

**PROGRAMA DE CURSO**  
**PENSAMIENTO ESTADÍSTICO**

**A. Antecedentes generales del curso:**

Departamento	<b>CIENCIAS DE LA COMPUTACION</b>					
Nombre del curso	Pensamiento Estadístico					
Nombre del curso en inglés	<i>Statistical Thinking</i>					
Código	CC6104		Créditos	6		
Horas semanales	Docencia	4,0	Auxiliares	0	Trabajo personal	7,0
Carácter del curso	Obligatorio			Electivo	<b>X</b>	
Requisitos	CC5206 Introducción a la Minería de Datos o MA3403 Probabilidades y Estadística o MA3401 Probabilidades.					

**B. Propósito del curso:**

El propósito de este curso es presentar a los y las estudiantes habilidades asociadas al pensamiento estadístico, esto es, una forma de pensamiento sistemático para describir fenómenos del “mundo real” tomando en cuenta su incertidumbre inherente y usando datos para la toma de decisiones.

El curso introduce los fundamentos del pensamiento estadístico, que son el análisis descriptivo de datos, la programación estadística en R y la teoría de probabilidades. Luego se enseñan las dos escuelas más importantes de la inferencia estadística: la inferencia frecuentista y la inferencia Bayesiana. Finalmente, se discuten temas avanzados como la evaluación de modelos usando teoría de información, modelos gráficos dirigidos para modelar dependencias condicionales entre variables y modelos multi-nivel.

El contexto del desarrollo de habilidades de aplicación metodológica será a través del planteamiento de métodos estadísticos y aplicaciones reales, con soluciones bien delimitadas respecto a su alcance y tamaño.

Desde el punto de vista teórico, el curso busca que los estudiantes sean capaces de leer literatura en inglés en el área. Desde el punto de vista práctico, se apunta a que los estudiantes sean capaces de implementar modelos estadísticos a problemas concretos utilizando entornos computacionales.

En resumen, se espera que los estudiantes desarrollen una metodología de trabajo que los lleve resolver problemas del mundo real usando datos y razonamiento estadístico.

Las competencias específicas (CE) y genéricas (CG) a las que tributa el curso son:

**CE1:** Analizar problemas computacionales, construir modelos, expresándolos en representaciones y lenguajes formales adecuados.

**CE4:** Extraer información relevante, utilizando el proceso de descubrimiento de conocimiento de datos.

**CG1:** Comunicación académica y profesional

Comunicar en español de forma estratégica, clara y eficaz, tanto en modalidad oral como escrita, puntos de vista, propuestas de proyectos y resultados de investigación fundamentados, en situaciones de comunicación compleja, en ambientes sociales, académicos y profesionales.

**CG2:** Comunicación en inglés

Leer y escuchar de manera comprensiva en inglés variados tipos de textos e informaciones sobre temas concretos o abstractos, comunicando experiencias y opiniones, adecuándose a diferentes contextos de acuerdo a las características de la audiencia.

**CG3:** Compromiso ético

Actuar de manera responsable y honesta, dando cuenta en forma crítica de sus propias acciones y sus consecuencias, en el marco del respeto hacia la dignidad de las personas y el cuidado del medio social, cultural y natural.

**CG4:** Trabajo en equipo

Ejecutar con su equipo, de forma estratégica, diversas actividades formativas propuestas, considerando la autogestión de sí mismo y la relación con el otro, asumiendo diversos roles: de líder, colaborador u otros, según requerimientos y objetivos, sin discriminar por género u otra razón.

### C. Resultados de aprendizaje:

Competencias específicas	Resultados de aprendizaje
CE1	RA1: Descompone, de forma irreductible, un modelo estadístico, reconociendo analíticamente, sus componentes principales y deduciendo sus datos de entrada y salida.
	RA2: Plantea y explica un modelo de solución, delimitado respecto a su alcance y tamaño, para problemas del mundo real, mediante la identificación de componentes principales con el fin de procesar la entrada y generar la salida deseada, usando técnicas específicas de estadística.
	RA3: Diseña e interpreta el significado de un experimento científico, según el área de aplicación correspondiente, con el fin de diseñar y presentar adecuadamente los resultados experimentales de una investigación empírica.
CE4	RA4: Implementa y ejecuta un programa computacional, según la tarea de inferencia estadística a resolver y el modelo propuesto, con el fin de obtener una solución ejecutable de dicho modelo.
	RA5: Evalúa la solución implementada, delimitada según alcance y tamaño, usando métricas de evaluación con el fin de validar y/o rectificar el modelo propuesto para el problema en cuestión.
Competencias genéricas	Resultados de aprendizaje
CG1, CG4	RA6: Presenta de manera escrita propuestas de solución a problemas en estadística, a fin de explicar, de manera sintética y precisa, las soluciones propuestas por el equipo y su pertinencia.
CG2	RA7: Lee en inglés, de manera analítica y comprensiva, artículos científicos en estadística, a fin de relacionar dicha información y generar nuevos conocimientos aplicables a temas del pensamiento estadístico.
CG3, CG4	RA8: Realiza, con su equipo, las actividades comprometidas, de manera responsable y honesta en los plazos comprometidos, citando fuentes y referencias de donde se extrae la información, a fin de elaborar propuestas propias sin incurrir en plagio.

**D. Unidades temáticas:**

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
1	RA1-RA4-RA5-RA6-RA8	<b>Fundamentos</b>	3 semanas
Contenidos		Indicador de logro	
<p>1.1. Introducción al pensamiento estadístico.</p> <p>a. ¿Qué se puede hacer con la estadística?</p> <p>b. Las grandes ideas de la estadística</p> <p>i. Aprender de datos</p> <p>ii. Agregación</p> <p>iii. Incertidumbre</p> <p>iv. Muestreo</p> <p>v. Causalidad</p> <p>1.2. Programación estadística en R</p> <p>a. Variables</p> <p>b. Funciones</p> <p>c. Vectores</p> <p>d. Listas</p> <p>e. Data.frames</p> <p>1.3 Estadística descriptiva</p> <p>a. Estadísticas de resumen</p> <p>i. Medidas de tendencia central</p> <p>ii. Medidas de dispersión</p> <p>iii. Percentiles</p> <p>iv. Correlaciones</p> <p>b. Visualización</p> <p>i. Scatterplots</p> <p>ii. Boxplots</p> <p>iii. Histogramas</p> <p>1.4 Probabilidades</p> <p>a. Experimento</p> <p>b. Espacio muestral</p> <p>c. Evento</p> <p>d. Variable aleatoria</p>		<p>El/la estudiante:</p> <ol style="list-style-type: none"> <li>Explica los fundamentos del pensamiento estadístico, considerando sus componentes principales.</li> <li>Manipula y visualiza conjuntos de datos usando el entorno de programación R.</li> <li>Reconoce de forma analítica las estadísticas de resumen y las técnicas básicas de visualización de datos.</li> <li>Describe y explica los fundamentos de la teoría de probabilidad.</li> <li>Lee de manera comprensiva diversos textos y artículos en inglés sobre estadística, determinando ideas principales aplicables a ejemplos que se le presentan.</li> <li>Programa, con su equipo, en R para resolver problemas probabilísticos, respetando obligaciones y acuerdos adquiridos como compromisos para el logro de la actividad.</li> <li>Planifica y presenta sus trabajos, basándose en sus capacidades, sin incurrir en plagio o copia respecto del análisis de datos.</li> <li>Produce un texto sobre donde reporta resultados de un problema de análisis descriptivo de datos, considerando a las diferentes audiencias posibles (especializada, no especializada, pares, académicos, entre otros) para una adecuada recepción del mensaje.</li> </ol>	

<ul style="list-style-type: none"> <li>e. Función de masa de probabilidad.</li> <li>f. Función de densidad de probabilidad.</li> <li>g. La distribución Normal</li> <li>h. Probabilidad conjunta</li> <li>i. Probabilidad condicional</li> <li>j. Teorema de Bayes</li> <li>k. Probabilidades totales</li> <li>l. Esperanza</li> <li>m. Varianza</li> <li>n. Ley de los grandes números</li> <li>o. Teorema central del límite.</li> </ul>	
<p>Bibliografía de la unidad</p>	<p>[1] Capítulo 1-5. [2] Capítulo 1. [3] Capítulos 1-7.</p>

Número	RA al que tributa	Nombre de la unidad	Duración en Semanas
2	RA1-RA2-RA3-RA4-RA5-RA6-RA8	Inferencia Frecuentista	5 semanas
Contenidos		Indicador de logro	
<ul style="list-style-type: none"> <li>2.1. Introducción a la inferencia estadística</li> <li>2.2. Estimación puntual</li> <li>2.3. Distribución muestral</li> <li>2.4. Error estándar</li> <li>2.5. Estimador de máxima verosimilitud.</li> <li>2.6. Intervalo de confianza</li> <li>2.7. Test de hipótesis               <ul style="list-style-type: none"> <li>a. Test de Wald</li> <li>b. Test de T</li> <li>c. P-valor</li> </ul> </li> <li>2.8. Diseño de experimentos</li> <li>2.9. Regresión lineal               <ul style="list-style-type: none"> <li>a. Mínimos cuadrados</li> <li>b. Regresión lineal simple</li> <li>c. Regresión lineal múltiple</li> </ul> </li> </ul>		<p>El/la estudiante:</p> <ul style="list-style-type: none"> <li>1. Reconoce de forma analítica, ejemplos de problemas inferencia estadística.</li> <li>2. Identifica aplicaciones de la inferencia estadística a problemas del mundo real y utiliza métodos pertinentes de inferencia frecuentista para resolverlos.</li> <li>3. Describe los fundamentos de la inferencia frecuentista y la distribución muestral de un estimador.</li> <li>4. Explica el funcionamiento de técnicas de estimación puntual.</li> <li>5. Explica el mecanismo para calcular un intervalo de confianza.</li> <li>6. Aplica test de hipótesis en la resolución de problemas de inferencia verificando que se cumplan todos los supuestos correspondientes.</li> <li>7. Explica y utiliza el modelo de regresión lineal para modelos predictivos de salida continua.</li> <li>8. Implementa métodos de inferencia frecuentista usando el entorno de trabajo R.</li> </ul>	

	<p>9. Respeta los acuerdos tomados con el equipo para la resolución de problemas de inferencia y las tarea que esta actividad conlleva.</p> <p>10. Trabaja en la resolución de problemas de inferencia frecuentista con su equipo de trabajo, planificando la actividad para cumplir a tiempo con los compromisos basándose en sus capacidades por el bien del equipo.</p> <p>11. Relaciona, jerarquiza e integra en sus escritos información proveniente de múltiples fuentes donde reporta resultados de problemas sobre inferencia frecuentista, usando reglas de citado correspondiente.</p>
Bibliografía de la unidad	<p>[1] Capítulo 6, 9, 10 y 13.</p> <p>[3] Capítulos 9, 10, 14 y 15.</p>

Número	RA al que tributa	Nombre de la unidad	Duración en Semanas
3	RA1-RA2-RA3-RA4-RA5-RA6-RA7-RA8	Inferencia Bayesiana	5 semanas
Contenidos		Indicador de logro	
3.1. Introducción a la Inferencia Bayesiana.	<p>a. Distribución prior</p> <p>b. Distribución posterior</p> <p>c. Verosimilitud promedio</p> <p>d. Priors conjugados</p> <p>e. Aproximación cuadrática del posterior.</p> <p>f. MAP y su relación con el estimador de máxima verosimilitud.</p>	<p>E/lal estudiante:</p> <ol style="list-style-type: none"> <li>1. Reconoce y analiza problemas de inferencia donde es factible de utilizar el enfoque Bayesiano.</li> <li>2. Explica la lógica de la inferencia Bayesiana y sus componentes principales: la distribución prior, la función verosimilitud, el posterior y la verosimilitud promedio.</li> <li>3. Reconoce y utiliza herramientas disponibles para extraer información a partir de una distribución posterior en problemas de inferencia Bayesiana.</li> <li>4. Extrae y aproxima una distribución posterior, usando el algoritmo MCMC.</li> </ol>	
3.2. Usando el Posterior	<p>a. Tomar muestras de un posterior.</p> <p>b. Intervalos de credibilidad</p> <p>c. Intervalo de densidad posterior mayor.</p> <p>d. Distribución posterior predictiva.</p>	<ol style="list-style-type: none"> <li>5. Implementa modelos Bayesianos para la resolución de problemas de inferencia estadística, usando el lenguaje de programación R.</li> <li>6. Organiza con el equipo el quehacer para la resolución de problemas de inferencia bayesianos, cumpliendo las obligaciones y acuerdos de cada actividad.</li> <li>7. Plantea a su equipo, de manera clara, precisa y constructiva, su posición acerca de cómo resolver tareas sobre modelos bayesianos para cumplir de forma efectiva la tarea emprendida.</li> </ol>	
3.3. Modelos lineales Bayesianos			

<p>3.4. MCMC: a. Algoritmo MCMC para extraer muestras de una distribución posterior.</p>	<p>8. Relaciona, jerarquiza e integra en sus escritos información proveniente de múltiples fuentes, según normas de citado correspondientes. 9. Plantea soluciones e ideas sobre inferencia Bayesiana, a través de una argumentación oral consistente que evidencia claridad, precisión y coherencia.</p>
<p>Bibliografía de la unidad</p>	<p>[1] Capítulo 11 [2] Capítulos 2-5 y 9. [3] Capítulo 11.</p>

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
4	RA2-RA4-RA5-RA6-RA7-RA8	Tópicos Adicionales en Inferencia Estadística	2 semanas
Contenidos		Indicador de logro	
<p>4.1. Teoría de Información para la evaluación de modelos. a. Entropía b. Divergencia c. Criterio de información Akaike d. Deviance e. Criterio de información Bayesiano.</p> <p>4.2. Modelos gráficos probabilísticos a. Independencia condicional b. Grafos dirigidos acíclicos c. D-separación</p> <p>4.3 Modelos Multi-nivel a. Introducción a los modelos multi-nivel.</p>		<p>El/la estudiante:</p> <ol style="list-style-type: none"> <li>Explica los elementos fundamentales de la teoría de información para la evaluación de modelos estadísticos.</li> <li>Modela las dependencias de un modelo multivariable, considerando la descripción de e un grafo dirigido acíclico</li> <li>Analiza el funcionamiento de un grafo dirigido acíclico para la modelación de independencias condicionales.</li> <li>Explica el funcionamiento de modelos multi-nivel interpretando su naturaleza jerárquica para la resolución de problemas complejos en inferencia estadística.</li> <li>Determina las ventajas de usar modelos multi-nivel por sobre hacer agregación.</li> <li>Implementa modelos multi-nivel Bayesianos usando el ambiente de trabajo R.</li> <li>Cumple obligaciones y acuerdos, respetando los compromisos adquiridos para trabajar con eficiencia y eficacia en la resolución de problemas de inferencia estadística.</li> <li>Plantea a su equipo, de manera clara, precisa y constructiva, su posición acerca de cómo resolver problemas en tópicos avanzados de inferencia estadística, para cumplir de forma efectiva la tarea emprendida en problemas de inferencia estadística.</li> </ol>	

	9. Plantea soluciones e ideas sobre modelos multi-nivel, a través de una argumentación oral consistente que evidencia claridad, precisión y coherencia.
Bibliografía de la unidad	[1] Capítulo 17. [2] Capítulos 6, 7 y 13.

#### E. Estrategias de enseñanzas:

La metodología de enseñanza y aprendizaje fomenta la participación del estudiante en el aula, las clases son principalmente:

- Clase expositiva, en donde el estudiante identifica los problemas fundamentales de la estadística, así como modelos y técnicas para abordarlos.
- Resolución de problemas. En cada unidad el estudiante es expuesto a librerías de programación que permiten implementar modelos de solución a problemas en estadística.

A lo anterior se le suman tareas individuales que deben ser desarrolladas con el computador y enviadas a través de U-cursos.

#### F. Estrategias de evaluación:

El curso tiene las siguientes instancias de evaluación:

Práctica:

- 5 tareas individuales que involucran preguntas teóricas, resolución de problemas aplicados y lectura de artículos en inglés.

Teórica:

- 1 Examen teórico.

*La nota final se calcula así: 70% para la nota tareas y 30% para la nota del examen. Para aprobar el curso se debe tener una nota final igual o superior a 4.0.*



### G. Recursos bibliográficos:

#### Bibliografía obligatoria:

- (1) Wasserman, L. (2013). All of statistics: a concise course in statistical inference. Springer Science & Business Media.
- (2) McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press.
- (3) Poldrack, R. (2020). Statistical Thinking for the 21st Century An open source textbook for statistics, with companions for R and Python. Disponible en: <https://statstheinking21.org/>

#### Bibliografía Complementaria:

- (4) Matthew J. C. Crump Answering questions with data: <https://crumplab.github.io/statistics/>
- (5) Mine Çetinkaya-Rundel Introduction to Modern Statistics: <https://openintro-ims.netlify.app/index.html>

### H. Datos generales sobre elaboración y vigencia del programa de curso:

Vigencia desde:	Primavera 2021
Elaborado por:	Felipe Bravo
Validado por:	Académico Par: Sergio Ochoa – Jefe Docente del DCC Validación CTD Computación.
Revisado por:	Área de Gestión Curricular, SGD