



Biblioteca del Instituto de Estadística y Cartografía de Andalucía

Resúmenes de revistas
Junio 2016



Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN, CIENCIA Y EMPLEO



Estadística española : revista del Instituto Nacional de Estadística, ISSN 0014-1151
Número 189 (enero – abril 2016)

Series enlazadas de empleo y VAB para España, 1955-2014

P. 5-41

Angel de la Fuente Moreno

Resumen

En este trabajo se elaboran series homogéneas de distintos agregados de empleo y de VAB a precios corrientes y constantes para el conjunto de España durante el período 1955-2014. Las series se construyen mediante el enlace de diversas bases de la CNE y de la Contabilidad Trimestral, introduciéndose también una corrección tentativa para reconciliar las series de empleo de la CNE con las de la EPA.

Unbiased corrected classic estimates

P. 43-56

Mariano Ruiz Espejo

Abstract

The author studies the efficiency of an unbiased corrected linear regression estimator, which can be more efficient than the classic linear regression estimator, with simulation techniques. Also we provide four new unbiased corrected classic estimates from linear regression, product, and mean-of-the-products estimators.

Estudio para la estimación del número de trabajadores en convenios colectivos de trabajo

P. 57-79

Irene Patricia Méndez Candela, Begoña Varela Merino

Resumen

El presente estudio ofrece una estimación del número de trabajadores sujetos a convenios colectivos de ámbito sectorial alternativa a la aportada por las comisiones negociadoras en la Hoja Estadística. Para ello se utiliza la información contenida en el registro de Cuentas de Cotización de la Seguridad Social relativa a convenios colectivos de trabajo y en la base de datos de la Estadística de Convenios Colectivos de Trabajo. Este estudio responde a la necesidad indicada en una de las recomendaciones del Plan Estadístico Nacional para la mejora de la estimación del número de trabajadores afectados por convenios colectivos.

Medición de la percepción social sobre el funcionamiento de la administración tributaria en España

P. 81-94

Leonel Cerno

Resumen

El presente artículo hace referencia a la metodología utilizada y a los principales resultados del Barómetro Fiscal o medición de la opinión acerca de la fiscalidad en España. Dicho estudio, llevado a cabo anualmente por el Instituto de Estudios Fiscales (IEF), utiliza la metodología de recogida a través de encuestas de opinión, en donde revela y describe la evolución de ciertas actitudes que tienen los españoles hacia la Hacienda Pública y la Administración en general, midiendo la manera ciudadana hacia los servicios públicos, la eficacia de la Agencia Tributaria, las causas y

Consistencia en la desagregación de la población. El problema del ruido y el age heaping

P. 95-117

Silvia Bermudez Parrado, Rafael Blanquero Bravo

Resumen

La población desagregada en edad simple es una herramienta básica para las oficinas estadísticas, pues es usada, por ejemplo, como denominador en el cálculo de indicadores. Sin embargo, las cifras de población para algunos ámbitos territoriales solo están disponibles en forma agrupada: normalmente, esta distribución de la población se publica en grupos quinquenales de edad más un intervalo superior abierto donde se acumula la población de mayor edad. Un problema importante al que se suelen enfrentar las diferentes oficinas estadísticas, tanto de ámbito estatal como autonómico, es la desagregación de los datos de población en grupos de edad simple, permitiendo, cuando sea requerido, incluir conocimiento demográfico y mantener la consistencia de los resultados obtenidos con las agregaciones de población de los ámbitos territoriales superiores al desagregado o, incluso, la consistencia en la evolución de ésta a lo largo del tiempo. En este trabajo se consideran técnicas de optimización para dar respuesta a este problema real, aunque muy poco estudiado en la literatura. Los procedimientos propuestos permiten también tratar un problema habitual en este tipo de fuentes estadísticas, como es la presencia de ruido de distinta naturaleza en los datos disponibles, y, en particular, el fenómeno conocido como *age heaping*.



Mapping: revista internacional de Ciencias de la Tierra, ISSN 1131-9100

Volumen 25, número 175 (enero – febrero 2016): "VI Jornadas Ibéricas de Infraestructuras de Datos Espaciales (Parte I)"

IDEARAGON. Infraestructura marco del Sistema Cartográfico de Aragón

P. 8-17

Rafael Martínez Cebolla, Fernando López Martín, David Portolés Rodríguez

Resumen

El desarrollo formal del Decreto 208/2010 que aprueba el Reglamento de Ordenación de la Información Geográfica en Aragón inició una nueva etapa para la información geográfica generada en el territorio aragonés. La regulación del Sistema Cartográfico de Aragón (SCA) parte de los principios de coordinación, cooperación y eficiencia interadministrativa considerando que las actividades (cartográficas o procedimientos administrativos vinculados) que regula son ante todo un servicio público. Los instrumentos que componen este sistema son los siguientes: el Plan Cartográfico de Aragón como instrumento básico de programación de la actividad cartográfica, la Cartografía Oficial, la Norma Cartográfica, el Registro Cartográfico, el Nomenclátor Geográfico, la Cartoteca de Aragón y la Infraestructura de Datos Espaciales de Aragón (IDEARAGON). Esta última se erige como el marco aglutinador y soporte tecnológico integrador de los instrumentos del SCA al objeto de ordenar, gestionar y difundir la información geográfica mediante las oportunas piezas tecnológicas interoperables de manera que pueda ser controlada a través del Registro Cartográfico, organizada y almacenada en la base de datos geográfica, documentada mediante la confección de sus correspondientes metadatos, descubierta mediante el buscador geográfico y explotada mediante servicios Web y aplicaciones geográficas. El resultado final es que el componente tecnológico de IDEARAGON ha logrado difundir las acciones realizadas por el componente político (organización, marco legal e instrumentos) a través de la explotación del componente geográfico para que sea la sociedad la gran beneficiada en el uso de la información geográfica.

Explotación de datos LiDAR en la nube Servicios de acceso, procesamiento y visualización remota

P. 18-25

Salvador Bayarri, Rafael Gaitán, Jordi Torres

Resumen

El rápido incremento en la disponibilidad de datos LiDAR, con nuevos y menos costosos métodos de captura, y el desarrollo de aplicaciones en campos más allá de los sistemas de información geográfica convencionales (por ejemplo, en ingeniería civil, arqueología y patrimonio, atención a emergencias, gestión forestal, etc.) conlleva la necesidad de disponer de sistemas de almacenamiento, distribución, procesamiento y visualización para estos datos masivos. La tecnología y los costes para el alojamiento de datos y procesos convierten a las arquitecturas basadas en nube en una opción muy atractiva para implementar estos servicios, especialmente para aquellas organizaciones que no pueden permitirse una infraestructura dedicada. En el artículo se discuten las limitaciones para la implementación de servicios adecuados mediante los estándares OGC y se describe el desarrollo de otros sistemas con servicios no estándares, como PointCloudViz. Se presentan algunos casos prácticos de uso en las áreas de diseño de infraestructuras, gestión de desastres naturales y modelado geológico, destacando las ventajas de la aproximación basada en nube en el contexto de esas aplicaciones.

geoEuskadi. Infraestructura de Datos Espaciales de Euskadi, una gran plataforma tecnológica y organizativa a disposición de la administración, instituciones y ciudadanía

P. 26-31

Juan Carlos Barroso, Javier Díaz de Guereñu, Xabier Garitano, Ion Martínez de Ilarduya

Resumen

geoEuskadi comenzó en 2004 como iniciativa corporativa de recopilación de datos geográficos y definición de herramientas corporativas para facilitar el acceso a los mismos. Desde entonces ha ido evolucionando guiado por las directrices INSPIRE y el avance en tecnologías y estándares en pro de la interoperabilidad, difusión y explotación de los datos espaciales. Es un proyecto transversal dentro de la Administración Pública de Euskadi que sirve de plataforma organizativa y tecnológica para dinamizar el uso y explotación de la información geográfica en diferentes ámbitos y disciplinas, tanto dentro como fuera de la organización. Está liderado desde el Servicio de Información Territorial de la Dirección de Planificación Territorial y Urbanismo de Gobierno Vasco. Es un proyectomadura que cuenta ya con servicios ampliamente requeridos donde los datos de calidad armonizados, la tecnología y la formación y difusión al exterior son todos componentes imprescindibles.

Futuro del planeamiento urbanístico en la IDE Extremadura Planeamiento digital y firma electrónica

P. 32-39

Loreto del Viejo Trejo, Alberto Aparicio Ríos, Fernando Ceballos Zúñiga-Rodríguez, Javier Rubio Muriel

Resumen

Nos encontramos ante una sociedad que cada vez demanda más información en red, de forma inmediata, fiable y válida. Esto supone que sea necesario contar con dicha información en un soporte digital que permita hacer llegar dicha información al usuario final, preferiblemente de una forma fiel a la realidad. El problema que se plantea a la hora de mostrar documentos de planeamiento urbanístico en red, radica principalmente en la falta de capacidad de ofrecer esta información con la validez legal que intrínsecamente tiene. Actualmente conseguimos mostrar la información urbanística principalmente a nivel gráfico como temáticos a través de servicios estándar (WMS) en la Infraestructura de Datos Espaciales de Extremadura, pero la geometría es sólo una parte de este tipo de información, la otra parte importante la constituye la carga normativa asociada a los ámbitos territoriales que la componen. Esta información en su conjunto tendrá mayor validez para el usuario si a su vez tiene un determinado peso legal. El primer paso que se dio en Extremadura a este respecto consistió en escanear el planeamiento vigente en toda la región, y facilitar así el libre acceso por parte de los administrados a los ficheros que integran los documentos de planeamiento a través de Internet, pero la información urbanística con validez legal continúa siendo el documento diligenciado en papel. El siguiente hito alcanzado ha sido contar con parte del planeamiento redactado completamente desde su origen en un formato digital, de manera sistematizada y conectando cada ámbito geográfico al conjunto de documentos y normativa que le afecte, lo que ha permitido que sea posible consultar la información urbanística a través de los ámbitos gráficos, utilizando como herramienta de consulta la IDE Extremadura. Esto supone contar con el soporte necesario para someter estos documentos a un proceso paralelo de validación legal a través de firmas digitales válidas, que permitirá en un futuro próximo ofrecer al usuario un documento digital con validez legal, respetando la precisión del mismo tanto a nivel gráfico como normativo, servido directamente a través de la IDE Extremadura.

Geoportal SIGNA v.3.0 del IGN-CNIG Funcionalidades y novedades

P. 40-47

Celia Sevilla Sánchez, Miguel Villalón Esquinas, Jaime Sánchez Fanjul

Resumen

El Sistema de Información Geográfica Nacional (SignA) es un proyecto estratégico del IGN-CNIG que tiene como finalidad la integración de los datos y servicios del IGN-CNIG en un SIG, para su gestión, análisis y consulta, a través de Internet, lo que a su vez implica el desarrollo de un geoportal propio versátil, interoperable y eficiente. El geoportal del SignA, se abrió al público en diciembre de 2010 y acaba de publicar la tercera versión, integrando lo mejor de los mundos SIG e IDE en una única herramienta. Tras casi 5 años, el geoportal ha ido aumentando su demanda y

actualmente recibe una media de 15.000 visitas al mes, siendo una herramienta de uso diario de muchos de nuestros usuarios para multitud de aplicaciones. En el artículo se hará un breve repaso de las características del portal, profundizando en las mejoras respecto a las versiones anteriores y en las novedades. En primer lugar, esta versión del geoportal se ha actualizado a la última versión del software Geospatial Portal de Intergraphy se ha migrado la base de datos de Oracle Spatial a PostGreSQL en su extensión PostGis. Entre las actualizaciones cabe destacar la de la base de datos del SignA a partir de la última versión de BTN100 y de Cartociudad, creando, además, un mecanismo para su actualización periódica. Otras de las actualizaciones son la conexión a los nuevos servicios de enrutamiento de Cartociudad, de cálculo de perfiles longitudinales, al buscador de nombres geográficos del IGN y la visualización de los datos a través de los servicios WMTS del IGN (mapa ráster, mapabase, ortofoto, etc.).

Nuevo servicio de localización INSPIRE en la IDE de Navarra Publicación de una nueva versión de su catálogo con GeoNetwork Opensource

P. 48-55

César Díaz, Pablo Echamendi, Pedro Mendive, Carlos Sabando

Resumen

Desde sus comienzos, la Infraestructura de Datos Espaciales de Navarra (IDENA), a través de su catálogo de metadatos, ha permitido a sus usuarios localizar datos y servicios geográficos disponibles en el Sistema de Información Territorial de Navarra (SITNA). Con el fin de facilitar el acceso a la información y mejorar la gestión del registro de metadatos, IDENA ha estrenado una nueva versión de su catálogo utilizando GeoNetwork opensource, versión 2.10.4. Este producto, ofrece una aplicación que permite gestionar un catálogo de recursos geográficos y cumplir al mismo tiempo con los estándares. Esta nueva versión del catálogo de IDENA aporta un potente buscador a través de filtros y listas predictivas que, utilizando la propia información contenida en los metadatos, implementa las búsquedas por organización, fechas o temática INSPIRE, entre otras. También son necesarios desarrollos y personalizaciones para que GeoNetwork quede perfectamente integrado en la infraestructura de IDENA. Al mismo tiempo, esta evolución ha permitido configurar un servicio de localización CSW para la consulta y localización on-line, conforme con las especificaciones de la Directiva INSPIRE. Dicho servicio actualmente soporta la posibilidad de realizar harvesting, operación mediante la cual nodos externos pueden localizar y recolectar los metadatos de un catálogo en cualquier momento, evitando la transferencia física de la información cuando esta es requerida. La URL de acceso al cliente es: <http://idena.navarra.es/catalogo/> y la de acceso al servicio CSW es <http://idena.navarra.es/ogc/csw>.

El proyecto europeo Minerals4EU, un caso práctico de implantación de la Directiva INSPIRE. La experiencia de hacer compatible los objetivos específicos de un proyecto y su conformidad con la Directiva INSPIRE en el tema de Recursos Minerales

P. 56-63

María J. Mancebo, María Teresa López, Margarita Patricia Sanabria

Resumen

El obligado cumplimiento de las Normas de Ejecución comunes (Implementing Rules) y las Guías Técnicas o Directrices (Technical Guidelines) dirige y facilita la implementación de la Directiva INSPIRE dentro de un marco de compatibilidad e interoperabilidad en un contexto comunitario. Aun así, deja abiertas distintas posibilidades técnicas y de gestión que hay que abordar en los proyectos individuales y que pueden dificultar el éxito en la implantación. A los problemas ya conocidos que supone la transformación de la información a los modelos de datos INSPIRE, hay que sumar otros aspectos como son la falta de experiencias técnicas y de gestión, y la necesaria obligación de hacer compatibles los objetivos específicos de un proyecto y su conformidad con la Directiva INSPIRE. Abordar un proyecto europeo como Minerals4EU ha supuesto una experiencia real en la que afrontar todas y cada una de las problemáticas descritas anteriormente para darles una solución, unas veces más exitosa que otras. Ha sido necesario profundizar en los aspectos de gobierno del proyecto, llegando a acuerdos para reforzar los documentos de especificaciones técnicas y acotar las vías tecnológicas de implantación del proyecto. En resumen, el desarrollo de este proyecto, además de demostrar que la implementación de INSPIRE es posible, ha permitido obtener un conocimiento y una experiencia muy valiosos de lo que supone en la práctica la implementación de INSPIRE en todos sus aspectos, y puede

constituirse como un buen ejemplo para la elaboración de guías de buenas prácticas, balance del software y de las soluciones utilizadas, usabilidad del sistema, entre otras.

La gestión y la tecnología de los Identificadores Persistentes. Un análisis de cómo se afronta el problema de los Identificadores Persistentes y cómo el Esquema Nacional de Interoperabilidad condiciona la implementación de la Directiva INSPIRE

P. 64-73

Francisco J. López-Pellicer, Jesús Barrera, Julián González, F. Javier Zarazaga-Soria, Emilio López, Paloma Abad

Resumen

Los Estados miembros de la UE tienen de plazo hasta diciembre de 2021 para implementar completamente la Directiva INSPIRE y sus Reglas de Implementación, potencialmente siguiendo las Guías Técnicas que están basadas en normas y estándares internacionales (ISO/TC211, CEN/TC 287 y OGC). Sin embargo, los diferentes enfoques en la aplicación de las normas y estándares, su evolución normal y los desafíos asociados a la coordinación de los cambios entre ellos, junto con las diferentes opciones tecnológicas que están siendo adoptadas y los diferentes marcos legislativos nacionales están causando problemas de interoperabilidad inesperados. Uno de los problemas detectados es la falta de una Guía Técnica para la gestión de los identificadores globales y persistentes, o PID, que publicados por un proveedor pueden ser utilizados por aplicaciones externas para referirse al objeto espacial. Algunos Estados miembros de la UE ya han creado estructuras de gobierno, procesos, reglas y herramientas para crear, gestionar, mantener y usar PID en sus IDE pero han seguido diferentes aproximaciones. Estos problemas han sido identificados en el proyecto ARE3NA del JRC que ya ha producido algunos informes sobre este tema. La comunidad IDE no es la única que se enfrenta al reto de los PID. Existen otras comunidades que gestionan PID con éxito (LSID en bioinformática y biodiversidad, ISBN en publicaciones, DOI en objetos digitales, IBAN en banca, etc.) y de las que podemos aprender. El objetivo del trabajo es investigar qué alternativas organizativas y tecnológicas existen para coordinar el mantenimiento y los sistemas de gestión de identificadores persistentes, y analizar si la decisión de INSPIRE basada en HTTP URI es consistente con dicha experiencia y compatible con el Esquema Nacional de Interoperabilidad.



Mapping: revista internacional de Ciencias de la Tierra, ISSN 1131-9100

Volumen 25, número 176 (marzo – abril 2016): "VI Jornadas Ibéricas de Infraestructuras de Datos Espaciales (parte II)"

Interoperabilidad práctica de los geoportales de la IDEE. Hacia la optimización de la parte cliente

P. 8-15

Antonio F. Rodríguez, Ana Cevidanes, Alejandra Sánchez, Paloma Abad, Marta Juanatey

Resumen

En este artículo se plantea una metodología de verificación de la interoperabilidad y usabilidad prácticas de los geoportales de la IDEE, basada en documentos anteriores de trabajo, la experiencia acumulada en el equipo IDE del CNIG y el Proyecto Fin de Carrera de Ana Cevidanes. Hasta ahora se ha dedicado mucha atención y se ha invertido un notable esfuerzo en el contexto de las IDE, pero creemos que se ha prestado poca atención a los geoportales y clientes que explotan dichos recursos. Por eso, nos parece oportuno en el momento actual abordar criterios objetivos de armonización e interoperabilidad de geoportales que contribuyan a optimizarlos desde criterios técnicos y objetivos que atiendan a su interoperabilidad y facilidad de uso. También se informa de los resultados obtenidos de la aplicación de la metodología definida a los geoportales de referencia, nacionales y regionales de la IDEE para evaluar su interoperabilidad y usabilidad, identificar áreas de mejora y contribuir a su optimización. Finalmente se realizan una serie de consideraciones que tratan de balancear el coste y el resultado de las posibles medidas correctivas derivadas del análisis realizado, con lo que se llega a esbozar un plan de implantación que aborda primero las medidas de menor coste y mayor impacto en la interoperabilidad de los recursos gestionados.

Panorama de datos y servicios abiertos en el campo de la información geográfica en España. Avances y novedades en la liberación de recursos de información geográfica

P. 16-25

Pedro Vivas, Antonio F. Rodríguez, Emilio López, Juan Manuel Rodríguez, Agustín Cabria, Marta Juanatey, Alejandra Sánchez

Resumen

En esta comunicación se repasa la situación en cuanto a datos y servicios abiertos en España en el ámbito de la información geográfica, teniendo en cuenta la definición de datos abiertos formulada por la Open Knowledge Foundation. En primer lugar, se resumen el contenido de la Directiva 2013/37/UE de Reutilización de la Información del Sector Público y la Ley que la traspone (Ley /2014). En segundo lugar se analizan la oportunidad y contenido de la International Open Data Charter, que trata de extender la G8 Open Data Charter. En tercer lugar se expone un resumen de la política de datos y servicios de los principales productores de cartografía oficial en España, fundamentalmente los grandes productores nacionales y los que desempeñan ese papel en el ámbito regional. Para finalizar, se analiza la situación en cuanto a servicios abiertos, se extraen unas conclusiones y se formula un conjunto de recomendaciones generales sobre este tema. Esta comunicación intenta contribuir a la puesta en marcha de un Grupo Técnico de Trabajo (GTT) de CODIIGE dedicado a políticas de datos y licencias, en el que estén representados los principales productores de datos geográficos y proveedores de geoservicios, para intercambiar experiencias, buenas prácticas y modelos de negocio y para establecer recomendaciones por consenso que contribuyan a avanzar y mejorar la situación en este campo.

Hubs virtuales para facilitar el acceso a Datos Abiertos

P. 26-33

Miguel Ángel Latre, Francisco J. López-Pellicer, Nargess Kamali, Michael Bauer, Mattia Previtali, Raffaella Brumana, Stefan Braumann, Helga Kuechly, Paolo Mazzetti, Stefano Naviti

Resumen

El proyecto europeo ENERGIC OD (European NETwork for Redistributing Geospatial Information to user Communities - Open Data), financiado por la Unión Europea dentro del Programa Marco para la Competitividad y la Innovación (CIP), comenzó su andadura de tres años en octubre de 2014. El objetivo del proyecto es el de facilitar el uso de datos geográficos abiertos procedentes de fuentes diversas a través de la creación de distribuidores virtuales de datos (hubs virtuales) y la creación de servicios y aplicaciones innovadores que los utilicen. Uno de los principales obstáculos para la reutilización de datos abiertos es la heterogeneidad existente tanto en datos y formatos como en servicios, que requiere a usuarios y desarrolladores de software grandes esfuerzos en términos de acceso a los datos y de armonización de los mismos para poder ser utilizados. La adopción de estándares, aunque la reduce, no consigue evitar completamente esta heterogeneidad, especialmente en contextos interdisciplinares. Los hubs virtuales de ENERGICOD hacen frente al problema adoptando un enfoque basado en la mediación: componentes especiales (intermediarios o brokers) armonizan interfaces de servicio, modelos de datos y metadatos, permitiendo descubrir y acceder a datos e infraestructuras heterogéneas. ENERGIC OD desplegará un total de cinco hubs virtuales a nivel nacional en Francia, Alemania, Italia, Polonia y España y otro más a nivel europeo y desarrollará sobre los mismos un conjunto inicial de diez aplicaciones multidisciplinares basadas en la explotación de datos geográficos abiertos. ENERGIC OD creará nuevas oportunidades de mercado, abriendo todavía más el acceso a la información del sector público y a datos de investigación.

Evolución de CartoCiudad: nuevo visualizador del proyecto

P. 34-39

Alicia González, Ana Velasco, Patricia Trigo, Julián González, Gloria Andrés

Resumen

El proyecto de producción colaborativa CartoCiudad (red viaria continua que discurre sobre cartografía urbana por todo el territorio nacional y contiene asociada la información postal y censal) es el resultado de la armonización e integración de datos oficiales de varias fuentes y evoluciona hacia el objetivo de satisfacer las principales demandas de los usuarios en relación con los datos y a los servicios desarrollados sobre ellos. La principal novedad que se presenta es la publicación del nuevo visualizador de CartoCiudad, que mejora tanto en diseño como en funcionalidad el anterior. Está desarrollado con el software libre OpenLayers 3 y se basa en jQuery UI, en concreto utilizando Widget Factory. Es el resultado de la integración de distintos componentes web de búsqueda y de cálculo de rutas sobre una API básica que muestra cartografía. Esta API está personalizada para los datos y funcionalidades que se ofrecen a través del geoportal de CartoCiudad. También permite la combinación de capas del proyecto con la carga de otras procedentes de otros servicios y el cálculo de rutas entre puntos a lo largo de la red viaria que los conecta. En paralelo a los desarrollos anteriormente mencionados, y continuando con la línea de evolución del producto para alcanzar la conformidad con la Directiva INSPIRE, que ya se inició en materia de direcciones y se materializó a través de la publicación de los correspondientes servicios WMS y WFS también, en materia de transporte, se ha creado el servicio WFS transportes por carretera a partir de los datos de CartoCiudad.

Ser o no ser interoperable, esa es la cuestión. Estudio sobre las barreras de alto nivel a la interoperabilidad de los servicios geográficos OGC de las Infraestructuras de Datos Espaciales de Europa

P. 40-45

Alejandro Guinea de Salas, Olga López de Turiso Martínez, Estibaliz Pascual Calvo

Resumen

En la búsqueda de datos geoespaciales ofrecidos por los diversos gobiernos europeos nos encontramos con muy diversos procesos de autenticación y autorización, que dificultan e incluso imposibilitan la generación de aplicaciones capaces de trabajar con recursos protegidos procedentes de fuentes diversas. Pese a que la Directiva INSPIRE aboga por la disponibilidad de información geográfica relevante que permita la generación y seguimiento de las políticas de

impacto o de dimensión territorial de la Unión Europea, deja abiertos los métodos de control de acceso que es posible aplicar. Los diversos países siguen distintos criterios en este control, que se entrelaza con la gestión de derechos del autor y licencias. Podemos encontrar desde países abiertamente decididos a la compartición de datos de forma gratuita a otros donde la información está sujeta al abono de tarifas. Así, la restricción de accesos parece obedecer a motivos diversos: desde la necesidad de controlar la identidad de la persona física o jurídica que accede a los datos, por motivos de seguridad u otros motivos, hasta la recaudación económica en contrapartida al uso de los datos, siendo ambas situaciones no incompatibles entre sí. Al tiempo, existen casos en los que, tras procesos de autenticación difícilmente automatizables, nos encontramos con datos gratuitos y abiertos. En el artículo se presentarán las tipologías de autenticación encontradas en los más de 1000 recursos estudiados a nivel europeo en el marco de Copernicus, cómo estas influyen en el desarrollo de aplicaciones interoperables, y qué recomendaciones son aplicables para mejorar la interoperabilidad real de los servicios de las IDE.

Generación de Información Geográfica de Referencia (IGR) de Poblaciones: automatización de procesos

P. 46-57

Tania Gullón Muñoz-Repiso, José Antonio Merino Martín, Lorenzo Camón Soteres

Resumen

El Instituto Geográfico Nacional (IGN) está inmerso en un cambio en el proceso productivo cuyo objetivo es que la producción se realice por capas temáticas que constituirán la Información Geográfica de Referencia (IGR) del IGN, de la cual derivarán el resto de productos, evitándose así duplicidades. De esta manera se siguen los principios de la Directiva INSPIRE para el establecimiento de una infraestructura de información espacial en Europa y de la correspondiente Ley sobre las Infraestructuras y los Servicios de Información Geográfica en España (LISIGE) que traspone dicha directiva. La metodología de generación de IGR pretende ser lo más homogénea y automatizada posible, manteniendo la máxima exactitud geométrica que se puede alcanzar basándose en los datos disponibles. La mencionada IGR comprende diferentes temas: redes de transporte, hidrografía, poblaciones, etc. Este artículo se centra en la IGR referente a poblaciones, entendiendo población como el área geográfica que delimita un territorio claramente diferenciado por la ocupación de los asentamientos humanos y está identificada de manera inequívoca por un nombre. En este artículo se describen los procesos automáticos que se han desarrollado en el IGN para la obtención de las geometrías de las poblaciones de una forma precisa, sostenible, objetiva y homogénea para todo el territorio nacional dando forma a unas especificaciones cuyas características semánticas y geométricas satisfagan los requisitos técnicos impuestos por las necesidades actuales de los usuarios de datos geoespaciales nacionales y europeos en esta materia. Como resultado de la aplicación de la metodología automática y de la integración de datos de diversas fuentes de datos se consigue asignar cada parcela catastral a una población determinada y generar automáticamente un contorno para cada población. Dicho contorno es coincidente con el borde exterior de las parcelas catastrales, englobando tanto parcelas destinadas a uso residencial como parcelas destinadas a otros usos, industrial, zonas verdes, etc. quedan servicio a los habitantes de la población así como el entramado urbano que cohesionan y estructura dicho territorio.

Generación de Información Geográfica de Referencia (IGR) de Hidrografía: estado actual de la producción

P. 58-65

Celia Sevilla Sánchez, Eduardo Núñez Maderal, Nuria Valcárcel Sanz, Julian Delgado Hernández, Gema Martín-Asín, Ana de las Cuevas, Miguel Villalón Esquinas, Antonio F. Rodríguez Pascual, Jaime Sánchez Fanjul

Resumen

Disponer de Información Geográfica de Referencia (IGR) responde a la necesidad básica de conocimiento del territorio, para referenciar y localizar cualquier fenómeno espacial de modo unívoco. El Instituto Geográfico Nacional (IGN) puso en marcha hace un año un plan de producción coordinada de IGR de alta resolución conforme a INSPIRE, alineado con las decisiones sobre Gestión de la Información Geoespacial de Referencia de las Naciones Unidas, y acorde a los requerimientos de IGR a nivel nacional, europeo y global. El IGN, a través de grupos de trabajo internos, colabora con los Grupos Técnicos de Trabajo Técnicos creados dentro del Consejo Directivo de la Infraestructura de Información

Geográfica en España (CODIIGE). El artículo presenta los trabajos que se han llevado a cabo dentro del grupo interno de hidrografía del IGN para satisfacer las necesidades de los usuarios en materia de información hidrográfica. El objetivo es alcanzar las directrices básicas para una producción, actualización y explotación. Esta metodología debe garantizar IGR hidrográfica lo más exacta, objetiva, interoperable y actualizada, y producida de la forma más automática posible. Los primeros pasos consistieron en recoger y analizar los requerimientos y las necesidades de un primer conjunto de usuarios, y la legislación vigente; a continuación se realizó el análisis de las especificaciones de INSPIRE de hidrografía y se crearon las especificaciones del producto de datos con su correspondiente catálogo de objetos geográficos. Actualmente se está trabajando en procesos de producción e integración de datos, en nuevas técnicas de extracción automática de IGR de hidrografía a partir de datos LIDAR, en los controles de calidad y en la carga de datos.

Generación de Información Geográfica de Referencia (IGR) de Redes de Transporte: estado actual de la producción

P. 66-72

Alicia González Jiménez, Ana Velasco Tirado, Lorenzo Camón Soteres, Gloria Andrés Yusa, Patricia Trigo Gambaro-Espuig

Resumen

En cumplimiento con la Directiva 2007/2/CE para el establecimiento de una Infraestructura de Información Espacial en Europa (INSPIRE) y con la ley que la traspone Ley 14/2010, de 5 de julio, sobre las Infraestructuras y los Servicios de Información Geográfica en España (LISIGE) que fomentan la reutilización de los datos, el Instituto Geográfico Nacional se encuentra actualmente inmerso en un cambio de sus procesos productivos con el objetivo de que la generación de los distintos productos de información geográfica con temáticas comunes se realice por derivación de los datos almacenados en las Bases de Datos de Información Geográfica de Referencia (IGR). Los datos de dichas IGR deben corresponderse con la información geográfica de mayor exactitud geométrica, semántica y temporal posible, con objeto de satisfacer los requisitos de los usuarios de datos geoespaciales, además de permitir la derivación de los distintos productos que produce el IGN. En materia de transportes, el IGN está produciendo la Base de datos de «Redes e Infraestructuras del Transporte» (RT) bajo los criterios de ser conforme con la normativa. Se trata de una red tridimensional de transporte multimodal y cobertura nacional, generada en primera instancia por integración de datos procedentes de productos del IGN que contienen esta temática y de las fuentes de referencia existentes en cada modo de transporte. Tras la primera versión, el paso siguiente consistirá en la definición y aplicación de la metodología de actualización que permita aumentar la exactitud posicional y semántica así como la frecuencia de actualización, que sea viable y sostenible.



Plus diplômées, moins célibataires. L'inversion de l'hypergamie féminine au fil des cohortes en France

P. 705-730

Milan Bouchet-Valat

Resumen

L'hypergamie féminine, définie comme la propension des individus à former des couples au sein desquels la femme se trouve en infériorité par rapport à l'homme, est un phénomène largement observé. Cet article analyse la formation des premières unions à l'aide de l'enquête *Étude de l'histoire familiale 1999*. Les couples dans lesquels la femme est plus diplômée que son conjoint sont plus fréquents que le cas inverse en France depuis les cohortes nées à la fin des années 1950. Ce mouvement est principalement dû à l'allongement de la scolarité des femmes qui sont désormais plus diplômées que les hommes, mais va au-delà de ce qu'imposait l'évolution de la structure de la population (hypergamie relative), traduisant une modification des préférences des individus. Enfin, nous observons que le célibat définitif des femmes n'augmente plus avec leur diplôme, alors que les plus diplômées nées avant-guerre étaient fortement désavantagées sur le marché conjugal. À l'inverse, le célibat définitif des hommes non diplômés s'est accentué, signe de l'effet négatif persistant des difficultés d'insertion professionnelle sur la conjugalité masculine. Ces résultats indiquent un net recul de la norme d'hypergamie féminine en termes de diplôme – dont la portée demeure cependant incertaine.

Fécondité et niveau d'instruction des femmes pendant le socialisme d'État en Europe centrale et orientale

P. 731-769

Zuzanna Brzozowska

Resumen

Cet article analyse l'évolution de la descendance finale des femmes nées entre 1916 et 1960 dans sept pays d'Europe centrale et orientale (Croatie, République tchèque, Hongrie, Pologne, Roumanie, Slovaquie et Slovénie). L'effet de l'élévation du niveau d'instruction sur la fécondité des cohortes est analysé en tenant compte de l'infécondité et des naissances de rang élevé à l'aide des méthodes de décomposition et de standardisation appliquées aux données issues des recensements entre 1980 et 2000. Le recul de la fécondité des cohortes résulte de l'augmentation du niveau d'instruction d'une part et de la réduction du nombre de naissances de rang élevé d'autre part. Ces deux effets négatifs sont partiellement contrebalancés par la diminution de la proportion de femmes sans enfant. L'analyse montre que les évolutions de la fécondité des cohortes ont été largement partagées dans les différents pays, malgré des différences d'intensité. Il ne semble pas exister de tendance de fécondité spécifique aux pays socialistes, excepté la proportion de femmes sans enfant qui continue de diminuer (alors qu'elle augmente dans les pays occidentaux pour les femmes nées après 1940). En Roumanie, Pologne et Hongrie, la politique fortement pronataliste a pu ralentir la baisse des parités de rang élevé.

Résumen

Cet article propose une méthode d'estimation des substitutions linguistiques basée sur le principe de la cohorte fictive et de l'analyse de survie. À partir des données d'un seul recensement, cette méthode permet d'obtenir des taux de substitution linguistique pour différents groupes de population au Canada (selon la langue transmise à la naissance, le statut d'immigrant, l'âge à l'immigration, le niveau d'études) et pour toutes les régions concernées du pays. La robustesse de la méthode est validée en comparant les résultats obtenus à partir des recensements canadiens de 1991, 1996, 2001 et 2006. Les taux de substitution linguistique selon l'âge ou la durée d'immigration sont robustes dans le temps, mais varient significativement selon les groupes de population. Ils sont très faibles chez les immigrants allophones de première génération arrivés au Canada à l'âge adulte, mais peuvent atteindre 90 % dans la seconde génération. Ces taux varient peu d'une région à l'autre du Canada pour les allophones, mais davantage pour les minorités de langue officielle, atteignant dans certains cas des taux comparables à ceux de la seconde génération d'immigrants allophones. Au Québec, où le français et l'anglais constituent deux langues de convergence pour les allophones, la hausse des substitutions linguistiques effectuées vers le français est largement tributaire des changements dans la composition ethnolinguistique de l'immigration.

Les adoptions en France et en Italie : une histoire comparée du droit et des pratiques (xix^e-xxi^e siècles)

Jean-François Mignot

Resumen

La France et l'Italie ont des droits de l'adoption proches et font partie, depuis les années 1990, des pays qui adoptent le plus de mineurs à l'international. Mais, à plus long terme, dans quelle mesure les pratiques sont-elles similaires entre ces deux pays ? Cet article retrace l'histoire du droit et des pratiques de l'adoption en France et en Italie depuis le xix^e siècle. Il apparaît que si ces deux pays ont connu des histoires semblables du droit de l'adoption, les nombres actuels d'adoptions simples sont très différents, principalement en raison du fait que les recompositions familiales après divorce sont beaucoup plus rares en Italie qu'en France. En revanche, les nombres d'adoptions plénières et les profils des adoptés pléniers sont similaires dans les deux pays : après avoir été confrontés au faible nombre de mineurs nationaux adoptables, les candidats à l'adoption des deux pays sont confrontés, depuis 2011, à la baisse du nombre de mineurs adoptables à l'international. En conséquence, les adoptés internationaux en France et en Italie tendent à être plus âgés, plus souvent en fratrie, et plus souvent malades ou handicapés qu'auparavant.

Évolution de la mortalité au cours de la transition du socialisme planifié au capitalisme d'État à Shanghai

Jiaying Zhao, Edward Jow-Ching Tu, Guixiang Song, Adrian Sleight

Resumen

Contrairement aux dynamiques de mortalité des anciennes économies socialistes d'Europe durant leur période de transition, les changements de la mortalité en Chine depuis le début des réformes ont été peu étudiés. Cet article analyse les tendances de la mortalité des résidents permanents de Shanghai au moment où la Chine est passée du socialisme planifié au capitalisme d'État. L'amélioration constante de l'espérance de vie a connu un ralentissement entre 1992 et 1996. La mortalité des jeunes hommes adultes d'âge actif (20 à 44 ans) a augmenté, principalement à cause d'une progression des maladies cardiovasculaires et des morts violentes. Les décès dus aux transports ont connu une hausse, tandis que les suicides et les maladies du foie sont restés stables. Cette détérioration de la mortalité est liée aux changements structurels, parmi lesquels une détérioration de la protection sociale, une progression du chômage et du stress, ainsi que des transports et un environnement dégradés. Cependant, les effets défavorables sur la mortalité ont été atténués grâce à une stratégie de réformes graduelles, des institutions fortes et une croissance économique rapide. L'expérience de Shanghai peut s'avérer utile pour d'autres pays socialistes qui souhaiteraient passer à une économie de marché.



Technometrics, ISSN 0040-1706
Volume 57, number 2 (may 2015)

Improved Split-Plot and Multistratum Designs

P. 145-154

Luzia A. Trinca & Steven G. Gilmour

Abstract

Many industrial experiments involve some factors whose levels are harder to set than others. The best way to deal with these is to plan the experiment carefully as a split-plot, or more generally a multistratum, design. Several different approaches for constructing split-plot type response surface designs have been proposed in the literature since 2001, which has allowed experimenters to make better use of their resources by using more efficient designs than the classical balanced ones. One of these approaches, the stratum-by-stratum strategy has been shown to produce designs that are less efficient than locally D -optimal designs. An improved stratum-by-stratum algorithm is given, which, though more computationally intensive than the old one, makes better use of the advantages of this approach, that is, it can be used for any structure and does not depend on prior estimates of the variance components. This is shown to be almost as good as the locally optimal designs in terms of their own criteria and more robust across a range of criteria. Supplementary materials for this article are available online.

Bridge Designs for Modeling Systems With Low Noise

P. 155-163

Bradley Jones, Rachel T. Silvestrini, Douglas C. Montgomery & David M. Steinberg

Abstract

For deterministic computer simulations, Gaussian process models are a standard procedure for fitting data. These models can be used only when the study design avoids having replicated points. This characteristic is also desirable for one-dimensional projections of the design, since it may happen that one of the design factors has a strongly nonlinear effect on the response. Latin hypercube designs have uniform one-dimensional projections, but are not efficient for fitting low-order polynomials when there is a small error variance. D -optimal designs are very efficient for polynomial fitting but have substantial replication in projections. We propose a new class of designs that bridge the gap between D -optimal designs and D -optimal Latin hypercube designs. These designs guarantee a minimum distance between points in any one-dimensional projection allowing for the fit of either polynomial or Gaussian process models. Subject to this constraint they are D -optimal for a prespecified model.

Statistical Methods for Estimating the Minimum Thickness Along a Pipeline

P. 164-179

Shiyao Liu & William Q. Meeker

Abstract

Pipeline integrity is important because leaks can result in serious economic or environmental losses. Inspection information from a sample of locations along the pipeline can be used to estimate corrosion levels. The traditional parametric model method for this problem is to estimate parameters of a specified corrosion distribution and then to use these parameters to estimate the minimum thickness in a pipeline. Inferences using this method are, however, highly sensitive to the distributional assumption. Extreme value modeling provides a more robust method of estimation if a sufficient amount of data is available. For example, the block-minima method produces a more robust method to estimate the minimum thickness in a pipeline. To use the block-minima method, however, one must carefully choose

the size of the blocks to be used in the analysis. In this article, we use simulation to compare the properties of different models for estimating minimum pipeline thickness, investigate the effect of using different size blocks, and illustrate the methods using pipeline inspection data.

Statistical Methods for Degradation Data With Dynamic Covariates Information and an Application to Outdoor Weathering Data

P. 180-193

Yili Hong, Yuanyuan Duan, William Q. Meeker, Deborah L. Stanley & Xiaohong Gu

Abstract

Degradation data provide a useful resource for obtaining reliability information for some highly reliable products and systems. In addition to product/system degradation measurements, it is common nowadays to dynamically record product/system usage as well as other life-affecting environmental variables, such as load, amount of use, temperature, and humidity. We refer to these variables as dynamic covariate information. In this article, we introduce a class of models for analyzing degradation data with dynamic covariate information. We use a general path model with individual random effects to describe degradation paths and a vector time series model to describe the covariate process. Shape-restricted splines are used to estimate the effects of dynamic covariates on the degradation process. The unknown parameters in the degradation data model and the covariate process model are estimated by using maximum likelihood. We also describe algorithms for computing an estimate of the lifetime distribution induced by the proposed degradation path model. The proposed methods are illustrated with an application for predicting the life of an organic coating in a complicated dynamic environment (i.e., changing UV spectrum and intensity, temperature, and humidity). This article has supplementary material online.

The Uncertainty of Storm Season Changes: Quantifying the Uncertainty of Autocovariance Change-points

P. 194-206

Christopher F. H. Nam, John A. D. Aston, Idris A. Eckley & Rebecca Killick

Abstract

In oceanography, there is interest in determining storm season changes for logistical reasons such as equipment maintenance scheduling. In particular, there is interest in capturing the uncertainty associated with these changes in terms of the number and location of them. Such changes are associated with autocovariance changes. This article proposes a framework to quantify the uncertainty of autocovariance change-points in time series motivated by this oceanographic application. More specifically, the framework considers time series under the locally stationary wavelet (LSW) framework, deriving a joint density for scale processes in the raw wavelet periodogram. By embedding this density within a hidden Markov model (HMM) framework, we consider changepoint characteristics under this multiscale setting. Such a methodology allows us to model changepoints and their uncertainty for a wide range of models, including piecewise second-order stationary processes, for example, piecewise moving average processes.

Large-Vector Autoregression for Multilayer Spatially Correlated Time Series

P. 207-216

Rodrigue Ngueyep & Nicoleta Serban

Abstract

One of the most commonly used methods for modeling multivariate time series is the vector autoregressive model (VAR). VAR is generally used to identify lead, lag, and contemporaneous relationships describing Granger causality within and between time series. In this article, we investigate the VAR methodology for analyzing data consisting of multilayer time series that are spatially interdependent. When modeling VAR relationships for such data, the dependence between time series is both a curse and a blessing. The former because it requires modeling the between time-series correlation or the contemporaneous relationships which may be challenging when using likelihood-based methods. The latter because the spatial correlation structure can be used to specify the lead-lag relationships within and between time series, within and between layers. To address these challenges, we propose an $L_1 \setminus L_2$ regularized likelihood estimation method. The lead, lag, and contemporaneous relationships are estimated using an efficient

algorithm that exploits sparsity in the VAR structure, accounts for the spatial dependence, and models the error dependence. We consider a case study to illustrate the applicability of our method. In the supplementary materials available online, we assess the performance of the proposed VAR model and compare it with existing methods within a simulation study.

Bayesian Estimation of Density via Multiple Sequential Inversions of Two-Dimensional Images With Application to Electron Microscopy

P. 217-233

Dalia Chakrabarty, Nare Gabrielyan, Fabio Rigat, Richard Beanland & Shashi Paul

Abstract

We present a new Bayesian methodology to learn the unknown material density of a given sample by inverting its two-dimensional images that are taken with a scanning electron microscope. An image results from a sequence of projections of the convolution of the density function with the unknown microscopy correction function that we also learn from the data; thus, learning of the unknowns demands multiple inversions. We invoke a novel design of experiment, involving imaging at multiple values of the parameter that controls the subsurface depth from which information about the density structure is carried, to result in the image. Real-life material density functions are characterized by high-density contrasts and are highly discontinuous, implying that they exhibit correlation structures that do not vary smoothly. In the absence of training data, modeling such correlation structures of real material density functions is not possible. So we discretize the material sample and treat values of the density function at chosen locations inside it as independent and distribution-free parameters. Resolution of the available image dictates the discretization length of the model; three models pertaining to distinct resolution classes (at micrometer to nanometer scale lengths) are developed. We develop priors on the material density, such that these priors adapt to the sparsity inherent in the density function. The likelihood is defined in terms of the distance between the convolution of the unknown functions and the image data. The posterior probability density of the unknowns given the data is expressed using the developed priors on the density and priors on the microscopy correction function as elicited from the microscopy literature. We achieve posterior samples using an adaptive Metropolis-within-Gibbs inference scheme. The method is applied to learn the material density of a three-dimensional sample of a nano-structure, using real image data. Illustrations on simulated image data of alloy samples are also included.

Robust Optimization of Biological Protocols

P. 234-244

Patrick Flaherty & Ronald W. Davis

Abstract

When conducting high-throughput biological experiments, it is often necessary to develop a protocol that is both inexpensive and robust. Standard approaches are either not cost-effective or arrive at an optimized protocol that is sensitive to experimental variations. Here, we describe a novel approach that directly minimizes the cost of the protocol while ensuring the protocol is robust to experimental variation. Our approach uses a risk-averse conditional value-at-risk criterion in a robust parameter design framework. We demonstrate this approach on a polymerase chain reaction protocol and show that our improved protocol is less expensive than the standard protocol and more robust than a protocol optimized without consideration of experimental variation.

CUSUM Statistical Monitoring of M/M/1 Queues and Extensions

P. 245-256

Nan Chen & Shiyu Zhou

Abstract

Many production and service systems can be modeled as queueing systems. Their operational efficiency and performance are often measured using queueing performance metrics (QPMs), such as average cycle time, average waiting length, and throughput rate. These metrics need to be quantitatively evaluated and monitored in real time to continuously improve the system performance. However, QPMs are often highly stochastic, and hence are difficult to monitor using existing methods. In this article, we propose the cumulative sum (CUSUM) schemes to efficiently

monitor the performance of typical queueing systems based on different sampling schemes. We use M/M/1 queues to illustrate how to design the CUSUM chart and compare their performance with several alternative methods. We demonstrate that the performance of CUSUM is superior, responding faster to many shift patterns through extensive numerical studies. We also briefly discuss the extensions of CUSUM charts to more general queues, such as M/G/1, G/G/1, or M/M/c queues. We use case studies to demonstrate the applications of our approach. Supplementary materials for this article are available online.

Engineering-Driven Statistical Adjustment and Calibration

P. 257-267

V. Roshan Joseph & Huan Yan

Abstract

Engineering model development involves several simplifying assumptions for the purpose of mathematical tractability, which are often not realistic in practice. This leads to discrepancies in the model predictions. A commonly used statistical approach to overcome this problem is to build a statistical model for the discrepancies between the engineering model and observed data. In contrast, an engineering approach would be to find the causes of discrepancy and fix the engineering model using first principles. However, the engineering approach is time consuming, whereas the statistical approach is fast. The drawback of the statistical approach is that it treats the engineering model as a black box and therefore, the statistically adjusted models lack physical interpretability. This article proposes a new framework for model calibration and statistical adjustment. It tries to open up the black box using simple main effects analysis and graphical plots and introduces statistical models inside the engineering model. This approach leads to simpler adjustment models that are physically more interpretable. The approach is illustrated using a model for predicting the cutting forces in a laser-assisted mechanical micro-machining process. This article has supplementary material online.

Cook's Distance Measures for Varying Coefficient Models With Functional Responses

P. 268-280

Qibing Gao, Mihye Ahn & Hongtu Zhu

Abstract

The aim of this article is to develop Cook's distance measures for assessing the influence of both atypical curves and observations under varying coefficient model with functional responses. Our Cook's distance measures include Cook's distances for deleting multiple curves and for deleting multiple grid points, and their scaled Cook's distances. We systematically investigate some theoretical properties of these diagnostic measures. Simulation studies are conducted to evaluate the finite sample properties of these Cook's distances under different scenarios. A real diffusion tensor tract dataset is analyzed to illustrate the use of our diagnostic measures.

The Mahalanobis Distance for Functional Data With Applications to Classification

P. 281-291

Pedro Galeano, Esdras Joseph & Rosa E. Lillo

Abstract

This article presents a new semidistance for functional observations that generalizes the Mahalanobis distance for multivariate datasets. The main characteristics of the functional Mahalanobis semidistance are shown. To illustrate the applicability of this measure of proximity between functional observations, new versions of several well-known functional classification procedures are developed using the functional Mahalanobis semidistance. A Monte Carlo study and the analysis of two real examples indicate that the classification methods used in conjunction with the functional Mahalanobis semidistance give better results than other well-known functional classification procedures. This article has supplementary material online.



Technometrics, ISSN 0040-1706
Volume 57, number 4 (November 2015)

General Blending Models for Data From Mixture Experiments

P. 449-456

L. Brown, A. N. Donev & A. C. Bissett

Abstract

We propose a new class of models providing a powerful unification and extension of existing statistical methodology for analysis of data obtained in mixture experiments. These models, which integrate models proposed by Scheffé and Becker, extend considerably the range of mixture component effects that may be described. They become complex when the studied phenomenon requires it, but remain simple whenever possible. This article has supplementary material online.

Stochastic Polynomial Interpolation for Uncertainty Quantification With Computer Experiments

P. 457-467

Matthias Hwai Yong Tan

Abstract

Multivariate polynomials are increasingly being used to construct emulators of computer models for uncertainty quantification. For deterministic computer codes, interpolating polynomial metamodels should be used instead of noninterpolating ones for logical consistency and prediction accuracy. However, available methods for constructing interpolating polynomials only provide point predictions. There is no known method that can provide probabilistic statements about the interpolation error. Furthermore, there are few alternatives to grid designs and sparse grids for constructing multivariate interpolating polynomials. A significant disadvantage of these designs is the large gaps between allowable design sizes. This article proposes a stochastic interpolating polynomial (SIP) that seeks to overcome the problems discussed above. A Bayesian approach in which interpolation uncertainty is quantified probabilistically through the posterior distribution of the output is employed. This allows assessment of the effect of interpolation uncertainty on estimation of quantities of interest based on the metamodel. A class of transformed space-filling design and a sequential design approach are proposed to efficiently construct the SIP with any desired number of runs. Simulations demonstrate that the SIP can outperform Gaussian process (GP) emulators. This article has supplementary material online.

Robust Parameter Design With Computer Experiments Using Orthonormal Polynomials

P. 468-478

Matthias Hwai Yong Tan

Abstract

Robust parameter design with computer experiments is becoming increasingly important for product design. Existing methodologies for this problem are mostly for finding optimal control factor settings. However, in some cases, the objective of the experimenter may be to understand how the noise and control factors contribute to variation in the response. The functional analysis of variance (ANOVA) and variance decompositions of the response, in addition to the mean and variance models, help achieve this objective. Estimation of these quantities is not easy and few methods are able to quantify the estimation uncertainty. In this article, we show that the use of an orthonormal polynomial model of

the simulator leads to simple formulas for functional ANOVA and variance decompositions, and the mean and variance models. We show that estimation uncertainty can be taken into account in a simple way by first fitting a Gaussian process model to experiment data and then approximating it with the orthonormal polynomial model. This leads to a joint normal distribution for the polynomial coefficients that quantifies estimation uncertainty. Supplementary materials for this article are available online.

Optimal Sliced Latin Hypercube Designs

P. 479-487

Shan Ba, William R. Myers & William A. Breneman

Abstract

Sliced Latin hypercube designs (SLHDs) have important applications in designing computer experiments with continuous and categorical factors. However, a randomly generated SLHD can be poor in terms of space-filling, and based on the existing construction method that generates the SLHD column by column using sliced permutation matrices, it is also difficult to search for the optimal SLHD. In this article, we develop a new construction approach that first generates the small Latin hypercube design in each slice and then arranges them together to form the SLHD. The new approach is intuitive and can be easily adapted to generate orthogonal SLHDs and orthogonal array-based SLHDs. More importantly, it enables us to develop general algorithms that can search for the optimal SLHD efficiently.

Constructing General Orthogonal Fractional Factorial Split-Plot Designs

P. 488-502

Bagus Sartono, Peter Goos & Eric Schoen

Abstract

While the orthogonal design of split-plot fractional factorial experiments has received much attention already, there are still major voids in the literature. First, designs with one or more factors acting at more than two levels have not yet been considered. Second, published work on nonregular fractional factorial split-plot designs was either based only on Plackett–Burman designs, or on small nonregular designs with limited numbers of factors. In this article, we present a novel approach to designing general orthogonal fractional factorial split-plot designs. One key feature of our approach is that it can be used to construct two-level designs as well as designs involving one or more factors with more than two levels. Moreover, the approach can be used to create two-level designs that match or outperform alternative designs in the literature, and to create two-level designs that cannot be constructed using existing methodology. Our new approach involves the use of integer linear programming and mixed integer linear programming, and, for large design problems, it combines integer linear programming with variable neighborhood search. We demonstrate the usefulness of our approach by constructing two-level split-plot designs of 16–96 runs, an 81-run three-level split-plot design, and a 48-run mixed-level split-plot design. Supplementary materials for this article are available online.

The Spatial LASSO With Applications to Unmixing Hyperspectral Biomedical Images

P. 503-513

Daniel V. Samarov, Jeeseong Hwang & Maritoni Litorja

Abstract

Hyperspectral imaging (HSI) is a spectroscopic method that uses densely sampled measurements along the electromagnetic spectrum to identify the unique molecular composition of an object. Traditionally HSI has been associated with remote sensing-type applications, but recently has found increased use in biomedicine, from investigations at the cellular to the tissue level. One of the main challenges in the analysis of HSI is estimating the proportions, also called abundance fractions of each of the molecular signatures. While there is great promise for HSI in the area of biomedicine, large variability in the measurements and artifacts related to the instrumentation has slow adoption into more widespread practice. In this article, we propose a novel regularization and variable selection method called the spatial LASSO (SPLASSO). The SPLASSO incorporates spatial information via a graph Laplacian-based penalty to help improve the model estimation process for multivariate response data. We show the strong performance of this approach on a benchmark HSI dataset with considerable improvement in predictive accuracy over

the standard LASSO. Supplementary materials for this article are available online.

Informative Sensor and Feature Selection via Hierarchical Nonnegative Garrote

P. 514-523

Kamran Paynabar, Judy Jin & Matthew P. Reed

Abstract

Placing sensors in every station of a process or every element of a system to monitor its state or performance is usually too expensive or physically impossible. Therefore, a systematic method is needed to select important sensing variables. The method should not only be capable of identifying important sensors/signals among multistream signals from a distributed sensing system, but should also be able to extract a small set of interpretable features from the high-dimensional vector of a selected signal. For this purpose, we develop a new hierarchical regularization approach called hierarchical nonnegative garrote (NNG). At the first level of hierarchy, a group NNG is used to select important signals, and at the second level, the individual features within each signal are selected using a modified version of NNG that possesses good properties for the estimated coefficients. Performance of the proposed method is evaluated and compared with other existing methods through Monte Carlo simulation. A case study is conducted to demonstrate the proposed methodology that can be applied to develop a predictive model for the assessment of vehicle design comfort based on the tested drivers' motion trajectory signals. This article has supplementary material online.

Matrix Discriminant Analysis With Application to Colorimetric Sensor Array Data

P. 524-534

Wenxuan Zhong & Kenneth S. Suslick

Abstract

With the rapid development of nano-technology, a "colorimetric sensor array" (CSA) that is referred to as an optical electronic nose has been developed for the identification of toxicants. Unlike traditional sensors that rely on a single chemical interaction, CSA can measure multiple chemical interactions by using chemo-responsive dyes. The color changes of the chemo-responsive dyes are recorded before and after exposure to toxicants and serve as a template for classification. The color changes are digitalized in the form of a matrix with rows representing dye effects and columns representing the spectrum of colors. Thus, matrix-classification methods are highly desirable. In this article, we develop a novel classification method, matrix discriminant analysis (MDA), which is a generalization of linear discriminant analysis (LDA) for the data in matrix form. By incorporating the intrinsic matrix-structure of the data in discriminant analysis, the proposed method can improve CSA's sensitivity and more importantly, specificity. A penalized MDA method, PMDA, is also introduced to further incorporate sparsity structure in discriminant function. Numerical studies suggest that the proposed MDA and PMDA methods outperform LDA and other competing discriminant methods for matrix predictors. The asymptotic consistency of MDA is also established. R code and data are available online as supplementary material.

Malware Detection Using Nonparametric Bayesian Clustering and Classification Techniques

P. 535-546

Yimin Kao, Brian Reich, Curtis Storlie & Blake Anderson

Abstract

Computer security requires statistical methods to quickly and accurately flag malicious programs. This article proposes a nonparametric Bayesian approach for classifying programs as benign or malicious and simultaneously clustering malicious programs. The analysis is based on the dynamic trace (DT) of instructions under the first-order Markov assumption. Each row of the trace's transition matrix is modeled using the Dirichlet process mixture (DPM) model. The DPM model clusters programs within each class (malicious or benign), and produces the posterior probability of being a malware which is used for classification. The novelty of the model is using this clustering algorithm to improve the classification accuracy. The simulation study shows that the DPM model outperforms the elastic net logistic (ENL) regression and the support vector machine (SVM) in classification performance under most of the scenarios, and also outperforms the spectral clustering method for grouping similar malware. In an analysis of real malicious and benign

programs, the DPM model gives significantly better classification performance than the ENL model, and competitive results to the SVM. More importantly, the DPM model identifies clusters of programs during the classification procedure which is useful for reverse engineering.

Confidence Regions and Intervals for Meta-Analysis Model Parameters

P. 547-558

Andrew L. Rukhin

Abstract

This article obtains confidence regions for the heteroscedastic, one-way random effects model's parameters the heteroscedastic, one-way random effects model. The confidence regions are based on canonical representations of the restricted and profile likelihood functions in terms of independent normal random variables and χ^2 random variables. These regions provide conservative confidence intervals for the common mean and heterogeneity variance. Mathematical details and the R code are available online as supplementary material.

An Exact Confidence Set for a Maximum Point of a Univariate Polynomial Function in a Given Interval

P. 559-565

Fang Wan, Wei Liu, Yang Han & Frank Bretz

Abstract

Construction of a confidence set for a maximum point of a function is an important statistical problem which has many applications. In this article, an exact $1 - \alpha$ confidence set is provided for a maximum point of a univariate polynomial function in a given interval. It is shown how the construction method can readily be applied to many parametric and semiparametric regression models involving a univariate polynomial function. Examples are given to illustrate this confidence set and to demonstrate that it can be substantially narrower and so better than the only other confidence set available in the statistical literature that guarantees $1 - \alpha$ confidence level.

A Nonparametric Kernel Approach to Interval-Valued Data Analysis

P. 566-575

Yongho Jeon, Jeongyoun Ahn & Cheolwoo Park

Abstract

This article concerns datasets in which variables are in the form of intervals, which are obtained by aggregating information about variables from a larger dataset. We propose to view the observed set of hyper-rectangles as an empirical histogram, and to use a Gaussian kernel type estimator to approximate its underlying distribution in a nonparametric way. We apply this idea to both univariate density estimation and regression problems. Unlike many existing methods used in regression analysis, the proposed method can estimate the conditional distribution of the response variable for any given set of predictors even when some of them are not interval-valued. Empirical studies show that the proposed approach has a great flexibility in various scenarios with complex relationships between the location and width of intervals of the response and predictor variables.

A Semiparametric Software Reliability Model for Analysis of a Bug-Database With Multiple Defect Types

P. 576-585

Vignesh T. Subrahmaniam, Anup Dewanji & Bimal K. Roy

Abstract

Software bug-databases provide an important source of data for assessing the reliability of a software product after its release. Statistical analysis of these databases can be challenging when software usage is unknown, that is, there is no information about the usage, either in the form of a parametric model, or in the form of actual measurements. Reliability metrics, when defined on a calendar time scale, would depend on this unknown and time-dependent usage of the software and hence cannot be estimated. This article proposes a semiparametric analysis that makes use of defect classifications into multiple types to enable estimation of a model without making strict assumptions about the

underlying usage of the software. New reliability metrics whose computation does not depend on the unknown usage of the software have been proposed and methods for estimating them have been developed. The proposed method has been illustrated using data retrieved from the bug-database of a popular scripting language, named Python. This illustration compares reliability of two versions of the software without making assumptions about their unknown usage. This article has supplementary material online.
