

Guía para la presentación de Programaciones de Cursos de Posgrado (RESOLUCIÓN N° 691/06 C.S.)

A. DATOS GENERALES DEL CURSO:

- 1. DENOMINACIÓN DEL CURSO:** Programación en R para análisis de datos.
- 2. UNIDAD ACADÉMICA RESPONSABLE:**
Facultad de Ciencias Agrarias - UNNE.
- 3. DURACIÓN:** 5 días, del 17 al 21 de febrero 2020.
- 4. CARGA HORARIA Y CRÉDITOS:**
Horas teórico-prácticas presenciales: 40. Créditos: 2
- 5. ARANCEL:** \$3800 (pesos tres mil ochocientos)
- 6. DESTINATARIOS DEL CURSO:**
Docentes – Investigadores en el área de las ciencias: Biológicas, Exactas, Sociales y Económicas.
- 7. CUPO:**
Mínimo: 15
Máximo: 30
- 8. CERTIFICACIONES A OTORGAR:**
Únicamente se otorgará certificado de aprobación del Curso.
- 9. DOCENTES A CARGO:**
Coordinadora:
Dra. Laura I. Giménez
Profesoras Dictantes:
Lic. en Química (Dra.) Roxana Noelia Villafañe (20 hs).
Lic en Biología (Dra.) María Florencia D'Andrea (20 hs).
Auxiliar:
Lic. en Sistemas Patricia Loto (10 hs).
- 10. FUENTES DE FINANCIAMIENTO:**
Los recursos con que se autofinanciará el Curso provienen exclusivamente del arancel.
- 11. PRESUPUESTO DETALLADO DE GASTOS (aproximados):**

| Descripción | Monto |
|---------------------|--------------|
| Coordinación | 1000 |
| Horas Cátedra: | |
| • Profesor 20 horas | 9000 |
| • Profesor 20 horas | 7000 |
| • Auxiliar 10 horas | 2000 |
| Refrigerio | 3000 |
| Certificados | 1520 |
| *Pasaje, hotel | 10560 |
| ½ viático (4 días) | 4530 |
| **Gastos (nafta) | 2500 |
| Posgrado | 5700 |
| Total | 46410 |

Los honorarios correspondiente a Profesores y Auxiliar solicitados superan la cantidad de horas del curso debido a que las clases prácticas requieren de la presencia de 2 docentes en simultaneo.

* El pasaje y Estadía corresponde a la Profesora M. Florencia D' Andrea.

**Los gastos de nafta trayecto Resistencia-Corrientes vehículo particular de la Auxiliar Patricia Loto

12. En caso de haber saldo remanente del curso irá para la Cátedra de Cálculo Estadístico y Biometría.

13. El curso se autofinanciará.

B. PROGRAMACIÓN DIDÁCTICA DEL CURSO:

1. FUNDAMENTACIÓN:

El lenguaje de programación R se ha vuelto popular en varias disciplinas. Siendo un fenómeno relativamente reciente, muchos de los científicos/os que lo emplean con fines estadísticos carecen de entrenamiento en programación. Esto conlleva a que herramientas útiles para la visualización y limpieza de los datos sean poco conocidas, ya que son consideradas accesorias.

¿Cuáles son las ventajas de adoptar R en nuestro flujo de trabajo? ¿Es R meramente una herramienta estadística? El objetivo de este curso es brindar una iniciación a este versátil lenguaje de

programación a través de sus potencialidades, que incluyen tanto generar gráficos interactivos, reportes, así como aplicaciones.

Siendo un problema en la actualidad la falta de la reproducibilidad en las publicaciones científicas, presentaremos algunos conceptos básicos en relación al análisis de datos reproducible. Además, se proveerá de materiales, sugerencias de foros y comunidades de usuarios para que los asistentes puedan acceder a fuentes de soporte continuo y facilitar así la inclusión de la programación como práctica activa en los flujos de trabajo cotidianos de sus investigaciones.

Por intermedio de la secretaría de posgrado se realizó una consulta online para evaluar posibles interesados la que dió como resultado un alto porcentaje de interés sobre un total de respuestas aproximado de 25 personas que cubren el cupo mínimo.

2. OBJETIVOS DEL CURSO:

GENERAL:

- ✓ Aprender generalidades de programación en el entorno del software R y Rstudio para análisis, visualización y manipulación de datos.

TERMINALES:

- ✓ Entender la lógica del pensamiento computacional.
- ✓ Desarrollar habilidades en el manejo de datos.
- ✓ Aprender los fundamentos de los datos ordenados.
- ✓ Aprender a realizar visualizaciones que sean efectivas y aptas para ser publicadas.
- ✓ Entender la necesidad de documentar el código que se desarrolla.

3. CONTENIDOS:

Nociones de programación

- ✓ Tipos de objetos.
- ✓ Estructuras de control.

Manipulación de datos:

- ✓ Ordenación de datos con tidyrr.
- ✓ dplyr.

Visualización de datos:

- ✓ Gramática de gráficos con ggplot2.

Modelización de datos

Documentación:

- ✓ RMarkdown.
- ✓ Prácticas para la ciencia reproducible

Detalle por temas:

Tema 1: Generalidades y nociones de programación. R e interfaz de Rstudio. Trabajo con proyectos. Paquetes. Programación orientada a objetos. Cómo obtener ayuda. Tipos de datos. Acceso a los elementos de los diferentes tipos de datos. Funciones. Estructuras de control. Programación modelo *tidyverse* y uso del operador `%>%`. Nociones de programación funcional y uso del paquete *purrr*. Ejercicios de aplicación

Tema 2: Manipulación de datos. Ingreso de datos. Datos ordenados (*tidy data*). Paquetes *tidyr* y *dplyr*. Funciones `select()`, `filter()`, `mutate()`, `summarise()`, `arrange()`, `group_by()`. Otras funciones de *dplyr*. Ejercicios de aplicación

Tema 3: Visualización de datos. Visualización efectiva. Función `plot()` y `qplot()`. Gramática de gráficos. Paquete *ggplot2*. Capas de un gráfico: datos, parámetros estéticos, parámetros geométricos, uso de facets, estadísticas, coordenadas, temas. Ejercicios de aplicación. Breve introducción a la visualización interactiva con el paquete *Plotly* y a web-apps con el paquete *Shiny*. Mapas.

Tema 4: Modelización. Generalidades de la modelización de datos. Nociones de regresión y clasificación. Ejercicios de aplicación. Uso del paquetes para modelización de datos en *tidymodels*.

Tema 5: Documentación y ciencia reproducible. El camino hacia la ciencia reproducible: algunas prácticas deseables al momento de generar código para nuestras publicaciones. Reportes en *RMarkdown* y una introducción al control de versiones. Comunidades de usuarios.

4. METODOLOGÍAS DE ENSEÑANZA:

La modalidad de dictado será:

- ✓ Desarrollo de la parte teórica de parte de las docentes a cargo, que a su vez tendrán ejemplos de aplicación.
- ✓ Luego del desarrollo de la parte teórica de cada tema se realizarán actividades prácticas utilizando R con la interfase Rstudio. Cada participante realizará los ejercicios de la guía de trabajos prácticos y evaluará si el tema dado se puede aplicar a su propio trabajo.
- ✓ En cada uno de los temas dados se mostrarán Aplicaciones y ejemplos.

5. INSTANCIAS DE EVALUACIÓN:

Presentación de un trabajo integrador individual.

6. REQUISITOS DE APROBACIÓN DEL CURSO:

Ochenta por ciento de asistencia.

Aprobación del trabajo final escrito.

Escala numérica de cero a diez, requisito para la aprobación del curso nota mínima de seis.

7. Cronograma estimativo:

| Febrero 2020 | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 17 Lunes | 18 Martes |
| Tema 1 Teórico 9 a 13 hs. Dra. Villafañe, Roxana N. Dra. D'Andrea, M. Florencia. Práctico 14 a 15.30 Dra. Villafañe, Roxana N. 15:30 a 18 hs. Lic. Patricia Loto | Tema 1 Teórico 9 a 13 hs. Dra. Villafañe, Roxana N. Dra. D'Andrea, M. Florencia. Práctico 14 a 15.30 Dra. Villafañe, Roxana N. 15:30 a 18 hs. Lic. Patricia Loto |
| 19 Miércoles | 20 Jueves |
| Tema 2 y 3 Teórico 9 a 13 hs. Dra. Villafañe, Roxana N. Dra. D'Andrea, M. Florencia. Práctico 14 a 15.30 Dra. Villafañe, Roxana N. 15:30 a 18 hs. Lic. Patricia Loto | Tema 3 y 4 Teórico 9 a 13 hs. Dra. Villafañe, Roxana N. Dra. D'Andrea, M. Florencia. Práctico 14 a 15.30 Dra. Villafañe, Roxana N. 15:30 a 18 hs. Lic. Patricia Loto |
| 21 Viernes | |
| Tema 5. Teórico 9 a 13 hs. Dra. Villafañe, Roxana N. Dra. D'Andrea, M. Florencia. Práctico 14 a 18 hs. Lic. Patricia Loto Dra. Villafañe, Roxana N. | |

8. INFRAESTRUCTURA Y EQUIPAMIENTO NECESARIOS:

Aula, Cañón.

9. BIBLIOGRAFÍA:

- ✓ Bryan, J. (2015) How to name files. Disponible en <https://speakerdeck.com/jennybc/how-to-name-files> Visitado el 20/06/2019
- ✓ Bryan, J. (2017) Happy Git and GitHub for useR. Disponible en <https://happygitwithr.com/> Visitado el 20/06/2019.
- ✓ Bryan, J. (2019) What they forgot to teach you about R. Disponible en <https://whattheyforgot.org/> Visitado el 20/06/2019.
- ✓ Gandrud, C. (2016). *Reproducible research with R and R studio*. Chapman and Hall/CRC.
- ✓ Grolemund, G; Wickham, H. (2017) *R for data science*. O'Reilly. Disponible en <https://r4ds.had.co.nz/>
- ✓ Grolemund, G. (2014) *Hands on programming with R. Write your own functions and simulations*. O'Reilly. Disponible en <https://rstudio-education.github.io/hopr/>
- ✓ Hicks, S C y Peng, R D. (2019) Elements and Principles of Data Analysis. arXiv. Disponible en <https://arxiv.org/pdf/1903.07639.pdf>
- ✓ Ismay C. y Kim A.Y. (2019) Statistical Inference via Data Science: A modern dive into R and the tidyverse. Disponible en: <https://moderndive.com/index.html>
- ✓ Matloff, N. (2011). *The art of R Programming. A tour of statistical software design*. No Starch Press.
- ✓ Peng, R.D. (2014) *R Programming for data science*. Lulu.com
- ✓ R Core Development Team (2000) Introducción a R. Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos. Disponible en <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- ✓ ROpenSci Reproducibility in Science. A Guide to enhancing reproducibility in scientific results and writing. Disponible en <http://ropensci.github.io/reproducibility-guide/> Visitado el 20/06/2019
- ✓ Santana, J.S; Farfán, E.M. (2014) *El arte de programar en R. Un lenguaje para la estadística*. Disponible en [https://cran.r-project.org/doc/contrib/Santana El arte de programar en R.pdf](https://cran.r-project.org/doc/contrib/Santana%20El%20arte%20de%20programar%20en%20R.pdf)
- ✓ Wickham, H. (2009) *ggplot2. Elegant graphics for data analysis*. Springer.
- ✓ Wickham, H. (2014) *Tidy data*. Journal of statistical software. doi: 10.18637/jss.v059.i10

- ✓ Wickham, H. (2019) *Advanced R*. Second Edition. CRC Press. Disponible en <https://adv-r.hadley.nz/>
- ✓ Wilke, C. O. (2019) *Fundamentals of Data Visualization: A primer on making informative and compelling figures*. O'Reilly. Disponible en <https://serialmentor.com/dataviz/>
- ✓ Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., ... & Waugh, B. (2014). *Best practices for scientific computing*. PLOS Biology doi: 10.1371/journal.pbio.1001745
- ✓ Wilson, G.; Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., Teal, T. (2017). *Good enough practices in scientific computing*. PLOS Computational Biology. doi: 10.1371/journal.pcbi.1005510
- ✓ Xie Y; Allaire J.J., Golemund G. (2019) *R Markdown: The Definitive Guide*. CRC Press. Disponible en <https://bookdown.org/yihui/rmarkdown>
- ✓ Zumel, N., Mount, J. (2016) *Practical Data Science with R*. Manning Publications.