

TRABAJOS DE FIN DE GRADO OFERTADOS -GRADO EN ESTADÍSTICA- CURSO 2018-19

	<u>TÍTULO</u>	<u>TUTOR</u>	<u>RESUMEN</u>
1.	LA INTEGRAL DE LEBESGUE Y APLICACIONES EN ESTADÍSTICA	ÁNGEL TOCINO GARCÍA	La integración en espacios de medida abstractos permite aunar en una sola teoría diferentes conceptos y resultados estudiados por los estudiantes del Grado en Estadística, como son la integral de Riemann en intervalos cerrados, las integrales impropias, esperanzas de variables aleatorias discretas y continuas, etc. Se hará una introducción general a los espacios de medida (como ejemplo, se detallará la medida de Lebesgue en \mathbb{R}) y a las funciones medibles. Se estudiará en concepto de integral respecto a una medida, así como los teoremas fundamentales (Fatou, convergencia monótona y dominada). Finalmente se ilustrarán estos resultados teóricos aplicándolos a los casos de variables aleatorias bien conocidas de cursos anteriores.
2.	INFORMES ESTADÍSTICOS CON R MARKDOWN: APLICACIÓN EN INFORMES DE BÚSQUEDA DE GENES DIFERENCIALMENTE EXPRESADOS EN PACIENTES CON ALZHEIMER	JOSÉ M. SÁNCHEZ SANTOS	A medida que el software estadístico R (https://www.r-project.org/) ha incrementado su utilización en todo tipo de campos, se ha hecho patente la necesidad de generar informes estadísticos claros, dinámicos y elegantes en todo tipo de formato informático (word, pdf, html, etc.) ya que hasta hace unos años dichos informes se generaban en texto plano sin formato, obtenido directamente de la consola del software R. El principal objetivo del trabajo es el de recopilar, exponer y detallar las funciones más importantes del paquete “rmarkdown” de R (https://rmarkdown.rstudio.com/), que es una herramienta para la creación de ficheros *.Rmd que, mediante el lenguaje “markdown”, facilitan la aplicación de un formato claro y elegante a los informes estadísticos que se crean con R. Como aplicación se utilizará dicho paquete para la elaboración de informes sobre la búsqueda y el análisis de genes que presenten diferencias significativas en sus señales de intensidad génica en microarrays de individuos sanos frente a microarrays de pacientes con Alzheimer. Para ello se utilizará la herramienta “SAM” (significance analysis of microarray) como generalización del contraste clásico de la t de Student. Se usarán datos de microarrays extraídos de bases de datos publicadas y herramientas de Bioconductor y R-project. Para el desarrollo del proyecto se utilizará el software estadístico R y la interfaz RStudio.
3.	EL MÉTODO DE LA RAZÓN DE VEROSIMILITUDES EN CONTRASTES DE HIPÓTESIS	RAMÓN Á. ARDANUY ALBAJAR	Se trata de hacer un estudio del Método de la Razón de Verosimilitudes para construir Contrastes de Hipótesis, su definición, la relación con los contrastes de hipótesis de máxima potencia, su relación con las pruebas Bayesianas, determinar la distribución asintótica, y ver sus aplicaciones en Estadística Clásica, tanto para los casos de medias y varianzas de distribuciones normales, como para pruebas de adherencia de ajuste, tablas de contingencia y pruebas de homogeneidad. En el estudio se tendrá en cuenta el teorema de Neyman y Pearson para contrastes de máxima potencia de hipótesis simples así como la extensión al caso de hipótesis compuestas.
4.	TÉCNICAS ESTADÍSTICAS EN MODELOS DE PREDICCIÓN EN MACHINE LEARNING.	DESARA RODRÍGUEZ, INÉS SITTÓN M ^a TERESA SANTOS	El Machine Learning es una disciplina dentro de la Inteligencia Artificial que crea sistemas de aprendizaje automático. El aprendizaje se realiza a través de algoritmos que analizan y revisan los datos para predecir comportamientos futuros. En este trabajo se pretende analizar las técnicas estadísticas más frecuentes en este campo y calcular los principales parámetros de evaluación en modelos de predicción aplicados a conjuntos de datos provenientes de la Industria 4.0, para lo cual se diseñará un programa en R o Python que permita la visualización de los resultados.
5.	DISEÑO ÓPTIMO DE EXPERIMENTOS PARA LA ECUACIÓN DE WESTLEY DE LAS REACCIONES ENZIMÁTICAS CON INHIBICIÓN POR EXCESO DE SUSTRATO.	M ^a JESÚS RIVAS LÓPEZ.	Lo habitual en una reacción enzimática es que aumentando la cantidad de sustrato se consiga un aumento de la cantidad de producto. Existen, sin embargo, reacciones enzimáticas que presentan inhibición por exceso de sustrato consistente en que, a partir de cierta concentración de sustrato, el proceso presenta inhibición. El trabajo consiste en el estudio de las técnicas de diseño óptimo de experimentos para modelos



VNIVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL



800 AÑOS

1218 - 2018

TRABAJOS DE FIN DE GRADO OFERTADOS -GRADO EN ESTADÍSTICA- CURSO 2018-19

			no lineales y su posterior aplicación a la estimación de forma precisa de los distintos parámetros que aparecen en la ecuación de Westley, que más utilizada a la hora de modelar las reacciones enzimáticas que presentan este tipo de inhibición.
6.	MINERÍA DE DATOS, DATOS MASIVOS Y LEY DE PROTECCIÓN DE DATOS EUROPEA	Mª TERESA CABERO MORÁN	: La minería de datos y los datos masivos han evolucionado y generalizado muy deprisa en los últimos años. El impacto ha sido inmejorable en el campo científico y en todo tipo de investigaciones. La revolución en este apartado de la estadística ha llegado a metas que parecían insospechadas. Sin embargo, las consecuencias de estas disciplinas pueden ser profundísimas. La recolección masiva de los datos de las personas y la mala gestión pueden afectar a los derechos de los ciudadanos. Este hecho es desconocido aún para muchos y sumamente importante a la hora de utilizar la información. El trabajo propuesto consistirá en dar una visión actual de lo mencionado anteriormente, incluyendo cómo afecta la Ley de Protección de Datos europea, como marco regulador de la seguridad y privacidad.
7.	PROPIEDADES GEOMÉTRICAS DEL DISEÑO ÓPTIMO DE EXPERIMENTOS	JUAN M. RODRÍGUEZ DÍAZ	El Diseño Óptimo de Experimentos es una disciplina relativamente joven, creada a principios del S. XX y que en los últimos tiempos está experimentando un enorme desarrollo. Existe un gran interés en encontrar los mejores puntos en los que tomar las observaciones para obtener unas estimaciones óptimas de los parámetros del modelo de regresión. Existen diversos criterios de optimización, algunos de ellos resultantes de las distintas maneras de minimizar la matriz de covarianzas de los estimadores. En varios casos existe una caracterización geométrica de los diseños óptimos, que puede ser aplicada en la búsqueda de los mismos. Este trabajo se ocupará de recopilar y describir adecuadamente esas situaciones, ilustrándolas con convenientes ejemplos.
8.	ALGUNAS CARACTERÍSTICAS DE LAS OFERTAS DE TRABAJO EN INTERNET.	ÁNGEL ALCALÁ HERNÁNDEZ	El trabajo que se propone consiste en encontrar y analizar las características más sobresalientes de las ofertas de trabajo que aparecen en las diferentes páginas Web, principalmente Infojobs. Empleando análisis tanto descriptivos como inferenciales de las diferentes ofertas, en este caso, será necesario utilizar el Método de Muestreo más adecuado, así como la utilización de los estimadores que lleven a los mejores resultados posibles. El estudio se puede realizar en diferentes ámbitos geográficos, España, comunidades autónomas, provincias, así como comparaciones entre ellas. También es posible estudiar las características de las ofertas según la categoría de las mismas, comerciales, dependientes, ingenieros, etc., y en definitiva emplear cualquier filtro, de los que se utilizan en dichas páginas, para elegir un tipo determinado de oferta laboral a estudiar. El procedimiento más sencillo de muestreo a utilizar en estos casos es el Muestreo Sistemático, pero en cualquier caso necesitamos un tamaño de muestra suficiente, lo cual significa un esfuerzo considerable. Las principales dificultades que podemos tener es la falta de homogeneidad en la presentación de la información de las ofertas, incluso dentro de las mismas páginas web, como consecuencia nos veremos obligados a uniformar estos datos con cierta subjetividad, con la consiguiente pérdida de calidad. El alumno debe manejar Estadística Descriptiva y los diferentes Métodos de Muestreo, así como los estimadores más habituales en cada caso.
9.	APLICACIÓN WEB INTERACTIVA DE TÉCNICAS ESTADÍSTICAS MULTIVARIANTES MEDIANTE SHINY.	ANA BELÉN NIETO LIBRERO NEREA GONZÁLEZ GARCÍA	El avance de la tecnología ha dado lugar a bases de datos formadas por un gran número de variables en múltiples disciplinas. Para el análisis de esta información, la informática y las técnicas de análisis multivariante se han convertido en un requisito imprescindible. Por sus múltiples ventajas, R es actualmente uno de los softwares más utilizados para la implementación y uso de muchas de estas técnicas. Ahora bien, para algunos investigadores, la falta de conocimientos en programación supone uno de los mayores inconvenientes de este software. Se propone aquí la implementación web interactiva de una aplicación multivariante en Shiny, permitiendo al usuario el acceso a algunas técnicas clásicas sin necesidad de tener que manipular código de programación. Una vez revisada la bibliografía

TRABAJOS DE FIN DE GRADO OFERTADOS -GRADO EN ESTADÍSTICA- CURSO 2018-19

			correspondiente, se implementarán en esta interfaz web técnicas multivariantes como el Análisis de Componentes Principales o el Análisis de Correspondencias.
10.	REVISIÓN Y ANÁLISIS COMPARATIVO DE LOS PRINCIPALES ALGORITMOS DE CLASIFICACIÓN DE LA FAMILIA DE K-MEANS	NEREA GONZÁLEZ GARCÍA Y ANA BELÉN NIETO LIBRERO.	Uno de los objetivos de la minería de datos es el uso de técnicas automáticas de reconocimiento de patrones para detectar una estructura de grupos subyacentes a los datos. De las técnicas disponibles, el método de clúster más utilizado en la investigación es el K-means. Sin embargo, esta técnica presenta algunos inconvenientes como la aleatoriedad en la inicialización o la necesidad de conocimiento, a priori, del número de grupos. En este trabajo se realizará una revisión bibliográfica, tanto teórica como de los software disponibles, de las principales alternativas de clasificación, familia del K-means, propuestas en la literatura.
11.	APLICACIÓN DE LOS ALGORITMOS GENÉTICOS AL PROBLEMA DE ASIGNACIÓN	QUINTÍN MARTÍN MARTÍN	El trabajo que se propone se centrará en la aplicación de los algoritmos genéticos a los problemas de asignación (personas-recursos). Para ello, el alumno deberá combinar los conocimientos de programación lineal entera con los de algoritmos genéticos. Además de los operadores genéticos clásicos se propondrán otros que puedan mejorar la solución de los problemas de asignación. Se realizará un estudio comparativo de los resultados obtenidos mediante algoritmos genéticos y los obtenidos por programación lineal entera (LINDO)
12.	EL CÁLCULO DEL TAMAÑO MUESTRAL EN DISEÑO DE EXPERIMENTOS BIOLÓGICOS.	JOSÉ L. VICENTE VILLARDÓN.	Una de las cuestiones importantes, y muchas veces olvidada, en el diseño de experimentos biológicos es determinar el número de observaciones necesario para sacar conclusiones con la suficiente precisión y la suficiente confianza. El tamaño muestral depende de muchos factores entre los que se incluyen el tipo de experimento, la forma en la que se lleva a cabo, los recursos disponibles y la precisión y confianza requeridas. El trabajo consistiría en una revisión general de los métodos o bien en una revisión de la utilización de los mismos en un campo concreto y en la realización de un protocolo aplicable a proyectos de experimentación en Biología, especialmente a aquellos experimentos que se realizan con animales de laboratorio.
13.	USO DEL BIPLLOT LOGÍSTICO NOMINAL EN DATOS PROCEDENTES DE ENCUESTAS O DE CUALQUIER OTRA FUENTE (EL CONJUNTO CONCRETO DE DATOS QUEDA A ELECCIÓN DEL ALUMNO).	JOSÉ L. VICENTE VILLARDÓN	En los últimos años en el Departamento de Estadística hemos trabajado en el desarrollo de técnicas biplot para datos nominales y ordinales. El paso siguiente consiste en aplicaciones de las técnicas desarrolladas en diversos campos, por ejemplo en las tradicionales encuestas de opinión. Como ejemplo, en el barómetro de Febrero de 2018 del CIS se tratan temas relacionados con la salud y con medicinas alternativas que pueden estudiarse con este tipo de técnicas. El barómetro puede descargarse en http://www.cis.es/cis/opencms/ES/formulario.jsp?dwld=/Microdatos/MD3205.zip Además de este ejemplo hay innumerables archivos de datos disponibles en internet. El trabajo consistiría en el la aplicación de las técnicas multivariantes para representar datos nominales y ordinales a archivos de datos elegidos por el propio alumno y para los que tenga cierto conocimiento previo del problema. El trabajo puede extenderse para varios alumnos siempre que no coincidan en el tema y la encuesta concreta. El objeto es la divulgación y extensión de las técnicas recientemente desarrolladas.
14.	LOS MÉTODOS DE MACHINE LEARNING (APRENDIZAJE AUTOMÁTICO) EN GENÓMICA, CARDIOLOGÍA, CÁNCER, GENÉTICA O CUALQUIER OTRA ESPECIALIDAD DE SU ELECCIÓN.	JOSÉ L. VICENTE VILLARDÓN	Es conocido que los métodos de Machine Learning, cuyo objeto es el desarrollo de algoritmos computacionales que aprendan con la experiencia, permiten que los ordenadores asistan a las personas en el análisis de conjuntos de datos grandes y complejos, en particular, en datos procedentes de la genómica y campos relacionados. En el análisis de datos de expresión génica hay varias áreas de análisis como son la predicción de clases (class prediction), la búsqueda de grupos homogéneos (class Discovery) o la búsqueda de genes desregulados. En todas estas áreas las técnicas de aprendizaje automático pueden ser de utilidad. En este trabajo trataremos de realizar una revisión de los métodos que se han utilizado hasta el momento en este contexto y trataremos de explorar el potencial de otros nuevos, tanto en el contexto de



VNIVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL



800 AÑOS

1218 - 2018

TRABAJOS DE FIN DE GRADO OFERTADOS -GRADO EN ESTADÍSTICA- CURSO 2018-19

			la investigación como en el de la detección de riesgos y sus variantes genéticas asociadas. Las principales técnicas serán aplicadas y comparadas en un conjunto de datos reales, por ejemplo, los procedentes del proyecto NCI60 para el estudio de distintas líneas celulares correspondientes a distintos tipos de cáncer.
15.	ANÁLISIS DE MODELOS DE REDES ALEATORIAS	M. TERESA SANTOS MARTÍN ÁNGEL MARTÍN DEL REY	<p>El Análisis de Redes Complejas es una disciplina relativamente reciente que nace como herramienta utilizada por antropólogos y sociólogos para el estudio de las estructuras sociales y culturales. Posteriormente atrajo la atención de investigadores de disciplinas más técnicas debido al uso natural de la Teoría de Grafos para fundamentar matemáticamente los conceptos empleados, y al empleo de las Ciencias de la Computación para el procesamiento masivo y visualización de la gran cantidad de datos que se pueden emplear en algunas de sus aplicaciones.</p> <p>Entre los diferentes tipos de redes, tienen especial importancia las redes aleatorias no sólo por su caracterización matemática y estadística sino por la gran cantidad de ámbitos de aplicación en donde aparecen.</p> <p>El objetivo de este trabajo es definir y presentar las principales propiedades de este modelo de redes complejas, centrando especialmente la atención en el modelo de Erdős y Renyi. Asimismo, se detallarán los principales algoritmos de construcción de las citadas redes y sus aplicaciones más importantes.</p>
16.	GENERACIÓN DE SECUENCIAS PSEUDOALEATORIAS DE BITS DE USO EN CRIPTOGRAFÍA	ÁNGEL MARTÍN DEL REY	<p>La generación de secuencias pseudoaleatorias de bits juega un papel fundamental en el diseño de criptosistemas de clave secreta (criptosistemas de cifrado en flujo y cifrado en bloque). Para determinar si un determinado algoritmo produce secuencias de bits con buenas propiedades criptográficas es necesario aplicar una batería de tests estadísticos que determinan su calidad.</p> <p>El objetivo de este trabajo fin de grado es realizar un análisis detallado de los principales generadores pseudoaleatorios de bits junto con las técnicas estadísticas utilizadas para la determinación de la calidad de las secuencias producidas</p>