

BIOINGENIERÍA

BIOINFORMÁTICA Y BIOESTADÍSTICA

04 CRÉDITOS



ÍNDICE

1. Asignatura	3
2. Datos generales	3
3. Profesores.....	3
3.1 Profesor coordinador del curso	3
3.2 Profesor(es) instructor(es) del curso.....	3
4. Introducción al curso	3
5. Objetivos	4
6. Competencias	5
7. Resultados de aprendizaje.....	5
8. Temas	6
9. Plan de trabajo.....	8
9.1 Metodología	8
9.2 Sesiones de teoría	8
9.3 Sesiones de práctica (laboratorio o taller)	9
10. Planificación	9
11. Sistema de evaluación	8
12. Referencia bibliográficas.....	8



UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

SILABO 2020-I

1. ASIGNATURA

BI0025 Bioinformática y Bioestadística

2. DATOS GENERALES

2.1 Créditos: cuatro (4) créditos

2.2 Horas de teoría: dos (2) semanales

2.3 Horas de práctica: cuatro (4) semanales

2.4 Duración del período: dieciséis (16) semanas

2.5 Condición: Obligatorio para Bioingeniería y Ciencia de la Computación

2.6 Modalidad: Presencial

2.7 Requisitos: Bioingeniería: Biología molecular y celular (BIO002)

* En general, se recomienda haber aprobado el curso de Estadística y Probabilidades (IN0054) y Biología Molecular y Celular (BIO002), y se desea haber aprobado el curso de Bioquímica (BIO004).

3. PROFESORES

3.1 Profesor coordinador del curso

Luis Jaramillo Valverde

ljaramillo@utec.edu.pe

Atención previa coordinación con el profesor.

3.2 Profesor(es) instructor(es) del curso

Luis Jaramillo Valverde

ljaramillo@utec.edu.pe

Atención previa coordinación con el profesor.

Antonio Quispe Gutiérrez

aquispe@utec.edu.pe

Atención previa coordinación con el profesor.

4. INTRODUCCIÓN AL CURSO

La bioinformática y la bioestadística se han convertido en herramientas esenciales para el desarrollo de los profesionales en el campo de la bioingeniería. La aplicación de estas herramientas permite por ejemplo a estos profesionales analizar la estructura de las proteínas y secuenciar genomas. Dado que muchos de estos problemas son complejos y requieren el manejo de grandes bases de datos en el módulo de Bioinformática los estudiantes aprenderán a usar una variedad de métodos y herramientas de software que les permitan realizar este tipo de datos y comprender de las estructuras biológicas.



De manera complementaria, en el módulo de bioestadística los estudiantes aprenderán las pruebas estadísticas y estrategias de análisis que les permitan probar hipótesis y responder a preguntas científicas específicas. Para los estudiantes de bioingeniería, este curso representa una oportunidad para adquirir y cultivar una serie de habilidades que más adelante les permitirán profundizar y desarrollarse en el campo de la bioinformática y bioestadística, sacándole el máximo provecho a la informática y a las tecnologías de la información. Para los estudiantes de informática, este curso representa una oportunidad invaluable para incrementar y poner en práctica sus conocimientos de bioinformática y bioestadística, lenguajes de programación, estadística, modelamiento de datos, gestión de la información y sistemas inteligentes.

5. OBJETIVOS

- Sesión 1 y 2: Conocer el desarrollo histórico de la Bioinformática, revisar el dogma de la vida: replicación, transcripción, traducción, síntesis de proteína y estudiar el genoma mediante técnicas específicas.
- Sesión 3 y 4: Conocer las bases de datos biológicas principales, revisar secuencias de nucleótidos y secuencias de aminoácidos mediante alineamiento y comparación de secuencias múltiples.
- Sesión 5 y 6: Describir la Filogenia, diseñar arboles filogenéticos y métodos de reconstrucción filogeográficos.
- Sesión 7 y 8: Entender las tecnologías de secuenciación de próxima generación, los algoritmos para el ensamblaje de secuencias y heurística de ensamblaje.
- Sesión 9: Conocer las estructuras moleculares y plegamiento de proteínas; así como, el enhebrado de proteínas y su alineamiento estructural
- Sesión 10: Estudiar modelos de probabilidad y Markov ocultos; así como, filogenia probabilística.
- Sesión 11: Entender los tipos de diseño de estudios y calcular el tamaño de muestra.
- Sesión 12: Realizar el análisis descriptivo y gráfico de los datos; así como, su control de calidad.



- Sesión 13 y 14: Conocer el análisis invariado y uso de pruebas estadísticas paramétricas y no paramétricas; así como el análisis multivariado y uso de los modelos generales lineales de regresión para el control de confesores y modificadores de efecto.
- Sesión 15 y 16: Conocer el análisis de regresión lineal y logística, regresión binomial y de Polisón; así como, análisis de sobrevida y regresión Cox.

6. COMPETENCIAS

- a1: Capacidad para aplicar el conocimiento de las matemáticas (nivel 2)
El estudiante resuelve una hoja de problemas matemáticos donde describe, selecciona y aplica lo aprendido de la teoría de bioinformática y bioestadística.
- a2: Capacidad para aplicar el conocimiento de la ciencia (nivel 2)
El estudiante resuelve una hoja de problemas donde describe, selecciona y aplica lo aprendido de la teoría de bioinformática y bioestadística.
- b2: Capacidad para analizar información (nivel 2)
A partir de la lectura de un artículo científico acerca de estudios de bioinformática y bioestadística, el estudiante discute las ideas principales del mismo y propone nuevas alternativas con investigación adicional.
- g3: Capacidad de comunicarse gráficamente (nivel 2)
El estudiante preparará una presentación y expondrá frente al auditorio acerca de un proyecto de investigación basado en herramientas bioinformática y bioestadística.
- k1: Capacidad para utilizar las técnicas, habilidades y herramientas de la ingeniería moderna necesarias para la práctica de la ingeniería (nivel 1)
El estudiante utilizará softwares bioinformáticos y bioestadísticos, usados en investigaciones de alta importancia científica, para su entrenamiento.

7. RESULTADOS DE APRENDIZAJE

Al final del curso el estudiante del curso Bioinformática y Bioestadística se espera que el estudiante sea capaz de:



RA1. Que el estudiante sea capaz de navegar por las principales bases de datos del NCBI (TAXONOMY, NUCLEOTIDE, POPSET, PUBMED, GENE, OMIM) y usar las herramientas que ofrece la plataforma NCBI (BLAST, GENOME VIEWER), NSEMBLE (SINTENIA).

RA 2. Que el estudiante sea capaz de comprender el proceso algorítmico que ejecuta un software para dar solución a un problema biológico valorando sus ventajas y desventajas.

RA 3. Que el estudiante sea capaz de crear y manejar bases de datos, realizar el análisis descriptivo de datos, plantear y probar hipótesis estadísticas, y calcular tamaños de muestra en respuesta a hipótesis específicas.

RA 4. Que el estudiante sea capaz de realizar análisis bivariados mediante el uso de pruebas estadísticas paramétricas y no paramétricas, usar análisis multivariados para el control de confusores y modificadores de efecto, y usar de modelos lineales generales de regresión, incluyendo los análisis de regresión lineal, logística, Poisson, Binomial y de Cox.

8. TEMAS

1. Introducción a la bioinformática

1.1. Desarrollo Histórico de la Bioinformática. Conceptos Generales. Repositorios de Información Molecular.

1.2. El dogma de la vida: replicación, transcripción, traducción, síntesis de proteínas.

1.3. Estudio del genoma: mapas y secuencias, técnicas específicas.

2. Comparación de secuencias

2.1. 2.1. Bases de datos biológicas PubMed, NCBI (genbank), EMBL, Swissprot, Protein Data Bank, Drug Data Bank, Software DNAtwin, Swisspdbviewer.

2.2. Secuencias de nucleótidos y secuencias de aminoácidos.

2.3. Alineación de secuencia, problema de alineación emparejado, búsqueda exhaustiva, programación dinámica, alineación global, alineación local, penalización de huecos.

2.4. Comparación de secuencias múltiples: suma de pares, análisis de complejidad por programación dinámica, heurística de alineación, algoritmo en estrella, algoritmos de alineación progresiva.

3. Árboles filogenéticos

3.1. Filogenia: Introducción y relaciones filogenéticas.

3.2. Árboles filogenéticos: definición, tipo de árboles, problema de búsqueda y reconstrucción de árboles.



3.3. Métodos de reconstrucción: métodos de parsimonia, métodos de distancia, métodos de máxima verosimilitud, modelos evolutivos, inferencia bayesiana.

4. Ensamblaje de secuencia de ADN

4.1. Base biológica: caso ideal, dificultades, métodos alternativos para la secuenciación del ADN.

4.2. Tecnologías de secuenciación de próxima generación.

4.4. Algoritmos para el ensamblaje de secuencias: representación de superposiciones, rutas para crear supercadenas, algoritmo voraz, gráficos acíclicos.

4.5. Heurística de ensamblaje: búsqueda de superposiciones, ordenación de fragmentos, alineaciones y consenso.

5. Estructuras secundarias y terciarias.

5.1. Estructuras moleculares: primaria, secundaria, terciaria, cuaternaria.

5.2. Predicción de estructuras secundarias de ARN: modelo formal, energía de pares, estructuras con bases independientes, solución con programación dinámica, estructuras con bucles.

5.3. Plegamiento de proteínas.

5.4. Enhebrado de proteínas: definiciones, algoritmo de ramificación, ramificación para el enhebrado de proteínas.

5.5. Alineamiento estructural: definiciones, algoritmo DALI.

6. Una introducción de modelos probabilísticos en biología molecular

6.1. Probabilidad: Variables aleatorias, cadenas de Markov, algoritmo de aceleración metropolitana, campos aleatorios de Markov y muestra de Gibbs, probabilidad máxima.

6.2. Modelos de Markov ocultos (HMM), estimación de parámetros, algoritmo de Viterbi y método de Baul-Welch, aplicación en alineamientos emparejados y múltiples, detección de motivos en proteínas, en ADN eucariota, en familias de secuencias.

6.3. Filogenia probabilística: modelos probabilísticos de evolución, probabilidad de alineamientos, probabilidad de inferencia, comparación de métodos probabilísticos y no probabilísticos.

7. Diseño de estudios y cálculo de tamaño de muestra

7.1. Diseño de estudios y magnitudes de asociación.

7.2. Introducción al análisis estadístico de datos con R

7.3. Cálculo de tamaño de muestra.



8. Análisis descriptivo de datos

8.1. Manejo y control de la calidad de los datos.

8.2. Análisis descriptivo de datos.

8.3. Análisis gráfico de datos.

9. Análisis bivariados y multivariados

9.1. Análisis de datos de estudios observacionales (transversales, casos y controles, y de cohortes) y experimentales (ensayos clínicos controlados aleatorizados y no aleatorizados).

9.2. Análisis bivariados y uso de pruebas estadísticas paramétricas y no paramétricas.

9.3. Análisis multivariados y uso de los modelos generales lineales de regresión para el control de confusores y modificadores de efecto.

10. Análisis de regresión

10.2. Análisis de regresión lineal y logística.

10.3. Análisis de regresión binomial y de Poisson.

10.4. Análisis de sobrevida y regresión de Cox.

9. PLAN DE TRABAJO

9.1 Metodología

Las sesiones serán desarrolladas bajo el aprendizaje basado en problemas, lo que significa que el estudiante adquiere conocimiento mediante la formulación de preguntas de manera cíclica. El uso de esta metodología permitirá el desarrollo del pensamiento crítico, mejora en la resolución de problemas, mayor motivación del alumno y mejora en la capacidad de transferir conocimiento.

9.2 Sesiones de teoría

El curso se ofrece a estudiantes de ingeniería e informática. El coordinador del curso (o maestro invitado) conduce las sesiones de conceptos para todos los estudiantes en el aula y es responsable de la gestión del contenido. Como curso interdisciplinario, las sesiones de conceptos también son apoyadas por otros profesores de carreras de ingeniería de UTEC e invitados. La evaluación será continua durante todas las sesiones teóricas.



9.3 Sesiones de práctica (laboratorio o taller)

Las sesiones de laboratorio adoptan una estrategia de aprendizaje activo y / o talleres que sirven para reforzar los conceptos vistos en las sesiones de teoría. El software especializado para bioinformática y el software R para bioestadística se proporcionan en el laboratorio. La evaluación será continua durante todas las sesiones de práctica.

10. PLANIFICACIÓN

- Ver el Anexo 1

11. SISTEMA DE EVALUACIÓN

EVALUACIÓN *La ponderación de la evaluación se hará si ambas partes están aprobadas	TEORÍA	PRÁCTICA Y/O LABORATORIO
	<u>Módulo Bioinformática</u> 1 Examen Teórico E1 (30%) 1 Practica calificada PC (5%)	<u>Módulo Bioinformática</u> 1 Examen Práctico E2 (20%) 1 Proyecto P1 (5%)
	<u>Módulo Bioestadística</u> 1 Examen Teórico E3 (20%) 1 Practica calificada PC (5%)	<u>Módulo Bioestadística</u> 1 Examen Práctico E4 (10%) 1 Proyecto P2 (5%)
	60%	40%
	100%	

El redondeo se realizará al entero más cercano. 0.5 se redondea

12. REFERENCIA BIBLIOGRÁFICAS

a) LIBROS

Módulo Bioinformática

- Jonathan Pevsner. Bioinformatics and Functional Genomics. 3rd Edition. 2015. Wiley Blackwell. ISBN-13: 978-1118581780



- Arthur Lesk. Introduction to Bioinformatics. 4th Edition. 2014. Oxford University Press. ISBN-13: 978-0199651566
- Paul M. Selzer, Richard J. Marhöfer, Oliver Koch. Applied Bioinformatics: An Introduction. 2nd Edition. 2018 Springer. ISBN-13: 978-3319682990
- Phillip Compeau & Pavel Pevzner. Bioinformatics Algorithms: An Active Learning Approach. 3rd Edition. 2018 Active Learning Publisher. ISBN-13: 978-0990374633
- Claverie, J. M., & Notredame, C. (2011). Bioinformatics for dummies. John Wiley & Sons

Módulo Bioestadística

- Daniel WW; Cross CL. Biostatistics- A foundation for analysis in the health sciences. 10th ed; New Jersey; Willey; 2013.
- Baldi B; Moore DS. The Practice of Statistics in the Life Sciences. 3rd ed; New York; W.H. Freeman; 2014.
- Sprent P; Smeeton NC. Applied non-Parametric Statistical Methods. 4th ed; Florida; Chapman & Hall/CRC; 2007.

b) ARTICULOS

Módulo Bioinformática

- Fenstermacher, D. (2005). Introduction to bioinformatics. Journal of the American Society for Information Science and Technology, 56(5), 440-446.
- Ghiurcuta, C. G., & Moret, B. M. (2014). Evaluating synteny for improved comparative studies. Bioinformatics, 30(12), i9-i18.
- Gogarten, J. P., & Olendzenski, L. (1999). Orthologs, paralogs and genome comparisons. Current opinion in genetics & development, 9(6), 630-636.
- Ji, S. (1999). The cell as the smallest DNA-based molecular computer. BioSystems, 52(1), 123-133.
- Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics 1. Annu. Rev. Genet., 39, 309-338.
- Passarge, E., Horsthemke, B., & Farber, R. A. (1999). Incorrect use of the term synteny. Nat Genet, 23(4), 387.

Módulo Bioestadística

- Rosner B. Fundamentals of Biostatistics, 8th Edition, 2015, Duxbury Press, Pacific Grove, CA



- Cleophas TJ. Clinical trials and p-values beware of the extremes. Clin Chem Lab Med 2004; 42(3): 300-4.
- Collett D. Modelling Survival Data in Medical Research, 3rd Edition, 2014, Chapman & Hall/CRC, NW
- Dobson AJ, Barnett AG. An Introduction to Generalized Linear Models, 4th Edition, 2018, Chapman & Hall/CRC, NW
- Dupont W D. Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data, 2009, Cambridge University Press, Cambridge, U.K
- Everitt BS, Torsten H. A Handbook of Statistical Analyses Using R, 2nd Edition, 2010, Chapman & Hall/CRC, NW
- Hiebert SM. Teaching simple experimental design to undergraduates: do your students understand the basics? Adv Physiol Educ 2007; 31(1): 82-92.
- Jeng M. Error in statistical tests of error in statistical tests. BMC Med Res Methodol 2006; 6: 45.
- Nyirongo VB, Mukaka MM ; Kalilani-Phiri LV). Statistical pitfalls in medical research. Malawi Med J 2008; 20(1): 15-8.
- Oliver D; Mahon SM. Reading a research article part II: parametric and nonparametric statistics. Clin J Oncol Nurs 2005; 9(2): 238-40.
- Overholser BR; Sowinski KM. Biostatistics primer: part I. Nutr Clin Pract 2007; 22(6): 629-35.
- Overholser BR; Sowinski KM. Biostatistics primer: part 2. Nutr Clin Pract 2008; 23(1): 76-84.
- Whitley E; Ball J. Statistics review 3: hypothesis testing and P values. Crit Care 2002; 6(3): 222-5.

