL DEL PARTICIPANTE

Profesionales o graduados relacionados con las ciencias de la acuicultura, incluyendo Acuicultores, Agropecuarios, Biólogos, Biólogos Marinos, Biotecnólogos, Ingenieros Pesqueros, Ingenieros Agrónomos, Oceanógrafos, Veterinarios o Zootecnistas.

### DURACIÓN

Horas sincrónicas: 50 horas

Horas asincrónicas: 54 horas

Total: 104 horas

## PROFESORES

### **José Andrés Gallardo Matus**

Biólogo Marino por la Universidad Católica de la Santísima Concepción, Chile.

Doctor en Ciencias por la Universidad de Chile.

Profesor Adjunto de la Pontificia Universidad Católica de Valparaíso.

### **María Angélica Rueda Calderón**

Licenciada en Matemáticas por la Universidad Industrial de Santander, Colombia.

Doctora en Ciencias Agropecuarias por la Universidad Nacional de Córdoba, Argentina.

Investigadora post-doctoral de la Pontificia Universidad Católica de Valparaíso.

## FECHAS DE CLASES

### **Presentación del curso y de los participantes.**

jueves 30 septiembre – 18:00 - 20:00 PM

### **Clases sincrónicas**

martes de 18:00 a 20:00 horas

octubre: 05, 12, 19, 26.

noviembre: 02, 09, 16, 23.

3 sábados al mes de 10:00 AM – 14:00 PM

02, 09, 16, 23 de octubre

06, 13, 20, 27 de noviembre

### **Cierre del curso y entrega de certificados de participación**

Jueves 16 de diciembre - 18:00-18:30 PM

## REQUISITOS DE CONOCIMIENTOS TEÓRICOS Y COMPETENCIAS PRÁCTICAS

### **Título profesional o licenciatura.**

**Programación básica con R:** Deseable pero no excluyente. Los alumnos sin experiencia previa en programación con R deben considerar 4 horas de estudio y autoaprendizaje adicional por semana para poder alcanzar una comprensión avanzada de los objetivos de aprendizaje del curso.

**Inglés:** Los softwares R, Rstudio, Rmarkdown, GitHub y todas las librerías de análisis estadístico que se usarán en el curso solo están disponibles en inglés. Alumnos sin competencias de lectura en inglés no deberían tomar el curso.

## CONTENIDOS

### UNIDAD 1. INVESTIGACIÓN REPRODUCIBLE Y ANÁLISIS EXPLORATORIO DE DATOS

#### *Introducción a la unidad*

Se entregan los fundamentos de la investigación reproducible y del análisis exploratorio de datos. Se discuten los aspectos beneficiosos de hacer investigación reproducible para el investigador, así como los principales criterios que determinan que una investigación sea reproducible o no. Respecto del análisis exploratorio de datos se definen y clasifican los distintos tipos de variables aleatorias, y se entregan herramientas que permiten observar y predecir el comportamiento de las variables aleatorias bajo distintos tipos de distribución de probabilidad como la distribución Normal, Bernoulli, Binomial negativa, la distribución normal multivariante, entre otros.

#### *Resultado de aprendizaje de la unidad*

Al finalizar la unidad, el estudiante será capaz de aplicar los fundamentos de la investigación reproducible y del análisis exploratorio de datos usando el lenguaje de programación R.

#### *Palabras clave*

R, Rstudio, Rmarkdown, variables aleatorias, distribución de probabilidad, análisis exploratorio de datos.

#### *Subtítulos*

Subtítulo 1.1.- Investigación reproducible con R, Rstudio, Rmarkdwon y GitHub.

Subtítulo 1.2.- Variables aleatorias y distribuciones de probabilidad.

Subtítulo 1.3.- Análisis exploratorio de datos.

### UNIDAD 2. PRUEBAS DE CONTRASTES DE HIPÓTESIS PARAMÉTRICAS Y NO PARAMÉTRICAS

#### *Introducción a la unidad*

Se entregan los fundamentos de la inferencia estadística y de las pruebas de contraste de hipótesis. Respecto del análisis de datos de acuicultura y sobre la base de estudios de caso se entregan herramientas para aplicar diferentes test estadísticos incluyendo: prueba de t-student para la correlación de variables continuas y para diferencia de 2 medias, prueba de F para la diferencia de más de 2 medias (análisis de varianza), prueba de  $\chi^2$  para asociación de dos o más variables categóricas (tablas de contingencia), prueba de Wilcoxon para comparación de 2 muestras independientes, entre otras pruebas no paramétricas de uso común en investigación en acuicultura.

#### *Resultado de aprendizaje de la unidad*

Al finalizar la unidad, el estudiante será capaz de aplicar los fundamentos de la inferencia estadística en el análisis de datos de acuicultura y de datos de ciencias agropecuarias mediante el lenguaje de programación R.

#### *Palabras clave*

Parámetro, estadístico, correlación, permutación, combinación, inferencia estadística, contraste de hipótesis y análisis de sobrevivencia.

### *Subtópicos*

Subtópico 2.1.- Pruebas de contraste de hipótesis paramétrica.

Subtópico 2.2.- Pruebas de contraste de hipótesis no paramétrica.

## **UNIDAD 3. MODELOS LINEALES Y ANÁLISIS MULTIVARIANTE**

### *Introducción a la unidad*

Se entregan los fundamentos del uso de los modelos lineales y del análisis multivariado en la investigación en acuicultura. Se discute el uso de la aplicación de modelos lineales y del análisis multivariado para explicar, modelar o predecir la relación de una variable respuesta  $Y$  con una o más variables predictoras  $X$ . Se revisan los supuestos de cada tipo de modelo lineal y las técnicas usadas en análisis multivariado como herramientas exploratorias en el análisis de datos. Aplicando un diagnóstico apropiado en cada caso. Se verán los modelos lineales mixtos o lineales generalizados como una alternativa para corregir el modelo lineal cuando no se cumplen los supuestos.

### *Resultado de aprendizaje de la unidad*

Al finalizar la unidad, el estudiante será capaz de seleccionar y aplicar diferentes modelos lineales y análisis multivariado en el contexto de la investigación en acuicultura y en ciencias agropecuarias.

### *Palabras clave*

Regresión lineal, regresión lineal múltiple, regresión logística, análisis de cluster, análisis de componentes principales.

### *Subtópicos*

Subtópico 3.1.- Modelos lineales.

Subtópico 3.2.- Modelos lineales generalizados.

Subtópico 3.3.- Modelos mixtos.

Subtópico 3.4.- Técnicas de Análisis Multivariante.

## **RECURSOS DE APRENDIZAJE**

- **Rstudio:** Cuenta Rstudio.cloud por 3 mes.

- **Material docente:** Clases, guías de aprendizaje y códigos de programación para el análisis de datos con R.