

## DIPLOMADO EN

### **Análisis de datos con R e investigación reproducible para Biociencias.**

*Ira versión*

**28 AGOSTO – 30 DE NOVIEMBRE 2022**

#### DESCRIPCIÓN DEL DIPLOMADO

El Diplomado en *Análisis de datos con R e investigación reproducible para Biociencias* entrega competencias teóricas y prácticas de la ciencia moderna de datos y de la investigación reproducible con énfasis en Biociencias. Durante el diplomado, los alumnos desarrollarán habilidades fundamentales para el almacenamiento, lectura, procesamiento, análisis y presentación de resultados de investigación usando los softwares R, Rstudio y Rmarkdown. Además, bajo el paradigma de la investigación reproducible los estudiantes utilizarán el software de control de versiones Github para el desarrollo de un proyecto personal de análisis de datos. Los contenidos del curso se explican usando diferentes casos de estudio con énfasis en Biociencias.

#### OBJETIVOS DE APRENDIZAJE

Al finalizar el Diplomado los alumnos serán capaces de:

- Realizar análisis exploratorio de datos.
- Aplicar modelos básicos de inferencia estadística paramétrica y no paramétrica.
- Seleccionar y aplicar modelos lineales, no lineales y multivariados de modelamiento predictivo.
- Compartir y presentar resultados de análisis de datos de manera moderna y profesional mediante reportes dinámicos generados con el software Rmarkdown.
- Trabajar bajo el paradigma de la investigación reproducible utilizando de forma autónoma el software de control de versiones GitHub.

#### PERFIL DEL PARTICIPANTE

Profesionales o graduados de Agronomía, Biología, Biología Marina, Bioquímica, Biotecnología, Microbiología, veterinaria y áreas afines a la investigación con organismos vivos y recursos biológicos.

## DURACIÓN

Horas sincrónicas: 50 horas

Horas asincrónicas: 54 horas

Total: 104 horas

## HORARIO DE CLASES

Cada martes de 18:00 – 20:00 PM

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

## REQUISITOS DE POSTULACIÓN

- **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.

## RECURSOS DE APRENDIZAJE

- **Rstudio:** Cuenta Rstudio.cloud por 4 mes.
- **Material docente:** Clases, guías de aprendizaje y códigos de programación para el análisis de datos con R.

## VALOR DEL DIPLOMADO

**Valor total:** CLP\$ 1.200.000; (US\$1.200)

**Postulación y matrículas:** Hasta el 28 de agosto de 2022 o hasta que se completen los cupos disponibles.

**Formas de pago:** Contado, cuotas (hasta 4), tarjetas de crédito y orden de compra empresas.

## BECAS

**Beca “ALUMNI PUCV”:** 30% de descuento para ex alumnos graduados de la Pontificia Universidad Católica de Valparaíso.

**Nueva Beca “MUJERES EN BIOCIENCIAS”:** 30% de descuento para mujeres que trabajen en Biociencias.

**Beca “POSTGRADO EN BIOCIENCIAS”:** 40% de descuento para estudiantes de magister o doctorado en disciplinas afines a las Biociencias, chilenos o extranjeros.

**Beca “FUNCIONARIO PÚBLICO”:** 50% de descuento para funcionarios de instituciones públicas o estatales chilenas o extranjeras.

Las becas son individuales y no son transferibles a otras personas o cursos, las becas no son acumulables.

## **DIRECTOR Y PROFESOR DEL PROGRAMA**

**José Gallardo Matus**

Doctor en Ciencias por la Universidad de Chile.

Profesor Adjunto de la Pontificia Universidad Católica de Valparaíso.

## **PROFESOR**

**María Angélica Rueda Calderón**

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

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

## **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.

##### *Palabras clave*

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

##### *Subtópicos*

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

Subtópico 1.2.- Variables aleatorias y distribuciones de probabilidad.

Subtópico 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. Con diferentes estudios de caso los alumnos podrán 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 Biociencias.

##### *Palabras clave*

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

#### *Subtópicos*

Subtópico 2.1.- Inferencia estadística y formulación de hipótesis.

Subtópico 2.2.- Pruebas de contraste de hipótesis.

Subtópico 2.3.- Anova, pruebas a posteriori.

Subtópico 2.4.- Evaluación de supuestos de las pruebas paramétricas.

Subtópico 2.6.- Tablas de contingencia.

Subtópico 2.7.- Pruebas de Wilcoxon y Kruskal – Wallis.

### **UNIDAD 3. MODELOS LINEALES, ANÁLISIS DE SUPERVIVENCIA Y ANÁLISIS MULTIVARIANTE**

#### *Introducción a la unidad*

Se entregan los fundamentos del uso de los modelos lineales, de los análisis de supervivencia y del análisis multivariado en la investigación en Biociencias. Se discute el uso de la aplicación de modelos lineales, de los análisis de supervivencia 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, 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.

#### *Palabras clave*

Regresión lineal, regresión lineal múltiple, regresión logística, análisis de supervivencia, 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.- Análisis de supervivencia.

Subtópico 3.5.- Técnicas de Análisis Multivariante.

### **EVALUACIÓN Y CERTIFICACIÓN**

La evaluación se realiza mediante un proyecto personal de análisis de datos. El proyecto se desarrolla en dos partes:

Parte 1. Análisis exploratorio de datos. Considera evidencias de aprendizaje de unidad 1 y pondera un 40% de la nota final.

Parte 2. Análisis estadístico y presentación de resultados. Considera evidencias de aprendizaje de unidad 2 y 3, pondera un 60% de la nota final.

**Nota de aprobación:** La nota mínima de aprobación es un 4,0, en una escala de 1-7, con un 60% de exigencia. Solo los alumnos que aprueban el diplomado recibirán certificado de aprobación.

