

UNIVERSIDAD DE CIENCIAS MÉDICAS DE LA HABANA

Escuela Nacional de Salud Pública

Título: Análisis de datos en fortificación de alimentos a gran escala con R.

PROGRAMA

TIPO DE ACTIVIDAD: Curso

COORDINADOR: Dr. Maicel Eugenio Monzón Pérez, Especialista de II grado en Bioestadística, Profesor auxiliar, M.Sc. en Informática en Salud

Centro de trabajo: Escuela Nacional de Salud Pública

CORREO ELECTRÓNICO: maicel@cencec.sld.cu, maicel@infomed.sld.cu

TELEFONOS: +54234317

COLECTIVO DE PROFESORES QUE IMPARTIRÁN EL CURSO

Nombre y apellidos	Categoría docente
Maicel Eugenio Monzón Pérez	Profesor auxiliar

MATRICULA PROPUESTA: 30

CENTRO ACREDITADO: ENSAP

ESPECIALIDAD QUE PROPONE: Bioestadística

FECHA PROPUESTA: 24-27 de marzo de 2025

LUGAR DONDE SE IMPARTIRÁ: Hotel Melia Habana

DESARROLLADO A TIEMPO: Completo, en modalidad presencial

TOTAL DE HORAS: 68

NIVEL: Internacional

DIRIGIDO A: Profesionales y técnicos que laboran la fortificación de alimentos.

FUNDAMENTACIÓN

La fortificación a gran escala de alimentos se ha consolidado como una intervención costo-efectiva clave para mejorar el acceso a micronutrientes esenciales en la dieta de poblaciones vulnerables a nivel global. Su éxito, sin embargo, depende críticamente de una evaluación rigurosa y basada en datos para asegurar que las estrategias implementadas alcancen de manera equitativa a quienes más lo necesitan. El análisis de datos, utilizando herramientas como R, es fundamental para modelar el impacto potencial de diferentes escenarios de fortificación, optimizar las estrategias existentes y dirigir los esfuerzos hacia las poblaciones más desatendidas.

Este curso ha sido diseñado para capacitar a profesionales de la salud pública, nutricionistas e investigadores en general en el uso de R para el análisis de datos relacionados con programas de fortificación de alimentos en Cuba. El curso surgió como necesidad del taller de fortificación de alimentos realizado en noviembre de 2024, impulsado por el Programa Mundial de Alimentos de las Naciones Unidas (PMA).

El curso ofrece una combinación de teoría, práctica intensiva y gamificación. Los participantes adquirirán las habilidades necesarias para leer datos desde ficheros con formatos comunes, ordenar y transformar los datos para realizar análisis relacionados con la fortificación de alimentos a gran escala, y presentar los resultados en forma de tablas y gráficos.

OBJETIVO GENERAL

Que los estudiantes sean capaces de aplicar técnicas para desarrollar análisis de datos consistentes en:

1. Importar datos desde ficheros con formato comunes (xls, csv, sav, dat, etc.).
2. Ordenar datos.

3. Realizar transformaciones.
4. Representar de forma gráfica y tabular los resultados de análisis estadísticos en fortificación de alimentos a gran escala.

PLAN TEMÁTICO

TEMA I: Generalidades del entorno y lenguaje R.

Objetivos temáticos:

Que los estudiantes conozcan:

- ¿Qué es R y por qué es útil en la fortificación de alimentos?
- Las principales características del lenguaje
- Instalación y configuración de R y RStudio.
- Interfaz de RStudio: Consola, script, ambiente, gráficos.
- Creación, listado y remoción de objetos en memoria
- Creación de objetos básicos (vectores, listas, data frames).
- Obtener ayuda
- Ejemplo de aplicaciones de R con interfaz visual (Rcmdr)

Contenidos:

Se tratarán contenidos clave como: ¿Qué es R y por qué es útil en la fortificación de alimentos? Instalación y configuración de R y RStudio. Interfaz de RStudio: Consola, script, ambiente, gráficos. Creación de objetos básicos (vectores, listas, data frames). Elementos de Markdown.

Horas Teóricas: 2 Horas Prácticas: 2- Horas Estudio Independiente: 8 Horas Totales: 12

TEMA II: Importación de datos

Objetivos temáticos:

Que los estudiantes sean capaces de manejar R para:

- Aprender a leer datos desde diferentes formatos.

Contenidos:

Se abordarán temas como:

1. Leer archivos CSV utilizando el paquete readr: Los estudiantes aprenderán a importar datos desde archivos CSV utilizando funciones como `read_csv()` y cómo manejar diferentes formatos de archivos.
2. Leer archivos Excel utilizando el paquete readxl: Los participantes serán capaces de importar datos desde archivos Excel (.xlsx) utilizando el paquete readxl y sus funciones asociadas.
3. Manejo de problemas comunes (codificación, separadores, valores perdidos): Al finalizar, los estudiantes podrán resolver problemas frecuentes al importar datos, como codificaciones incorrectas, separadores inesperados y valores perdidos.

H. Teóricas: 2 H. Prácticas: 2- H. Estudio Independiente: 8 H. Totales: 12

TEMA III: Ordenar datos

Objetivos temáticos:

Que los estudiantes sean capaces de manejar R para:

- Aprender a leer datos desde diferentes formatos.

Contenidos:

Se abordarán temas como:

1. Reorganizar columnas y filas con tidyr: Los estudiantes aprenderán a utilizar el paquete tidyr para reorganizar sus datos, asegurándose de que cada variable esté en su propia columna y cada caso en su propia fila.
2. Convertir datos de formato ancho a largo (y viceversa) con `pivot_longer()` y `pivot_wider()`: Los participantes serán capaces de transformar sus datos

entre formatos anchos y largos utilizando las funciones `pivot_longer()` y `pivot_wider()`, lo que facilitará el análisis y la visualización de los datos.

3. Separar y combinar columnas: Al finalizar, los estudiantes podrán separar columnas en varias utilizando `separate()` y combinar varias columnas en una sola con `unite()`, lo que les permitirá adaptar sus datos a las necesidades específicas de su análisis.

H. Teóricas: 2 H. Prácticas: 2- H. Estudio Independiente: 8 H. Totales: 12

TEMA IV: Manipulación de datos

Objetivos temáticos:

Que los estudiantes sean capaces de manejar R para:

- Aprender a transformar y resumir datos.

Contenidos:

Se abordarán temas como:

1. Filtrar, seleccionar y ordenar datos con `dplyr`: Los estudiantes aprenderán a utilizar las funciones `filter()`, `select()` y `arrange()` para seleccionar filas y columnas específicas de un data frame, filtrar datos según condiciones lógicas, y ordenar los datos en forma ascendente o descendente según una o varias columnas.
2. Crear nuevas variables con `mutate()`: Los participantes serán capaces de agregar nuevas columnas a un data frame utilizando la función `mutate()`, lo que les permitirá transformar y combinar variables existentes para crear nuevas variables relevantes para su análisis.
3. Resúmenes estadísticos básicos con `summarise()`: Al finalizar, los estudiantes podrán calcular estadísticas básicas como sumas, promedios y conteos para cada grupo de datos utilizando la función `summarise()`, lo que les facilitará resumir grandes conjuntos de datos de manera efectiva

H. Teóricas: 2 H. Prácticas: 2- H. Estudio Independiente: 8 H. Totales: 12

TEMA V: Representación tabular y gráfica.

Objetivos temáticos:

Que los estudiantes sean capaces de:

- Representar datos de forma tabular y gráfica

Contenidos:

Se abordarán temas como:

Creación de tablas con las bibliotecas gtsummary : Los estudiantes aprenderán a crear tablas resumidas y estadísticas utilizando las bibliotecas gtsummary y ggstatsplot, lo que les permitirá presentar sus resultados de manera clara y concisa.

Visualización de resultados con la biblioteca ggstatsplot: Los participantes serán capaces de crear gráficos atractivos y efectivos utilizando ggstatsplot, lo que les facilitará comunicar sus hallazgos de manera visual y comprensible.

Creación de informes con Markdown: Al finalizar, los estudiantes podrán generar informes profesionales utilizando Markdown, integrando texto, tablas y gráficos para presentar sus análisis de manera coherente y bien estructurada.

H. Teóricas: 2 H. Prácticas:2 - H. Estudio Independiente: 8 H. Totales: 12

PLAN CALENDARIO

TEMA	F.O.E.	HORAS Teóricas	HORAS Prácticas	Estudio Independiente	HORAS TOTALES
TEMA I: Generalidades del entorno y lenguaje R.	Conferencia	2	2	8	12

TEMA II: Importación de datos	Conferencia	2	2	8	12
TEMA III: Ordenar datos	Conferencia	2	2	8	12
TEMA IV: Manipulación de datos	Conferencia	2	2	8	12
TEMA V: Representación tabular y gráfica	Conferencia	2	2	8	12
Trabajo final				20	20
Total		10	10	48	68

Leyenda: **F.O.E. Forma de organización de la enseñanza**

Nota: todas las actividades serán impartidas por el profesor Maicel Monzón Pérez

TEMAS Y TOTAL DE HORAS A IMPARTIR POR PROFESOR

Profesores	Horas de docencia
Maicel Eugenio Monzón Pérez	24
Total	68

ESTRATEGIA DOCENTE

El curso se realizará durante una semana a tiempo completo de forma presencial en el Hotel Melia Habana desde el día martes 24 de marzo hasta el 27 de marzo de 2025. La actividad docente será estructurada por cuatro conferencias, además de una evaluación final. La tabla que se muestra a continuación describe los detalles.

Horario	Día 1 (martes)	Día 2 (miércoles)	Día 3 (Jueves)
Mañana	Introducción (15 min) Conf. Tema I (90 min)	Conf. Tema III (90 min)	Conf. Tema V (90 min)
Coffee breaks	-	-	-
Mañana	Video instalación (10 min) Prác.Tema I (90 min)	Conf. Tema IV (90 min)	Conf. Tema VI (90 min)
Almuerzo	-	-	-
Tarde	Video Rcmdr (10 min) Conf. Tema II (90 min)	Prác.Tema III y IV (90 min)	Prác. Tema V y VI (90 min)
Coffee breaks	-	-	-
Tarde	Prác.Tema II (90 min)	Prác. Tema III y IV (90 min)	Orientación EF

Legenda: **Conf:** conferencias, **Video:** Video demostrativo, **Prác:** Trabajo práctico
EF: Evaluación final

MEDIOS DE ENSEÑANZA

El fondo muendial de aliemntos propocionara a todos los asistentes a curso laptops, las intalaciones del aula cuntasn con una PC con proyector y acceso a internet para el profesor y los participantes.

SISTEMA DE EVALUACIÓN

- Evaluación formativa: Estará dada por ejercicios diarias a travez de una estrategia de gamificacion que se implementó para el curso.
- Evaluación final: El trabajo final es la resolución de un problema práctico que consiste en realizar un procesamiento estadístico a partir de una base de datos real sobre fortificacion de alieimentos. Esta deberá ser ordenada y transformada antes de realizar el análisis.

Al concluir el curso el estudiante recibirá un certificado si satisface las exigencias del programa. Las evaluaciones se expresarán con las calificaciones de Excelente (5), Bien (4), Aprobado (3) o Desaprobado (2).

BIBLIOGRAFÍA RECOMENDADA

1. Grolemond G, H Wickham, R para Ciencia de Datos. disponible en: <https://r4ds.had.co.nz/>, O' Reilly Media, 2017.
2. de Jonge E, An introduction to data cleaning with R. disponible en: https://cran.r-project.org/doc/contrib/de_Jonge+van_der_Loo-Introduction_to_data_cleaning_with_R.pdf

MINICURRICULO DEL CLAUSTRO

- Nombre(s) y apellidos: Dr. Maicel Eugenio Monzón Pérez
- Categoría docente: Profesor auxiliar
- Grado científico: MSc.
- Categoría investigativa: -
- Cargo y centro de trabajo: Responsable de Estadística, Centro Nacional Coordinador de Ensayos Clínicos
- Años de experiencia profesional: 17
- Especialidad: Dr en medicina, Especialista de primer grado en Bioestadística
- Maestría: En Informática Medica
- Número de postgrados recibidos: + 10
- Número de postgrados impartidos: +10
- Número de publicaciones: +10

Líneas investigativas en los últimos 5 años y títulos:

Ética en la investigación

Modelación de la propagación de epidemias

Ciencia de datos

Ensayos clínicos

Métodos estadísticos

Maicel Eugenio Monzón Pérez

Coordinador (a)