



FORMATO PARA PROPONER TÓPICOS	
1. Título del Tópico	
Introducción a R y bioestadística	
2. Tutor responsable	
Nombre completo	Osbaldo Resendis Antonio
Adscripción	CCG
Teléfono	777-3291899
Correo electrónico	resendis@ccg.unam.mx
3. Profesores invitados	
<i>Nombre completo</i>	Alejandra Medina Rivera
Adscripción	CCG
Teléfono	777-3132063
Correo electrónico	amedina@lcg.unam.mx
<i>Nombre completo</i>	Leonardo Collado Torres
Adscripción	IBt
Teléfono	777-3291605
Correo electrónico	lcollado@ibt.unam.mx
<i>Nombre completo</i>	Veronica Rohen
Adscripción	ITAM
Teléfono	
Correo electrónico	vrohen@lcg.unam.mx
4. Introducción/justificación del Tópico	
<p>La investigación en biología molecular está produciendo una amplia cantidad de datos gracias al desarrollo de nuevas tecnologías como la secuenciación masiva. El análisis y el procesamiento de datos requieren de la ayuda de análisis estadísticos capaces de manejarla. El uso de programas bioinformáticos especializados en estadística, se ha vuelto la herramienta indispensable para la interpretación, manejo y normalización de datos, por ello este curso pretende dar una introducción a lenguajes y paquetería especializada en analisis estadísticos. Actualmente el lenguaje más utilizado para el análisis de datos es R, ya que es un conjunto integrado de paquetes para manipulacion de datos, cálculo y gráficos. El objetivo de este tópico es dar a conocer las capacidades de R e introducir a los alumnos en el manejo y utilización de paqueterías especializadas y las bases estadísticas en las que se sustentan, de forma que puedan aplicar los conocimientos en el análisis de datos biológicos. Además, se pretende que los alumnos apliquen los conocimientos adquiridos en el análisis de sus datos. Se haran revisiones de artículos relevantes en el area de la bioestadística para genera un ambiente de discusion.</p>	

Debido a que será un curso teórico práctico de aplicaciones de programación y no de programación en sí, se pide que los alumnos manejen al menos un lenguaje de programación (perl, C, etc.).

5. Características para la impartición del Tópico * se requiere que los alumnos interesados manejen al menos un lenguaje de programación.

Indique el lugar en donde se realizará el Tópico	CCG
Número de sesiones y duración en horas por sesión (mínimo 36 horas)	30 sesiones de 2 horas cada una, miércoles y viernes.
Disponibilidad de impartirlo por videoconferencia	SI NO X
Número total de alumnos que puede aceptar	10
Número de alumnos del PDCB que puede aceptar	10

6. Método de evaluación

Por favor incluya en este apartado el % de la contribución relativa de:

Exámenes	10 %
Participación en clase	0
Tareas	30%
Presentación de un proyecto	60.00%
Otros	

7. Temario del tópico

Introducción. Unidad 1 (2 y 4 de Febrero AMR)

Como funciona R. La ayuda en línea. Emacs. Descripción e instalación. Operaciones básicas (comandos y sintaxis). Vectores. Objetos. Arreglos, matrices, listas. listado y remoción de objetos en memoria.

Unidad 2 (9 y 10 de Febrero AMR) Ejercicio para cubrir las bases del uso de R en consola. Modelos estadísticos. Gráficos. Leyendo datos desde un archivo. Guardando datos. Generación de datos. Manipulación de objetos. Conversión de objetos. Operadores. Cómo acceder los valores de un objeto: el sistema de indexación Accediendo a los valores de un objeto con nombres El editor de datos. Funciones aritméticas simples.

Cálculos con Matrices.

**Unidad 3 (16 y 18 de Febrero AMR)
Haciendo gráficas en R.**

Manejo de gráficos. Abriendo múltiples dispositivos gráficos. Disposición de una gráfica. Funciones gráficas. Comandos de graficación de bajo nivel. Parámetros gráficos. Los paquetes grid y lattice.

Programación práctica con R.

Bucles y Vectorización.
Escribiendo un programa en R.
Creando sus propias funciones.
Ejercicio/practica.

BIOESTADÍSTICA.

Unidad 4 y 5 (22 de Febrero VR 4 hrs).

Variables aleatorias. Discretas. Continuas. Medidas de tendencia central y dispersión de las variables aleatorias (media, mediana, valor esperado y varianza).

Unidad 6 (1 de Marzo VR). 4 de Marzo Practicas unidades 4,5 y 6 el AMR)

Población y muestreo.

Unidad 7 (8 Marzo VR, 11 de Marzo practicas LC)

Operaciones básicas con sucesos aleatorios (cálculo de probabilidades).

Unidad 8 y 9 (15 de Marzo VR./ 18 de marzo OR)

Distribución de variables aleatorias. Discretas (distribución de Bernoulli, binomial, geométrica, binomial geométrica, hipergeométrica y de Poisson). Continuas (distribución uniforme, exponencial, normal, χ^2 , t de Student, F de Snedecor). Momentos y comulantes.

Unidad 10 (22 de Marzo VR./ 25 de marzo OR)

Pruebas de significancia (paramétricas).
ANOVA. Corrección por multitesting.

Unidad 11 (29 de marzo VR) Pruebas no paramétricos. Aleatoriedad de una muestra (test de rachas). Normalidad de una muestra (test de D'Agostino). Contraste de Wilcoxon para muestras apareadas. Contraste de Kruskal-Wallis.

Unidad 12 (5 Abril Abril, VR) Contraste no parametrico.
(12 de Abril, VR) Contraste de poblaciones con datos categóricos.

Unidad 13 (27 y 29 de Abril, LC, 4 hrs) Regresiones (regresión lineal, bondad del ajuste).

Unidad 14 (4 de Mayo , AMR, 2 Hrs) Clustering, coeficiente de correlación lineal de Pearson.

Unidad 15 (6 de Mayo, OR) Probabilidad condicionada e independencia de sucesos. Teoremas del cálculo de probabilidades (teorema de la probabilidad compuesta, teorema de la probabilidad total, teorema de Bayes, Cadenas de Markov).

Unidad 16 (11 y 13 de Mayo, LC, 4 Hrs) Análisis estadístico de datos de secuenciación masiva.

Evaluación (18 y 20 de Mayo, exposición aplicación de R sobre los proyectos de estudiantes. Discusión de artículos. Propuestas de ejercicios o practicas con sus datos.

Tutoriales de R

<http://math.illinoisstate.edu/dhkim/rstuff/rtutor.html>

<http://cran.r-project.org/doc/manuals/R-intro.html>

<http://mercury.bio.uaf.edu/mercury/R/R.html>

8. Bibliografía

- 1.- Dalgaard Peter. Introductory statistics with R. ed. Springer 2002.
- 2.- Gentelman R., Huber W., Carey V.J., Irizarry R.A., Dudoit S., Bioinformatics and computational biology solutions using r and Bioconductor. ed. Springer 2005.
- 3.- Kohane I.S., Kho A.T., Butte A.J., Microarrays for an integrative genomics. Ed MIT press 2003.
- 4.- Rosner B., Fundamentals of Biostatistics (6a Ed), Iowa University State University Press 1980

Máximo 5 cuartillas

Las propuestas se deben entregar en formato electrónico en las oficinas de la entidad académica en donde el tutor está acreditado. El archivo se utilizará para incluir los Tópicos aprobados por Comité Académico en la página Web del PDCB.