

ISBN 978-987-688-265-1



XLVI COLOQUIO ARGENTINO DE
Estadística
4º Jornada de Educación



**Saber y
saber hacer
con estadística**

Alfredo Baronio, Silvia Cabrera y Nancy Scattolini

Compiladores

Resúmenes - Conferencias - Talleres - Ponencias - Pósters

Universidad Nacional de Río Cuarto

Facultad de Ciencias Económicas

31 de julio al 3 de agosto de 2018

Río Cuarto, Córdoba, Argentina

UniRío
editora

Saber y saber hacer con estadística : XLVI Coloquio Argentino de Estadística-XLVI CAE y 4ta Jornada de Educación Estadística Martha Aliaga-IV JEE / Alfredo Baronio ... [et al.] ; compilado por Alfredo Baronio ... [et al.]. - 1a ed. - Río Cuarto : UniRío Editora, 2018.

Libro digital, PDF - (Actas de congresos)

Archivo Digital: descarga y online

ISBN 978-987-688-265-1

1. Análisis Estadístico. 2. Ciencias Económicas. I. Baronio, Alfredo II. Baronio, Alfredo, comp.
CDD 310

Saber y saber hacer con estadística: XLVI Coloquio Argentino de Estadística-XLVI CAE y 4ta Jornada de Educación Estadística

Alfredo Baronio, Silvia Cabrera y Nancy Scattolini (Compiladores)

2018 © UniRío editora. Universidad Nacional de Río Cuarto
Ruta Nacional 36 km 601 – (X5804) Río Cuarto – Argentina
Tel.: 54 (358) 467 6309 – Fax.: 54 (358) 468 0280
editorial@rec.unrc.edu.ar
www.unrc.edu.ar/unrc/comunicacion/editorial/

ISBN 978-987-688-265-1

Primera Edición: julio de 2018

Esta obra está bajo una Licencia Creative Commons Atribución 2.5 Argentina.

http://creativecommons.org/licenses/by/2.5/ar/deed.es_AR

Consejo Editorial

Facultad de Agronomía y Veterinaria
Prof. Laura Ugnia y Prof. Mercedes Ibañez

Facultad de Ciencias Económicas
Prof. Nancy Scattolini y Prof. Silvia Cabrera

Facultad de Ciencias Exactas, Físico-Químicas
y Naturales
Prof. Sandra Miskoski y Prof. Julio Barros

Facultad de Ciencias Humanas
Prof. Gabriela Jure

Facultad de Ingeniería
Prof. Jorge Vicario

Biblioteca Central Juan Filloy
Bibl. Claudia Rodríguez y Prof. Mónica Torreta

Secretaría Académica
Prof. Ana Vogliotti y Prof. José Di Marco

Equipo Editorial

Secretaría Académica: *Ana Vogliotti*

Director: *José Di Marco*

Equipo: *José Luis Ammann, Daila Prado, Maximiliano Brito, Ana Carolina Savino
Soledad Zanatta, Lara Oviedo y Daniel Ferniot*

Contenido

Comisión Organizadora Local.....	8
Comité Científico.....	8
Unidad académica ejecutora	9
Financiamiento.....	9
Presentación	10
Conferencias.....	11
BUENAS PRÁCTICAS EN LAS ESTADÍSTICAS OFICIALES.....	12
PROCESOS MARKOVIANOS. SU NECESIDAD EN EL ANÁLISIS DE IMÁGENES DE TELEDETECCIÓN	13
ALGUNAS DE LAS PROBLEMÁTICAS DEL APRENDIZAJE EN ESTADÍSTICA. IMPLICANCIAS EN LA ENSEÑANZA Y EN LA INVESTIGACIÓN	13
SABER Y SABER HACER CON ESTADÍSTICA. <i>PIE</i>	18
CURSOS	23
INTRODUCCIÓN AL ANÁLISIS BAYESIANO CON APLICACIONES EN STAN	24
MODELOS GRÁFICOS GAUSSIANOS	25
TALLERES.....	28
DISEÑO DE PROBLEMAS PARA EL DESARROLLO DEL PENSAMIENTO ESTADÍSTICO EN LOS ALUMNOS.....	29
EL USO DEL DATO ESTADÍSTICO EN SALUD.....	33
RE-PENSAR LA ENSEÑANZA DE LA ESTADÍSTICA EN LA UNIVERSIDAD: APRENDIZAJE SIGNIFICATIVO, CONECTIVIDAD Y UBICUIDAD	36
Comunicaciones Orales.....	39
Control de procesos	40
ALGUNAS TENDENCIAS ACTUALES EN EL DESARROLLO DE METODOLOGIA ESTADISTICA PARA LA MEJORA CONTINUA.....	41
Datos de duración. Datos faltantes.....	44
ANÁLISIS DE SOBREVIDA MEDIANTE EL SOFTWARE R, APLICACIÓN A PACIENTES CON CÁNCER DE PÁNCREAS.....	45
Estadística espacial.....	47
COMPARACION DE MODELOS GEOESTADISTICOS: UNA APLICACIÓN AL ESTUDIO DE CONTAMINACION MAGNETICA	48
Inferencia Estadística	51
EVALUANDO A LOS ALUMNOS A TRAVES DE INDICADORES DE PENSAMIENTO ESTADÍSTICO	52
Métodos multivariados.....	55

ANÁLISIS PREDICTIVO Y EXPLORATORIO DE ABANDONO EN SECUNDARIA	56
CARACTERIZACIÓN SOCIO ECONÓMICAS Y ACADÉMICAS DE LOS ESTUDIANTES DE ANÁLISIS MATEMÁTICO I EN LA MODALIDAD A DISTANCIA-AÑO 2017- EN LA FCE DE LA UNRC.....	57
VALIDACION DE INSTRUMENTOS DE MEDICION MULTIDIMENSIONALES.....	60
Métodos robustos.....	64
ROBUSTEZ EN ENSAMBLES DE MODELOS LINEALES REGULARIZADOS	65
SM-ESTIMADORES PARA ANÁLISIS CANÓNICO.....	67
Modelos de regresión.	69
CALIBRACIÓN COMPARATIVA ROBUSTA.....	70
Muestreo.....	72
UNA PROPUESTA DE CURSO DE MUESTREO CON R	73
DETERMINACIÓN DEL TAMAÑO DE UNA MUESTRA Y ASIGNACIÓN EN PRESENCIA DE ERRORES DE CLASIFICACIÓN	74
IMPUTACIÓN DE LA NO RESPUESTA UTILIZANDO EL PROCEDIMIENTO RANDOM FOREST	77
DISEÑO MUESTRAL PARA UNA MUESTRA MAESTRA DEL ÁREA RURAL: (MMR- INDEC)	78
Probabilidad y Procesos estocásticos	81
METODOLOGÍAS ESTOCÁSTICAS PARA LA MEDICIÓN DEL RIESGO DE MERCADO EN ORGANIZACIONES BANCARIAS	82
Teoría sobre distribuciones de probabilidad	84
APLICACIÓN DE UNA CÓPULA MIXTA PARA LA ESTIMACIÓN DE LA PÉRDIDA TOTAL DE UNA CARTERA DE SEGUROS VEHICULARES.....	85
ANÁLISIS DE CORRELACIONES CONDICIONALES DINÁMICAS APLICADO AL ESTUDIO DE LA INTERRELACIÓN ENTRE MERCADOS BURSÁTILES.....	88
VARIACIÓN REGULAR EN VARIAS VARIABLES: APLICACIÓN A TEOREMAS DE CONVERGENCIA	89
Otras Categorías Metodológicas.....	92
LA EVALUACIÓN EN ESTADÍSTICA EN CARRERAS DE OTRAS DISCIPLINAS	93
ASOCIACION ESTADISTICA: ERRORES EN CONTENIDOS CONCEPTUALES Y PROCEDIMENTALES EN ESTUDIANTES UNIVERSITARIOS.....	96
ALGUNAS ACTIVIDADES PARA ENSEÑAR CONCEPTOS ESTADÍSTICOS.....	100
LA RELEVANCIA VALORATIVA DE LOS INTANGIBLES EN EL MERCADO DE CAPITALES ARGENTINO. ESTUDIO EMPIRICO APLICANDO UN MODELO PARA DATOS DE PANEL DE EFECTOS FIJOS.....	103
COMPARACIÓN DE LOS ESTILOS DE APRENDIZAJE DE LOS ALUMNOS DE LAS CARRERAS DE BIOLOGÍA Y DE MATEMÁTICA	106
MUESTREO SIMPLE Y DOBLE PARA EL MONITOREO DE UNIDADES NO CONFORMES EN PROCESOS DE ALTA CALIDAD.....	109

Posters.....	112
Biología.....	113
UN MODELO GEOESTADÍSTICO PARA ANALIZAR LA DISTRIBUCIÓN ESPACIAL DE LA FRACCIÓN COMERCIAL DE CENTOLLA (<i>Lithodes santolla</i>) EN LA PATAGONIA CENTRAL, ARGENTINA..	114
Ecología y Medio Ambiente	117
ESTUDIO DE LOS ÍNDICES DE VEGETACIÓN EN EL NORTE DE LA PROVINCIA DE SANTA CRUZ	118
Economía.....	120
COTIZADOR INMOBILIARIO CON RANDOM FOREST EN R.....	121
CONSTRUCCIÓN DE UN ÍNDICE ESPACIAL DE BANCARIZACIÓN. UNA PROPUESTA PARA LA REGIÓN CENTRO	122
ANÁLISIS DE SERIES DE LAS FINANZAS PÚBLICAS DE MENDOZA.....	124
MEDICIÓN DE LA PROPENSIÓN AL EMPRENDEDURISMO EN ESTUDIANTES DE CIENCIAS ECONÓMICAS.....	126
VARIABLES EXPLICATIVAS DE LA POBREZA EN LOS HOGARES ARGENTINOS: APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA.....	128
ACTITUD DE UNA COMUNIDAD LOCAL HACIA EL DESARROLLO DEL TURISMO MEDIANTE EL USO DE MODELOS DE ECUACIONES ESTRUCTURALES.....	134
Educación, Ciencia y Cultura	138
ADAPTACIÓN Y ANÁLISIS DE UNA ESCALA DE MEDIDA PARA ESTRATEGIAS DE APRENDIZAJE AUTÓNOMO EN EDUCACIÓN A DISTANCIA	139
ESTUDIO DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES UNIVERSITARIOS EN EL CICLO BÁSICO. EL CASO DE LAS ASIGNATURAS DEL DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA DE LA FACULTAD DE CIENCIAS ECONÓMICAS/UNRC.....	140
EVALUACIÓN DEL PERFIL DE LOS ESTUDIANTES DE UN CURSO DE GRADO DE ESTADÍSTICA	145
Enseñanza de la Estadística.....	147
COMPRESION DE INFORMACIÓN CUALI-CUANTITATIVA ASIGNATURA DEL INGRESO UNIVERSITARIO. UNA MATERIA INTERDISCIPLINARIA.....	148
ESTADÍSTICA DESCRIPTIVA. IMPLEMENTACIÓN DE R-COMMANDER.....	150
PAQUETE SHINY: UNA HERRAMIENTA R PARA LA ENSEÑANZA DE DISTRIBUCIONES DE PROBABILIDAD	152
Estadísticas Oficiales	154
EVALUACIÓN DEL DESARROLLO PROVINCIAL SEGÚN UN INDICADOR SINTÉTICO	155
Genética	158
ANÁLISIS FACTORIAL MÚLTIPLE PARA LA CARACTERIZACIÓN DE PLANTAS DE BANANA SEGÚN EL AMBIENTE DE CRECIMIENTO INICIAL.....	159

Industria y Mejoramiento de la Calidad.....	161
ESTUDIO DEL EFECTO DE ERRORES DE MEDIDA SOBRE EL ERROR DE PREDICCIÓN DE UN ÁRBOL DE REGRESIÓN EN UN PROBLEMA REAL.....	162
NUEVAS ESTRATEGIAS PARA LA CONSTRUCCION DE DISEÑOS 2 ^d -PLEGADOS.....	164
IMPACTO DE DISTINTOS MÉTODOS DE ESTIMACIÓN DE SCORES FRENTE A DATOS FALTANTES SOBRE EL GRÁFICO T2 DE HOTELLING CON COMPONENTES PRINCIPALES.....	167
Salud Humana	170
COMPARACIÓN DE MÚLTIPLES TRATAMIENTOS A PARTIR DE UNA REVISIÓN SISTEMÁTICA	171
DIAGNÓSTICO DEL PERFIL SOCIOECONÓMICO Y NUTRICIONAL DEL PACIENTE DIABÉTICO TIPO 2 A PARTIR DE MÉTODOS MULTIVARIANTES.	173
Ciencias Agropecuarias	175
MONITOREO DE HUERTAS Y GRANJAS DEL PARTIDO DE LUJÁN	176
ANÁLISIS DEL CONTENIDO VOLUMÉTRICO DE AGUA EN EL SUELO A PARTIR DE MÉTODOS DESCRIPTIVOS Y EXPLICATIVOS MULTIVARIANTES.....	177
ESTRATEGIAS DE ANÁLISIS PARA DATOS CATEGÓRICOS DE NATURALEZA BINARIA.....	180
ESTIMACIÓN ROBUSTA DE UN MODELO NO LINEAL MIXTO PARA EVALUAR EL CRECIMIENTO DEL MAIZ.....	182
Ciencias Exactas y Naturales	184
ESTUDIO DE LAS PROPIEDADES DISTRIBUCIONALES DE LOS ESTIMADORES LASSO	185
APLICACIÓN EN R PARA LA DETECCIÓN DE ASOCIACIÓN ENTRE IMÁGENES	187
Otras Ciencias de la Salud	189
MODELOS PARA DATOS DE CONTEO CON EXCESO DE CEROS: UNA APLICACIÓN SOBRE EL MAL DE CHAGAS	190
Otras Ciencias Económicas, Administración y Negocios.....	192
ANÁLISIS DE LA SATISFACCIÓN MEDIANTE ECUACIONES ESTRUCTURALES. EL CASO DE LOS MATRICULADOS EN EL CONSEJO PROFESIONAL DE CIENCIAS ECONÓMICAS DE LA PROVINCIA DE CÓRDOBA.....	193
Otras Ciencias Sociales y Humanas	195
TRIANGULACIÓN METODOLOGICA PARA EVALUAR LA PRESTACIÓN DE UN SERVICIO.....	196
ESTUDIO COMPARATIVO DE LOS PATRONES DE TRANSICIÓN A LA ADULTEZ.....	197
ESTUDIO DE LA INCIDENCIA DE FACTORES DEMOGRÁFICOS SOBRE LA INSERCIÓN LABORAL Y LA VIDA EN PAREJA DE LAS JÓVENES CORDOBESAS.....	199
Otras Aplicaciones.....	202

FACTORES DE RIESGOS DE ENFERMEDAD CARDIOVASCULAR EN PACIENTES CON DIABETES Y SU REPERCUSIÓN EN SUS ÓRGANOS BLANCO EN LA OBRA SOCIAL DE DOCENTES PARTICULARES (OSDOP).....	203
ESTUDIO COMPARATIVO DE LOS ALGORITMOS EM Y VB EN PROBLEMAS DE APRENDIZAJE AUTOMÁTICO DESARROLLADOS EN PYTHON.....	204
CONSTRUCCIÓN DE UN ÍNDICE DE DESIGUALDAD DE GÉNERO: APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES.	206
COMPARACIÓN DE METODOLOGÍAS PARA LA CONSTRUCCIÓN DE MAPAS DE PRECIPITACIONES EN LA PROVINCIA DE SANTA CRUZ.....	212

Comisión Organizadora Local

Presidente

Dr. Alfredo Baronio.

Integrantes

Cra. Silvia Butigué.

Lic. Favio Nicolás D'ércole.

Cr. Juan Manuel Gallardo.

Cra. Cecilia Ficco.

Cra. Gabriela García.

Lic. Juan Munt.

Lic. Ana Vianco.

Colaboradores

Docentes del Departamento de Matemática y Estadística de la FCE-UNRC.

Investigadores del Programa Producción de Datos y Econometría Aplicada.

SECyT UNRC.

Estudiantes de la FCE-UNRC.

Personal No Docente de la FCE-UNRC.

Comité Científico

Baronio, Alfredo Mario. (UNRC)

Blacona, María Teresa. (UNR)

Casparri, María Teresa. (UBA)

Cuesta, Cristina. (UNR)

Curti, Celina. (UNTREF)

Damilano, Gabriela L. (UNRC)

Filippini, Susana. (UNLU)

Herrera, María Inés. (UNRC)

Iñiguez, Patricia Alejandra. (UNRC)

Ivars, Daniel. (UNRC)

Kelmansky, Diana. (UBA)

Maldonado, Rita Ester. (UNRC)

Rabanal, Cristian. (UNViMe)

Vianco, Ana María. (UNRC)

Unidad académica ejecutora

Secretaría de Ciencia y Técnica, Secretaría de Extensión, Secretaría Académica, Dirección de Educación a Distancia y Departamento de Matemática y Estadística de la Facultad de Ciencias Económicas. Universidad Nacional de Río Cuarto. El encuentro será coorganizado con la Sociedad Argentina de Estadística.

Financiamiento

Ministerio de Ciencia y Tecnología de la Provincia de Córdoba. Programa de Apoyo a Eventos Científicos.

Secretaría de Ciencia y Técnica Universidad Nacional de Río Cuarto. Programa de Investigación Producción de Datos y Econometría Aplicada.

Presentación

La Sociedad Argentina de Estadística (SAE) es un organismo técnico – científico que promueve el desarrollo de la Estadística en el país. Fue creado en junio de 1952 y entre sus objetivos se encuentran: agrupar y vincular entre sí a todas las personas relacionadas con la Ciencia Estadística y con sus aplicaciones; fomentar y promover el estudio, investigación, desarrollo y perfeccionamiento de la Estadística y disciplinas conexas, mediante reuniones científicas y técnicas, cursos, becas, premios, concursos, publicaciones sin fines de lucro y otros medios adecuados; contribuir al mejoramiento de la enseñanza de la Estadística y mantener relaciones con instituciones afines, nacionales y extranjeras. Cada año, la SAE organiza Coloquios que intercaladamente comparte con otros países de Latinoamérica.

La FCE desarrolla, anualmente, las Jornadas de Intercambio de Conocimientos Científicos y Técnicos sustentadas en la idea de que la Ciencia es el resultado de un esfuerzo mancomunado de personas que comparten un interés común y una forma común de hacer. Específicamente, y complementando las concepciones de Thomas Kuhn y de Karl Popper, se ha considerado a la Ciencia como el resultado de una Comunidad que acuerda un método de expandir el conocimiento y coincide acerca de los temas a tratar.

Coincidentes con la tradición de la FCE, las Jornadas de Educación Estadística y el Coloquio Argentino de Estadística se plantean como un espacio para la socialización de los conocimientos, tanto de carácter científico como tecnológico, producidos a través de la investigación o el trabajo profesional; como así también, para el intercambio y debate referidos a aspectos metodológicos y a temáticas relativas a diferentes disciplinas. Con el objetivo de disponer un espacio para el intercambio de conocimientos y experiencias del saber y el saber hacer con Estadística, este encuentro se destina a investigadores, docentes, profesionales y estudiantes de las diversas disciplinas que desarrollan teoría y construyen o utilizan estadísticas para comunicar el resultado de su trabajo.

Este compendio reúne las propuestas presentadas, bajo el formato de cursos, talleres y comunicaciones orales y digitales. Las categorías metodológicas Métodos Multivariados, Modelos de Regresión, Control de Procesos; Estadística Espacial y Muestreo están presentes en el 56,25% de los casos; siendo las aplicaciones a Enseñanza de la Estadística, Ciencias Exactas y Naturales, Economía, Industria y Mejoramiento de la Calidad y Educación Ciencia y Cultura el 67,19% del total.

La Universidad Nacional de Río Cuarto, y la Facultad de Ciencias Económicas en particular, se complacen en compartir estas jornadas dedicadas al estudio y la difusión de aplicaciones de la estadística en las diversas disciplinas; donde, el Saber y Saber hacer con Estadística están presentes en cada una de las acciones.

CONFERENCIAS

BUENAS PRÁCTICAS EN LAS ESTADÍSTICAS OFICIALES

Lic. Hernán Muñoz

Director Nacional de Planificación, Relaciones Institucionales e Internacionales

Instituto Nacional de Estadística y Censos (INDEC)

hmunoz@indec.gob.ar

Resumen

La exposición se centrará en los principios fundamentales y las buenas prácticas recomendadas por los organismos internacionales para las Oficinas Nacionales de Estadística.

En particular, se hará foco en los Principios Fundamentales de las Estadísticas Oficiales (Naciones Unidas) y en las recomendaciones del Consejo de la Organización para la Cooperación y el Desarrollo Económico (OCDE) sobre Buenas Prácticas Estadísticas, a las que desde el año 2016 el INDEC ha adherido.

Asimismo, se hará referencia también a otras normas de buenas prácticas estadísticas, tales como los Estándares Especiales de Diseminación de Datos (Fondo Monetario Internacional), el Código de Buenas Prácticas de las Estadísticas Europeas (EUROSTAT) y el Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe (CEPAL).

Se analizarán las ventajas que representan para las Oficinas Nacionales de Estadística la aplicación de este conjunto de buenas prácticas, tanto en su arquitectura organizacional, así como para la mejora en la calidad de los indicadores estadísticos producidos.

Referencias bibliográficas

Comisión Económica para América Latina y el Caribe (2011). Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe.

EUROSTAT (2011). Código de Buenas Prácticas de las Estadísticas Europeas.

Fondo Monetario Internacional (2013). The special data dissemination standard: guide for subscribers and users.

Naciones Unidas (2014). Principios Fundamentales de las Estadísticas Oficiales.

Organización para la Cooperación y el Desarrollo Económico (2015). Recomendación del Consejo de la OCDE sobre Buenas Prácticas Estadísticas.

Palabras Clave: Buenas prácticas estadísticas – Legislación - Calidad – Transparencia
– Objetividad – Independencia funcional

Área de Aplicación: Estadísticas Oficiales

PROCESOS MARKOVIANOS. SU NECESIDAD EN EL ANÁLISIS DE IMÁGENES DE TELEDETECCIÓN

Oscar Bustos

FaMAF y CIEM - UNC

oscar.oh@gmail.com

Resumen

¿Qué es una imagen de teledetección espacial? Breve descripción de su formación. Exhibición por medio de un software especializado como ENVI-IDL.

Clasificación en imágenes de teledetección. Métodos habitualmente implementados en softwares. Métodos supervisados y no supervisados.

Información de contexto. Necesidad de tenerla en cuenta para lograr clasificaciones confiables.

Procesos markovianos en general.

El método ICM aplicado a la clasificación en imágenes.

Palabras Clave: Imágenes – Clasificación – Procesos markovianos.

ALGUNAS DE LAS PROBLEMÁTICAS DEL APRENDIZAJE EN ESTADÍSTICA. IMPLICANCIAS EN LA ENSEÑANZA Y EN LA INVESTIGACIÓN

Osmar Darío Vera

Universidad Nacional de Quilmes

overa17@gmail.com

Resumen

La estadística, la probabilidad y la combinatoria son hoy parte de los Diseños Curriculares de matemáticas para las clases de la escuela primaria y secundaria en muchos países, el nuestro no está excluido. Aunque esto ocurre, la capacitación para

enseñar estadísticas dista mucho de ser un componente universal en cursos de desarrollo profesional tanto para los maestros, para los profesores, y aún para los licenciados en matemática; sean en formato de cursos de grado como de postgrado. Aún la oferta de cursos de formación, sigue siendo muy escasa.

Debido a su utilidad en la vida diaria, su aplicación en muchas disciplinas, la necesidad del conocimiento estocástico en muchas profesiones, su importancia en el desarrollo crítico (Batanero, 2001), es que la estadística hoy es parte de los diseños curriculares en nuestras provincias, especialmente a partir del primer año de la enseñanza secundaria. Según los NAP (Núcleos de Aprendizaje Prioritarios, 2012), este comprende un eje, llamado: *En relación con las probabilidades y la estadística* (Res CFE N° 180/12), en relación con la enseñanza de la Matemática para el nivel secundario. Aunque en países de Iberoamérica (Batanero, Arteaga y Contreras, 2011) ya aparece lo estocástico como contenido para su enseñanza desde la escuela primaria, por ejemplo: dentro del Bloque: *Tratamiento de la información, azar y probabilidad* (MEC, 2006a, España); en nuestro país casi no se encuentran al momento recomendaciones para su enseñanza, y los estudiantes finalizan la escuela secundaria con escasa comprensión de los principios básicos relacionados con el análisis de datos, lo que explica muchos de los problemas que encontramos en el uso posterior de la estadística en su vida cotidiana o profesional, o en los cursos de estadística en la Universidad (Meletiou, 2003).

El papel de los profesores es esencial al momento de interpretar y adaptar los diseños (Ponte, 2001). En consecuencia, el cambio e incorporación en la enseñanza de la estadística en las escuelas dependerá del paradigma que nos podamos plantear para convencer a los profesores de que la estadística es uno de los temas más interesantes y útiles para nuestros estudiantes y que, todos ellos tienen capacidad para adquirir algunos conceptos elementales, dependiendo de su edad evolutiva.

Por otro lado, las actitudes, concepciones y creencias de los profesores hacia la estadística es un punto de atención para tener en cuenta, puesto que este pensamiento incidirá en las creencias y conocimiento de los estudiantes sobre la Matemática. Será un factor clave en cualquier movimiento para cambiar la enseñanza de las matemáticas (Chapman, 1999). Las actitudes se consideran en general un constructo mental, por ende no es directamente observable. Esta se infiere a partir de alguna escala

de valores que está relacionada con la observación de los sujetos en cuestión. Gal, Ginsburg y Garfield (1997) las definen como “una suma de emociones y sentimientos que se experimentan durante el período de aprendizaje de la materia objeto de estudio” (p. 40). Hay muy pocos trabajos sobre las actitudes que se centren en los profesores de estadística, o en la estadística para los profesores de matemática. Podemos destacar el trabajo de Estrada y Batanero (2008) realizado en España con profesores de Educación primaria, tanto en ejercicio como en formación; este trabajo ha sido relacionado también con otras variables. Uno de los resultados ha sido que las actitudes mejoran con conocimientos y práctica profesional.

Ball, Thames y Phelps (2005) quienes hacen estudios de evaluación del conocimiento matemático de los profesores, lo dividen en varios componentes: conocimiento común del contenido, conocimiento especializado del contenido, conocimiento del contenido y la enseñanza y conocimiento del contenido y los estudiantes. Si nos centrásemos en el conocimiento común del contenido estadístico, los estudios de evaluación suelen usar cuestionarios o entrevistas y tareas similares a las empleadas en investigaciones con estudiantes. Su interés es determinar, al menos si el nivel de conocimiento de los futuros profesores es suficiente para resolver las tareas que tendrán que proponer a sus futuros alumnos. Estrada y Batanero (2008) encontraron en su estudio una proporción preocupante de futuros profesores que no dominan los conceptos elementales que han de enseñar a sus futuros alumnos. Para citar un ejemplo de la investigación, Wild y Pfannkuch (1999) han resaltado la importancia del concepto de variación en estadística como el núcleo de su modelo sobre razonamiento estadístico. A pesar de ello, la investigación de Silva y Coutinho (2008) muestra que el razonamiento predominante de los profesores brasileños sobre la variación es verbal, lo que les impide enseñar a sus estudiantes el significado de medidas tales como la desviación típica, limitándose a la enseñanza de algoritmos para su cálculo. Les resulta difícil integrar el razonamiento sobre la media, desviación respecto a la media, intervalo de k desviaciones típicas desde la media (interpretación de la desigualdad de Tchebychev) y estimación de la frecuencia en el intervalo.

También es importante contar con un conocimiento profesional para enseñar estadística, es decir el que se relaciona con la educación estadística, ya que para muchos objetos no es posible transferir directamente el conocimiento profesional para enseñar que se aplica para el profesor de matemática. La investigación de González y Pinto (2008) muestra que los estudiantes para profesor de matemática tienen pocos conocimientos didácticos en relación a los gráficos estadísticos, para citar un ejemplo de la investigación. Por otro lado, los libros de texto y materiales curriculares preparados son insuficientes, en algunos casos, como soporte para el profesor al presentar una visión muy parcial de los conceptos (por ejemplo, solo la aproximación clásica a la probabilidad o la inferencia); en otros casos las aplicaciones se limitan a juegos de azar o no se basan en datos tomados de aplicaciones reales; en otros las definiciones de los conceptos son incorrectas o incompletas (Cardeñoso, Azcárate y Serradó, 2005). Garfield y Ben-Zvi (2008) describen el conocimiento profesional mediante seis componentes que se ha de desarrollar en la formación de profesores: ideas estadísticas fundamentales; uso de datos reales; uso de actividades para el aula; integración de las herramientas tecnológicas; implementación del discurso en el aula y uso de métodos alternativos de evaluación.

Por último, y sin poder ser exhaustivos no podemos dejar de descartar el papel esencial que tiene hoy el uso de la tecnología y más aún en la estadística y para su enseñanza. Esta fue reconocida por International Association for Statistical Education (IASE) y en los congresos sobre Enseñanza de la Estadística, donde largamente se ha discutido y se discute sobre posibles software para ser utilizados. Entre otras cosas, su uso implica una notable reducción en el tiempo que hasta ahora se ocupa para el cálculo, en el diseño de gráficos tanto del profesor como del estudiante, la simulación; lo cual implica dejar un mayor espacio para la interpretación, que es un objeto importante en estadística (Olivo, 2008); además se sustituyen los razonamientos formales por otros más intuitivos (inferencia informal, por ejemplo) (Batanero y Díaz, 2015).

Referencias bibliográficas

Ball, D. L., Thames, M. H. y Phelps, G. (2005). Articulating domains of mathematical knowledge personal.umich.edu/~dball/ for teaching. Online: www.personal.umich.edu.

- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Grupo de Investigación en Educación Estadística. *Didáctica de las Matemáticas*.
- Batanero, C. (2009). Retos para la formación estadística de los profesores. *Actas do II Encontro de probabilidades e estatística na escola*, 7-21.
- Batanero, C., & Díaz, C. (2015). Aproximación informal al contraste de hipótesis. *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2, 207-214.
- Batanero, C., Arteaga, P., & Contrera, J. M. (2011). El currículo de estadística en la enseñanza obligatoria. *Em Teia| Revista de Educação Matemática e Tecnológica Iberoamericana-ISSN: 2177-9309*, 2(2).Res CFE N° 180/12.
- Cardeñoso, J. M., Azcárate, P. y Serradó, A. (2005). Los obstáculos en el aprendizaje del conocimiento probabilístico: Su incidencia desde los libros de texto. *Statistics Education Research Journal* 4(2), 59-81. On line: <http://www.stat.auckland.ac.nz/~iase/serj/>.
- Chapman, O. (1999). Researching mathematics teacher thinking. En O. Zaslavsky (Ed.), *Proceedings of the 23th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 385-392). Haifa: The Technion – Israel Institute of Technology.
- de Educación, C. F. (2012). Núcleo de Aprendizajes Prioritario. Resolución CFE.
- Estrada, A. y Batanero, C. (2008). Explaining teachers' attitudes towards statistics. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education.
- Gal, I., Ginsburg, L. y Garfield, J. B. (1997). Monitoring attitudes and beliefs in statistics education. En: I. Gal y J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37-51). Voorburg: IOS Press.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media.
- Gal, I., Ginsburg, L. y Garfield, J. B. (1997). Monitoring attitudes and beliefs in statistics education. En: I. Gal y J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37-51). Voorburg: IOS Press.

- Meletiou, M. (2003). On the formalist view of mathematics: impact on statistics instruction and learning En A. Mariotti (Ed.), Proceedings of Third European Conference in Mathematics Education. Bellaria, Italy: ERME.
- Olivo, E., Batanero, C., & Díaz, C. (2008). Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación matemática*, 20(3), 5-32.
- Ponte, J. P. (2001). Investigating in mathematics and in learning to teach mathematics. In T. J. Cooney & F. L. Lin (Eds.), *Making sense of mathematics teacher education* (pp. 53-72). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Rossmann, A., Chance, B., & Medina, E. (2006). Some key comparisons: statistics and mathematics, and why teachers should care. In G. Burrill (Ed.), *Thinking and reasoning with data and chance, 68th NCTM Yearbook (2006)* (pp. 323–334). Reston, VA: National Council of Teachers of Mathematics.
- Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 221-248.
- Palabras Clave:** Estadística y escuela media - Enseñanza de la Estadística – Conocimiento profesional profesor de Estadística.

Área de Aplicación: Enseñanza de la estadística.

SABER Y SABER HACER CON ESTADÍSTICA. PIE

Alfredo Mario Baronio y Ana María Vianco

Facultad de Ciencias Económicas – Universidad Nacional de Río Cuarto

anavianco@yahoo.com.ar

Problemática

La Estadística, la Matemática y por añadidura la Econometría son herramientas que permiten hallar respuestas a los interrogantes de investigación y no compartimentos estancos e independientes del estudio de la ciencia. El desafío para el docente es incorporar esta noción en el estudiante que le permita satisfacer la necesidad, frecuentemente expresada, de conocer qué es una investigación y, paralelamente, introducir un método para el estudio empírico porque muestra cómo emplear la estadística fundamentando el *saber y saber hacer con estadística*.

Con este fin se diseña el Proceso de Investigación Econométrica (PIE); presentado como un conjunto encadenado de pasos y métodos que se entrelazan para alcanzar la respuesta que dio origen a la investigación, es la solución a la necesidad de entender aplicaciones estadísticas integradas para resolver una problemática en particular. Se considera que esta metodología de enseñanza es válida para todos los niveles de formación y para todas las disciplinas.

Objetivo General. Compartir la estrategia de enseñanza adoptada para el dictado de Econometría en la UNRC.

Objetivos particulares: (a) Mostrar una visión general del PIE; (b) Ilustrar las etapas y los tipos de análisis que pueden utilizarse; (c) Mostrar que la estadística está inserta en un método que, paso a paso, lleva al investigador a comprobar sus teorías. (d) Aplicar la estadística para desarrollar evidencia empírica.

Metodología

El PIE comienza por definir la investigación, continúa planteando la tabla de datos, luego diseña la fuente de información, sigue con recolección, organización y procesamiento de la información, para finalmente analizarla con métodos estadísticos y econométricos. La experiencia indica que, el impartir Econometría de esta manera, se compatibiliza el enseñar para saber y saber hacer con estadística y con datos. Estas jornadas se constituyen en el ámbito propicio para que, quienes están involucrados en la formación de nuevos profesionales, se den un espacio para la discusión del qué, del cómo y del cuándo enseñar estadística. El primer error lo constituye centrarse directamente en el análisis de la información.

El proceso científico de la teoría económica encuentra en la econometría, métodos para comprobar empíricamente sus modelos. Para llevar a cabo esta tarea, se emplea el PIE que se fundamenta en la metodología de la investigación y en la aplicación de la estadística. Para el econometrista, estos fundamentos tienen en cuenta la metodología de lo general a lo específico y el proceso generador de datos, respectivamente. Para ponerlo en un contexto amplio, se sigue lo indicado por Kinneary y Taylor (1993), en el sentido de que el proceso aquí desarrollado hace uso de la estadística en cada paso de su aplicación. De esta manera, el investigador puede justificar el uso del instrumental econométrico tanto, desde la teoría estadística como desde la inferencia estadística y sus fundamentos matemáticos.

Siguiendo a Klein (1957), específicamente, se está frente a un problema econométrico cuando los métodos se llevan al nivel de reunir datos -y el trabajo empírico es adecuadamente enlazado con la teoría de la probabilidad y con la inferencia estadística-, haciendo que aquellos constituyan una muestra razonable.

Si la econometría es, fundamentalmente, operativa tiene que acercarse más a la realidad, a sus acciones. El investigador debe introducirse en los pormenores de ella para conocer cuál es su metodología específica; es decir, cómo utilizarla desde el comienzo hasta el final de su trabajo.

¿Cuál es el fundamento metodológico de este razonamiento? El econometrista actúa como un investigador y debe establecerse como tal. Hernández Sampieri, (2000), señala que el proceso de investigación está constituido por una serie de partes íntimamente relacionadas; donde el marco teórico y el análisis desempeñan un papel fundamental.

La primera etapa recibe, generalmente, el nombre de Definición de la investigación y reporta tanto los objetivos e hipótesis de investigación como el marco teórico de referencia, usado por el investigador en la formulación de su proyecto de investigación.

En una investigación económica, las fuentes de información para elaborar las hipótesis de investigación son fundamentalmente, la teoría, la investigación exploratoria y la experiencia. Además, en cuanto a la teoría habrá que considerar la teoría estadística. De esta forma, se define el método de la econometría como aquél que tiene por finalidad comprobar hipótesis y teorías formuladas por la ciencia económica. Esta comprobación se hace con el auxilio, imprescindible, de las matemáticas y de la estadística.

Según esta concepción de los modelos económicos y de la econometría, el método econométrico utiliza las técnicas de la estadística matemática, fundamentada en la teoría de la probabilidad, para describir analíticamente una situación de la economía, hacer pronósticos o analizar políticas, en espacio y tiempo específico.

Resultados

Con estos fundamentos se explican los pasos del proceso de investigación econométrica (PIE), siguiendo la metodología de lo general a lo específico sostenida por Hendry a finales del Siglo XX. Estas etapas se subdividen en partes que determinan el método econométrico. La metodología señalada contempla las pautas de la nueva econometría, incorporándola a las fases de aplicación de la metodología tradicional.

Desde este punto de vista, el proceso de investigación econométrica se basa, fundamentalmente, en el proceso generador de datos (PGD).

Conforme a lo expresado, es un proceso de investigación empírico que, en general, se ajusta a un método que sigue las siguientes etapas:

1. Definición de la investigación, una exposición lo más precisa posible de la información que se necesita para desarrollar la evidencia empírica, en espacio y tiempo específico, de acuerdo al problema de investigación de que se trate.
2. Planteamiento de una tabla de datos, unidades de observación por variables, con observaciones, temporales o espaciales (ubicadas en filas), y características o atributos observados de las mismas, que pueden ser variables cuantitativas o cualitativas, (ubicadas en columnas)
3. Diseño de fuentes de información, de muestreo para las fuentes de información primaria y pautas para la obtención de fuentes de información secundarias.
4. Recolección, procesamiento y organización de los datos, metodología para la organización de datos provenientes de fuentes de información secundaria y para la recolección y procesamiento de datos provenientes de fuentes primarias, resumiendo las principales operaciones de campo, que comporta todo proceso de observación; esto es, a partir de una muestra o de fuentes secundarias, aplicar un proceso de recolección y procesamiento de la información para medir sus características.
5. Análisis de la información, técnicas apropiadas al análisis de las variables: descriptivos, exploratorios, correlacionales, explicativos y confirmatorios y métodos para determinar la semejanza entre las unidades de observación evaluando la asociación o relación entre las características seleccionadas y observadas, temporal o espacialmente, aplicando modelos econométricos.

Referencias bibliográficas

- Hendry, David F. "Econometrics-Alchemy or Science?" The London School of Economics. *Economica, New Series*, Vol. 47, No. 188 (Nov., 1980), pp. 387-406
- Hernández Sampieri, R.; Fernández Collado, C. y Baptista Lucio, P. "Metodología De La Investigación". México: McGraw Hill, 2000.
- Kinney, T. y Taylor, J. "Investigación De Mercado. Un Enfoque Aplicado". McGraw Hill, 1993.
- Klein, L. R. "The Scope and Limitations of Econometrics." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1957, 6(1), pp. 1-17.

- Loria, E. "Econometría con Aplicaciones". México: Pearson Prentice Hall, 2007.
- Morgan, M. S. "The History of Econometric Ideas: Historical Perspectives on Modern Economics". Cambridge: Cambridge and New York: Cambridge University Press, 1990.
- Navarro, A. M. "Reflexiones sobre la relación entre Economía, Econometría y epistemología," Anales. Buenos Aires: Academia Nacional de Ciencias Económicas, 1997.
- Otero, J. M. Econometría. "Series Temporales y Predicción". Madrid: Editorial AC, 1990.
- Pulido San Román, A. "25 Años de Experiencia en Econometría Aplicada." Estudios de Economía Aplicada, 1993a, v.0, pp. 82-98.
- Samuelson, P. y Nordhaus, W. "Macroeconomía". Madrid: McGraw Hill, 2001.

Fuentes

Sitio web de los autores <http://www.econometricos.com.ar/>

Revista Econometrica, volumen I

Palabras Clave: PIE – Metodología – Econometría

CURSOS

INTRODUCCIÓN AL ANÁLISIS BAYESIANO CON APLICACIONES EN STAN

Ignacio Álvarez-Castro, Juan José Goyeneche

Instituto de Estadística

Facultad de Ciencias Económicas y de Administración

Universidad de la República, Uruguay

nachalca@iesta.edu.uy

Resumen

La inferencia Bayesiana es una importante rama de la Estadística, su principal objetivo es obtener una distribución de probabilidad que modela la incertidumbre sobre las cantidades de interés (por ejemplo, parámetros) condicional a los datos observados. Esta distribución es llamada distribución posterior (o a posteriori), y se obtiene mediante la regla de Bayes al combinar el modelo estadístico para los datos (función de verosimilitud) y la distribución previa (o a priori) que modela la incertidumbre sobre las cantidades de interés sin condicionar en los datos. La estadística Bayesiana ha ganado popularidad desde la década del 90s debido, entre otras cosas, al desarrollo de distintos métodos de aproximación a la distribución posterior, usando algoritmos de MCMC (Monte Carlo usando Cadenas de Markov). Los avances en el poder de cómputo y técnicas de simulación transformaron en viable la aplicación de inferencia Bayesiana en casi cualquier problema concreto.

Este curso introduce el paradigma bayesiano para hacer inferencia a través de algunas aplicaciones desde modelos simples con unos pocos parámetros a estimar que en general pueden resolverse analíticamente, hasta modelos con estructuras más complejas en los que la posterior es aproximada computacionalmente.

El objetivo central es que los participantes tengan un panorama de las etapas para realizar un análisis Bayesiano de datos moderno. Se usará el programa STAN (<http://mc-stan.org/>), introduciendo a los participantes a la programación en este lenguaje para estimar los modelos estadísticos de interés.

Destinatarios: profesionales que utilicen la Estadística; docentes de Estadística, Econometría, Biometría; otras personas que trabajen con análisis de datos.

Programa

1. Introducción al Análisis Bayesiano. Regla de Bayes, distribución *a posteriori*.

2. Modelos de un parámetro: Binomial, Poisson, Normal.
3. Distribuciones *a priori* o previas: conjugadas, no informativas, débilmente informativas.
4. Inferencia con cálculo numérico: integración numérica, aproximación con grillas, simulación Monte Carlo, simulación MCMC. Software bayesiano.
5. Diagnóstico y verificación de modelos.

Los puntos 1 a 3 del programa se cubrirán en la sesión del primer día y los puntos 4 y 5 en el segundo día.

Referencias bibliográficas

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B., 2013. *Bayesian Data Analysis*. CRC Press.

McElreath, R., 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall.

Stan Development Team (2017) *Stan Modeling Language: User's Guide and Reference Manual*. Version 2.17.0.

Palabras Clave: Estadística Bayesiana – Monte Carlo con Cadenas de Markov – Lenguaje STAN

Categoría Metodológica: Métodos Bayesianos.

MODELOS GRÁFICOS GAUSSIANOS

Marcelo Ruiz

Universidad Nacional de Río Cuarto

mruiz@exa.unr.edu.ar

Resumen

Los modelos gráficos Gaussianos constituyen una potente herramienta de modelización de propiedades de objetos pertenecientes a diversos campos científicos como genética, salud humana, epidemiología, entre otros.

Sea $\mathcal{G} = (V, E)$ un grafo con $V \neq \emptyset$ el conjunto de nodos o vértices y E , el conjunto de lados, un subconjunto de $V \times V = V^2$. Dos nodos i y j son denominados conectados, adyacentes o vecinos si $(i, j) \in E$. El grafo \mathcal{G} es no dirigido si satisface la condición $\forall (i, j) \in V^2 : (i, j) \in E$ si y sólo si $(j, i) \in E$. Para simplificar asumiremos que $V = \{1, \dots, p\}$.

Un *modelo gráfico* (MG) es un grafo tal que V indexa un conjunto de variables $\{X_1, \dots, X_p\}$ y el conjunto de lados E se define como $(i, j) \in E$ si y sólo si X_i y X_j son condicionalmente dependientes dadas las restantes variables $\{X_l : l \in V \setminus \{i, j\}\}$.

Dado un nodo $i \in V$, su entorno $A_i = \{l \in V \setminus \{i\} : (i, l) \in E\}$ indica cuáles son los nodos conectados con i y, en consecuencia, un MG puede ser descrito por su sistema de entornos $\{A_i\}_{i=1}^p$.

Un modelo gráfico $\mathcal{G} = (V, E)$ se denomina *Gaussiano* (MGG) cuando $(X_1, \dots, X_p)^T \sim N(0, \Sigma)$ con matriz de covarianza, $\Sigma = (\sigma_{ij})_{i,j=1,\dots,p}$, definida positiva. Los MGG constituyen el objeto de estudio en este curso.

En un MGG $\mathcal{G} = (V, E)$, el conjunto de lados representa la estructura de dependencia condicional del vector $\{X_1, \dots, X_p\}$ y, para trasladarla a un problema estadístico, es necesario hallar parametrizaciones de E . Una de ellas se da a través de la matriz de precisión, $\Omega = \Sigma^{-1} = (\omega_{ij})_{i,j=1,\dots,p}$. Más específicamente, se puede probar que para un MGG el conjunto de lados queda caracterizado por la siguiente equivalencia

$$(i, j) \in E \text{ si } \omega_{i,j} \neq 0.$$

De este modo, hallar el conjunto de variables que son condicionalmente dependientes es equivalente a determinar el conjunto de elementos no nulos de la matriz de precisión.

La *selección de covarianza* (ver Dempster, 1972) es una denominación para el conjunto de procedimientos que persiguen el objetivo de identificar la dependencia condicional en un MGG a partir de una muestra.

En *alta dimensión*, cuando p es mayor que el número de observaciones, n , la covarianza muestral S no es invertible y el estimador de máxima verosimilitud de Σ no existe. Incluso cuando $p/n \leq 1$, pero cercano 1, S es invertible pero está mal condicionada, incrementando el error de estimación (ver Ledoit, 2004). Para tratar con este problema, varios procedimientos de selección de covarianzas han sido desarrollados asumiendo que la matriz de precisión es *rala*.

En este curso estudiaremos procedimientos de selección de covarianza consistentes en *selección de entornos* basados en estrategias de *regularización* tipo lasso y aproximaciones *stepwise*. Para la primera estrategia ver Meinshausen y Bühlmann

(2006) y Friedman et al. (2008), entre otros y para la segunda ver Lafit, Nogales, Ruiz y Zamar (2018), entre algunos autores.

El curso tendrá un recorrido en el cual se entrelazan teoría y aplicaciones.

Objetivos:

- Introducir la teoría y procedimientos estadísticos recientes en modelos gráficos gaussianos de alta dimensión.
- Mostrar la importancia que la selección de covarianza posee en diferentes aplicaciones.

Destinatarios: Estudiantes de grado y egresados de carreras afines.

Programa:

Unidad I. Grafos y Modelos Gráficos.

Unidad II. Parametrizaciones de la matriz de precisión.

Unidad III. Regularización

Unidad IV. Selección de entornos en un MGG.

Referencias bibliográficas

Bühlmann, P., Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer.

Friedman, J., Hastie, T., Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. (2008). *Biostatistics*, 9 (3), 9 (3), 432–441.

Lafit, G., Nogales, J., Ruiz, M., Zamar, R. (2018). A Stepwise Approach for High-Dimensional Gaussian Graphical Models. En escritura.

Lauritzen, L. (1996). *Graphical Models*. Oxford University Press.

Dempster, A. (1972). Covariance selection. *Biometrics*.157–175.

Kalisch, M., Bühlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *The Journal of Machine Learning Research*, 8, 613–636.

Meinshausen, N. Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34 (3), 1436–1462.

Ren, Z., Sun, T. Zhang, C., Zhou, H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *The Annals of Statistics*, 43 (3), 991–1026.

Palabras Clave: Modelos gráficos, Normal multivariada, grafos raros, selección de covarianza, regularización, stepwise

Categorías Metodológicas: Métodos multivariados y Grandes Datos (Big Data).

TALLERES

DISEÑO DE PROBLEMAS PARA EL DESARROLLO DEL PENSAMIENTO ESTADISTICO EN LOS ALUMNOS

Noemi Maria Ferreri, Graciela Haydée Carnevali

Departamento de Matemática, Escuela de Formación Básica, Facultad de Ciencias
Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario
nferreri@fceia.unr.edu.ar

Introducción:

En muchos ámbitos se viene insistiendo desde tiempo en la necesidad de que en los cursos de Estadística se trabaje para ir desarrollando el pensamiento estadístico en los alumnos, buscando que ellos sean capaces de comprender la gran cantidad de información que circula por los medios de comunicación masivos y también que puedan ayudar a resolver los problemas de naturaleza estadística que tienen lugar en diferentes campos de aplicación. Este reclamo no se circunscribe únicamente a los cursos de Estadística del nivel superior sino también a los cursos de Estadística o de Matemática del nivel medio e incluso al nivel primario.

El pensamiento estadístico es un pensamiento complejo que no se adquiere en el corto plazo, pero todo lo que se haga en los cursos de Estadística va a favorecerlo.

Una de las dimensiones principales de este pensamiento está asociada a poder transitar adecuadamente las diferentes etapas de la resolución de un problema de naturaleza estadística, que siguiendo a los autores Wild y Pfannkuch, se puede denominar PPDAC (Planteo del Problema, Planificación del Estudio Estadístico, Recolección de los Datos, Análisis de los Datos, Conclusiones). Para que los alumnos puedan transitar adecuadamente este ciclo es necesario que en los cursos puedan resolver más o menos frecuentemente problemas de naturaleza estadística y puedan recorrer la mayor cantidad de etapas del ciclo, sin descuidar el desarrollo de los conceptos estadísticos asociados.

¿Qué características deben presentar, entonces, los problemas que se les propongan? En primer lugar, no deben plantear situaciones muy simplificadas, de manera que para resolverlos, sólo tengan que aplicar una fórmula, hacer unas cuentas y obtener conclusiones que no puedan relacionar con algún contexto de aplicación. Por el contrario, deben tener la mayor cantidad de “ingredientes” posibles, de manera que en su resolución los alumnos vayan transitando las distintas etapas del proceso. En segundo

lugar, deben pertenecer a algún área de aplicación del futuro ejercicio profesional, en el caso de la enseñanza superior o bien de interés de los alumnos, en el caso de la enseñanza a nivel secundario. Esto favorece el compromiso del alumno con la problemática y facilita la tarea. Por otra parte, en el momento de evaluar el desempeño de los alumnos en la tarea de resolución, es importante que el docente pueda recorrer todas las etapas del proceso y detectar en cuál o en cuáles se presentan las mayores dificultades, para poder re-orientar su labor docente y las próximas actividades a proponer.

Tanto para la confección de los problemas para proponer a los alumnos como para evaluar su trabajo, se requieren indicadores de las distintas etapas del proceso de resolución de problemas y de los conceptos básicos de Estadística. El taller se orienta en esa línea y sus objetivos son:

- Destacar la necesidad de desarrollar el pensamiento estadístico en los alumnos tanto de nivel secundario como de nivel superior.
- Presentar indicadores que guíen a los docentes en la confección de problemas para el trabajo con los alumnos y en la tarea de evaluación de su trabajo.
- Elaborar problemas a partir de ejercicios de libros y diseñar las guías de evaluación

Dirigido a: docentes o interesados en la docencia de la Estadística en cualquier nivel educativo

Duración del Taller: 1 encuentro de 2/3horas

Número máximo de participantes:40

Contenidos:

El pensamiento estadístico. Su importancia en la interpretación y en la resolución de problemas que se presentan en diferentes áreas de aplicación. Necesidad de su desarrollo en los alumnos de diferentes niveles educativos. Aplicación de indicadores para guiar la confección de problemas y para la evaluación del desempeño de los alumnos en su resolución

Conocimientos previos requeridos:

Se requiere estar familiarizado con los contenidos básicos de un curso de Estadística.

Propuesta de evaluación:

En el taller se pondrán a disposición de los asistentes diferentes textos con definiciones y caracterizaciones del pensamiento estadístico, y luego se intercambiarán

opiniones y reflexiones acerca de la necesidad de los docentes de ayudar a su desarrollo en los alumnos.

Luego se propondrán indicadores que guíen el trabajo de los docentes con los alumnos para el desarrollo y la evaluación del pensamiento estadístico. Se brindarán ejemplos de casos con indicadores ya aplicados, tanto para el trabajo en clase como para la evaluación.

Finalmente, se trabajará en grupos de dos o tres personas, a partir de problemas tomados de libros de Estadística para el nivel medio y para el nivel superior (diferentes especialidades). En dichos problemas se identificarán los indicadores presentes y luego se propondrán las modificaciones correspondientes para transformarlos en “casos” con mayor cantidad de “ingredientes”.

El trabajo de los asistentes se evaluará a partir del análisis del problema propuesto y de su transformación, aplicando los indicadores propuestos u otros que puedan surgir.

Referencias bibliográficas

- Behar, R. y Grima, P.; La Estadística en la Educación Superior. ¿Formamos Pensamiento Estadístico? Ingeniería y Competitividad, Vol 5, N° 2, pp 84-90 (2004)
- Behar, R.; El rol de la estadística en el trabajo del ingeniero. Ingeniería y Competitividad, Vol 4, N° 1, pp 47-54 (2002)
- Ben-Zvi, D. y Garfield, J.; Statistical Literacy, Reasoning and Thinking: Goals, Definitions and Challenges. En D. Ben-Zvi y J.Garfield (Eds.) The Challenge of Developing Statistical Literacy, Reasoning and Thinking, Dordrecht, Holanda: Springer (2004)
- Carnevali, G. H. y Ferreri, N.M. Resolución de problemas de naturaleza estadística: aplicación de indicadores para su evaluación, XIX EMCI Nacional y XI EMCI Internacional, San Nicolás (2015)
- Carnevali, G.H. y Ferreri, N. M. Resolución de problemas de decisión estadística: diseño y aplicación de indicadores para su desarrollo y evaluación, XVIII EMCI Nacional y X EMCI Internacional, Mar del Plata (2014)
- Carnevali, G. H. y otros Desarrollo del pensamiento estadístico: una experiencia con alumnos de Ingeniería Industrial. EMCI XIV, Mendoza (2008)
- Hoerl, R; Snee, R. : Statistical thinking and methods in quality improvement: A look to

- the future. *Quality Engineering*, 22,3, 119-129 (2010)
- Isaac Godinez, C. y Lamar Meneses, F. El pensamiento estadístico en la formación del ingeniero. En 15 Convención Científica de Ingeniería y Arquitectura, La Habana, Cuba (2010)
- Krishnamoorthi, K.S. Statistical thinking for Engineers. What, Why and How?. En IEEE Conference on Industrial Engineering and Engineering Management, Xiamen, China.(2010)
- Medina Martínez, N; Estrategia didáctica para la formación del pensamiento estadístico en los estudiantes. *Pedagogía Universitaria*, 16, (4), 136-159(2011)
- Montgomery, D., Runger, G. & Hubele, N.; *Engineering Statistics*, New Jersey, Estados Unidos: John Wiley & Sons (2010)
- Moore, D.: *Estadística Aplicada Básica*. Barcelona, España: Antoni Bosch Editor (2005)
- Moore, D.; What shall we teach beginners? (Invited discussion of a paper from Wild and Pfannkuch). *International Statistical Review*, Vol 67, N° 3, pp.250-252 (1999)
- Pfannkuch, M.; Training teachers to develop statistical thinking. En C. Batanero y otros (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*, Monterrey, México (2008)
- Snee, R. : Statistical thinking and its contribution to total quality. *The American Statistician*, 44, 2, 116-121 (1990)
- Wild, C. & Pfannkuch, M. Towards an Understanding of Statistical Thinking. En D. Ben-Zvi y J.Garfield (Eds.) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, Dordrecht, Holanda: Springer(2004)
- Wild, C.; Pfannkuch, M.: Statistical Thinking in Empirical Enquiry (with discussion). *International Statistical Review*, Vol 67, N° 3, pp.223-265 (1999).

Palabras Clave: Pensamiento Estadístico – Indicadores – Ciclo PPDAC – Problemas

Área de Aplicación: Enseñanza de la Estadística

EL USO DEL DATO ESTADISTICO EN SALUD

Juan Pablo Incocciati (1)

Diana Rosana Kucukbeyaz (2)

SALUD PUBLICA LA MATANZA Y MINISTERIO DE SALUD (1)

UNTREF (2)

jpabloinco@gmail.com-ranush2003@yahoo.com.ar

Introducción:

La Función Pública, en Salud es, ante todo, un derecho inalienable y trascendental para el ser humano.

Los organismos encargados de llevar adelante políticas de Salud Pública, necesitan información precisa sobre la situación que atraviesa su sistema de Salud en relación con la comunidad a la cual sirve. Crear un instrumento que permita recopilar y analizar toda esta información, derivó en este proyecto, “Gestión de la Salud” el cual se viene desarrollando desde el año 2005 hasta el presente en la Secretaría de Salud Pública del Municipio de La Matanza (Provincia de Buenos Aires).

El proyecto parte desde la premisa de implementar un sistema de gestión de la Salud que permita brindar la mayor calidad de atención posible, y la mejor calidad del dato estadístico.

Se generó para desarrollar este proyecto de gestión “La Dirección de Auditoría Interna en Salud”, la cual es hoy como desde sus inicios, un área que persigue, y tiene como finalidad, ser un instrumento de información, de comunicación y de relevamiento de los organismos que articulan las Políticas Sanitarias de una comunidad.

Para llegar a ser un instrumento eficaz de información, fue necesario aprender cómo recolectar, analizar, archivar dichos datos, convertir estos datos en información pertinente, sensible, específica y factible; logrando la creación de un tablero de indicadores técnicamente diseñados con herramientas estadísticas.

Los procedimientos para el diseño de cada indicador, se plasmaron en una normativa, “Manual de Auditoría Interna en Salud”, junto con los formularios correspondientes al relevamiento de los datos necesarios para la construcción de cada indicador del Tablero de Indicadores de consulta permanente.

Objetivos:

Que el participante

- Conozca el marco normativo de las estadísticas oficiales para el desarrollo de la Dirección de Auditoría Interna en Salud Interpretar el procedimiento para la recopilación de o de los datos necesarios para la construcción de indicadores de la gestión en Salud que permitan la toma de decisiones, en políticas de Salud
- Aplique los formularios de recopilación de datos correspondientes a los diferentes servicios de Salud en situaciones concreta de campo.
- Adquiera capacidad de realizar informes sólidos, que presenten los datos disponibles de forma clara, sintética y comprensible para una decisión buena en Salud

Destinatarios:

Dicho taller está destinado a toda persona estadístico o no, que se desempeñe dentro del área afectada por las estadísticas sistematizadas del sistema de salud.

Cronograma y Programa:

Primer encuentro del taller:

-Definiremos por medio de una presentación: “El rol del Estadístico en la Salud”.

-Se organizarán grupos de trabajo con los presentes, se distribuirán (en forma impresa) las normativas de control estadísticos para los diferentes servicios de salud junto con los formularios oficiales de recolección de datos vigentes en la Salud Pública.

Se realizarán reflexiones internas en los grupos de trabajo sobre las variables que se presentan en ellos.

-Responderán: ¿Qué representan estas variables en un análisis de la Gestión en Salud? ¿Son suficientes para una decisión en Salud? ¿Hay otras variables que influyen y no están en estos formularios?

-A partir del debate surgirá la falta de planillas estadísticas de trabajo interno, con variables que no se encuentran en los materiales trabajados y necesarios para la construcción de indicadores que permitan evaluar los diferentes servicios de salud y tomar la mejor decisión dentro de la gestión.

Segundo encuentro del Taller:

-Retomando el tema de “necesidad de planillas estadísticas de trabajo interno” con el que culminó el encuentro anterior, y en sus mismos grupos de trabajo, analizarán diferentes ejemplos vigentes de modelos de planillas de relevamiento de datos contruidos para la tarea interna de gestión de la Dirección de Auditoría Interna en Salud de La Matanza (Provincia de Buenos Aires).

- Los presentes reflexionarán en sus grupos de trabajo, sobre estas planillas que contienen los datos estadísticos que son utilizados dentro del proceso de análisis de los servicios de Salud para lograr la medición de las prácticas sanitarias.
- Compartimos las reflexiones
- Propondremos diseñar indicadores a partir de los datos de los formularios discutidos en los equipos de trabajo.
- Puesta en común
- Como cierre, del encuentro, se expondrán los indicadores hoy utilizados en la Gestión en Salud, de los servicios de Radiología, Laboratorio y Consultorios Externos por la Dirección de Auditoría Interna en Salud de La Matanza (Provincia de Buenos Aires)

Referencias bibliográficas

- Cappelletti, Carlos. Elementos de Estadística. Cesarini Hns. 2da Edición. Año 1983
- Kelmansky Diana M. Estadística para todos – 1ª Edición –Buenos aires; Ministerio de la Nación. Instituto Nacional de Educación Tecnológica, 2009.
- Indicadores para la gestión de los servicios de Salud Pública. Villalba, Joan R. Vol 24, Issn 0213-9111.
- César N. Aguirre. Estadística aplicada en las Ciencias Sociales y Humanas. Estadística I. Colección: Cuadernos de cátedra. Editorial Universitaria de Misiones. Año 2005
- Instructivo para el llenado para el formulario 102 R, Resumen Mensual de Radiología, Ministerio de Salud de la Provincia de Buenos Aires, 1997
- Sistema de Información Odontológico, Programa Provincial de Salud Bucal, Provincia de Buenos Aires, Ministerio de Salud, 1999.
- Actualización de Factores de Conversión para Evaluación de Laboratorios Hospitalarios, Ministerio de Salud de la Provincia de Buenos Aires,1995

Palabras Clave: Salud –Estadísticas – Indicadores – Control

Área de Aplicación: Salud Humana

RE-PENSAR LA ENSEÑANZA DE LA ESTADÍSTICA EN LA UNIVERSIDAD: APRENDIZAJE SIGNIFICATIVO, CONECTIVIDAD Y UBICUIDAD

Gabriela Pilar Cabrera y Andrea Peña Malavera

IAP de Ciencias Básicas y Aplicadas. Universidad Nacional de Villa María.

gabriela.pilar.cabrera@gmail.com

Resumen

Repensar la enseñanza de la Estadística para cursos introductorios de nivel universitario en carreras con escasa formación matemática, supone un reto. En respuesta a esto, en el marco de las investigaciones llevadas a cabo por el Grupo de Investigación y Desarrollos Didácticos (GIDED) de la UNVM, se trabajó en la generación de un dispositivo didáctico centrado en el desarrollo del razonamiento estadístico. Se entiende al razonamiento estadístico (Ben-Zi, 2016) como la forma en que las personas crean y dan sentido a la información estadística e infieren e interpretan críticamente los resultados estadísticos. Este razonamiento supone una comprensión conceptual de ideas estadísticas fundamentales: datos, variabilidad, aleatoriedad, centro, distribución, muestreo, inferencia y probabilidad. La concepción de dicho dispositivo se sustenta en los principios de Aprendizaje Significativo (Moreira, 2012), del Conectivismo (Siemens, 2014) y la consideración de que la ubicuidad caracteriza la situación actual (Igarza 2016, p.13). También se adopta la idea de Taquini (2016, p.7) en relación a que el siglo XXI debe forjarse en una Nueva Educación que aproveche las tecnologías disponibles para seguir formando personas autónomas y conscientes de su rol en el mundo, que puedan afrontar una realidad cambiante y desenvolverse en armonía con el medioambiente y con sus pares. El desarrollo didáctico en cuestión propicia el carácter dialógico-reflexivo y cooperativo en cada clase, el acceso y tratamiento de conceptos y procedimientos a partir de redes conceptuales, el aprendizaje ubicuo, el uso del software estadístico *InfoStat* (Di Rienzo *et al.* 2018), aplicaciones diseñadas con el paquete Shiny del entorno de programación R, los juegos y simulaciones de experimentos aleatorios. Además, procura a través del E-portafolio y de las evaluaciones “ensayo”, un proceso de evaluación continua por parte del equipo docente, de co-evaluación de pares y auto-evaluación de cada estudiante. Sumado a esto, una serie de videos tutoriales creados por el equipo docente para cada clase y disponibles en *YouTube*. Estos videos tutoriales refieren tanto a los conceptos y procedimientos desarrollados en la clase, como al modo

en el que fueron abordados. Es así que los estudiantes disponen de todas las clases para ser “re-visadas” y “re-construidas” en un tiempo diferente al que se suceden las mismas (Cabrera 2018, p.70). De este modo se compone un dispositivo didáctico dinámico, colaborativo, interactivo, *hipertextual*, situado con una alta idoneidad didáctica global.

Objetivos:

- Propiciar un espacio de reflexión en relación a la enseñanza y aprendizaje de la estadística en la universidad y favorecer la innovación educativa.

Destinatarios:

- Docentes que dictan Estadística en carreras universitarias.
- Docentes universitarios en general.

Ejes temáticos:

Módulo 1:

- Aprendizaje significativo, conectividad y ubicuidad. Aula extendida.
- Razonamiento estadístico. Ideas estadísticas importantes.
- Supuestos del Razonamiento Inferencial Informal.
- Caracterización de los componentes del dispositivo didáctico. Módulo 2:
- Dinámica estructural del dispositivo didáctico.
- Modalidad de implementación del dispositivo didáctico.
- Análisis de las componentes del dispositivo didáctico.
- Aplicación de criterios de idoneidad didáctica para la evaluación del

dispositivo.

Referencias bibliográficas:

BenZvi, D., & Makar, K. (2016). International perspectives on the teaching and learning of statistics. In Ben-Zvi and Makar. (Eds.), *The teaching and learning of statistics: International perspectives* 1-10. Springer International Publishing Switzerland.

Cabrera, G. P. (2018). *Un dispositivo didáctico para cursos de Estadística en el Nivel Superior*. En Marcel David (2018) *Relatos de investigación y experiencias docentes en Educación Matemáticas* (2018). GIDED. UNVM. ISBN978-987-42-7307-9. Recuperado de <http://gided.unvm.edu.ar/index.php/book/relatos-de-investigacion-y-experiencias-docentes-en-educacion-matematica/>

Di Rienzo J., Casanoves F., Balzarini M. G., González L., Tablada M., Robledo W.
InfoStat Versión 2016. Grupo InfoStat. FCA, UNC. Argentina. URL
<http://www.infostat.com.ar>

Igarza, Roberto (2016). Futuro busca presente: la educación de la Nube. *Ciencia e Investigación*, 66(5), 9-18.

Moreira, M. A. (2012). ¿Al final, qué es aprendizaje significativo. *Curriculum: Revista de teoría, investigación y práctica educativa*, 25, 29-56.

Siemens, G. (2014). Connectivism: A learning theory for the digital age.

Taquini, Alberto (2016). Hacia una nueva educación. *Ciencia e Investigación*, 66(5), 3-9.

Palabras Clave: Razonamiento estadístico – Dispositivo didáctico – Aprendizaje Ubicuo.

Área de Aplicación: Enseñanza de la Estadística.

COMUNICACIONES ORALES

A horizontal green arrow pointing to the right, with the text 'COMUNICACIONES ORALES' centered inside it.

Control de procesos

A horizontal green arrow pointing to the right, with the text "Control de procesos" written inside it in a bold, blue, serif font.

ALGUNAS TENDENCIAS ACTUALES EN EL DESARROLLO DE METODOLOGIA ESTADISTICA PARA LA MEJORA CONTINUA

Andrea F. Righetti, Silvia Joeques

Facultad de Ciencias Económicas. Universidad Nacional de Córdoba

Mail de contacto: analizamos@yahoo.com.ar

Resumen

El control estadístico de procesos (CEP) se emplea desde hace varias décadas con el propósito de controlar y reducir la variabilidad de procesos, productos y/o servicios, mejorando su calidad y productividad. La producción científica en esta área ha crecido significativamente después de 1990, lo que indica que el CEP continúa siendo una importante área de investigación. Sin embargo, algunos autores advierten sobre la brecha existente entre las investigaciones desarrolladas y su utilización práctica en los proyectos de mejora continua, entendiendo ésta como el procedimiento que se enfoca en prevenir los defectos y fallas que los procesos puedan producir, con el fin de obtener bienes y servicios de alta calidad con bajos costos.

Farmer, Woodall, Steiner y Champ (2014), Woodall y Montgomery (2008), sostienen que el monitoreo de procesos por sí mismo no es suficiente para producir una mejora significativa en ellos. Por este motivo, apoyan firmemente el uso de metodologías de gestión de la calidad, como Six Sigma, que mediante un enfoque disciplinado, orientado a proyectos y basado en metodología estadística, busca reducir la variabilidad y eliminar defectos y desperdicio de productos, procesos y servicios. Desde esta perspectiva es importante identificar los avances logrados recientemente en cuanto a las nuevas metodologías desarrolladas para optimizar el CEP.

Objetivo: Proporcionar una síntesis de los avances más importantes y las principales tendencias en el desarrollo de metodologías estadísticas para la mejora continua. La atención se centra en los trabajos realizados en la última década.

Metodología: Revisión de la bibliografía de los últimos diez años referida al desarrollo de nuevas metodologías estadísticas para la mejora continua.

Resultados o conclusiones: En los últimos años el desarrollo de metodologías, dentro del campo del control estadístico de la calidad, se ha centrado principalmente en dos aspectos: a) el efecto del incumplimiento de las hipótesis estadísticas básicas sobre los gráficos de control multivariados y b) en la construcción de técnicas estadísticas que

permitan la reducción eficiente de la dimensionalidad explotando la estructura de correlación poblacional.

En el primero, los estudios se focalizan principalmente en: i) Análisis de la robustez de los modelos ante alteraciones en la hipótesis de normalidad y en el estudio del comportamiento de los contrastes habituales ante distribuciones de la familia exponencial, ii) Métodos no paramétrico del control estadístico de la calidad, con nuevos gráficos de control y análisis comparativo de su efectividad con las técnicas paramétricas clásicas y iii) Desarrollo y análisis de técnicas de eliminación de la autocorrelación muestral, especialmente la aplicación de modelos de series temporales multivariadas.

En el segundo, se abordan líneas referidas a: i) Obtención de factores comunes dinámicos, que puede permitir la reducción de la dimensionalidad sin pérdida excesiva de la cantidad de información muestral. Concretamente, se está estudiando la distribución probabilística de los factores comunes, que debe permitir la obtención de límites estadísticos de control y el estudio de la potencia de los contrastes, ii) Desarrollo de las técnicas de obtención de tendencias comunes, que permitan explotar la estructura de correlación poblacional para obtener gráficos de control más eficientes que los clásicos en la detección de observaciones fuera de control estadístico.

Esta breve síntesis, muestra que cada día, tanto en las organizaciones como en la sociedad, existe una mayor preocupación por los temas relacionados con la calidad. De allí la gran evolución que han presentado las técnicas de control de calidad y, entre ellas, los métodos estadísticos de mejoramiento de la calidad.

Referencias bibliográficas:

- Capizzi, G. (2014) Recent Advances in Process Monitoring: Nonparametric and Variable-Selection Methods for Phase I and Phase II, *Quality Engineering*, 27:1, 44-67, DOI: 10.1080/08982112.2015.968046
- Chakraborti, S., Human, S. W., Graham, M. A., & Phase, I. (2008). Statistical process control charts: an overview and some results. *Quality Engineering*, 21(1), 52-62.
- Farmer L. A., Woodall W.H., Steiner, S. H. and Champ, C. W. (2014).An Overview of Phase I Analysis for Process Improvement and Monitoring.*Journal of Quality Technology*.46(3), 265-280.
- Hoerl, R. W., & Snee, R. (2010). Statistical thinking and methods in quality improvement: a look to the future. *Quality Engineering*, 22(3), 119-129.

Li, J., Tsung, F., and Zou, C. (2012). Directional Control Schemes for Multivariate Categorical Processes. *Journal of Quality Technology* 44(2), pp. 136–154.

Lizarelli, F. L., Bessi, N.C., Oprime, P. C., Amaral, R. M., & Chakraborti, S. (2016). A bibliometric analysis of 50 years of worldwide research on statistical process control. *Gestão & Produção*, 23(4), 853-870.

Ryan, T. P. (2011). *Statistical Methods for Quality Improvement*, 3rd Edition. Hoboken, NJ: John Wiley & Sons, Inc.

Woodall, W. H. and Montgomery, D. C. (2008). An Overview of Six Sigma. *International Statistical Review*, 76, 3, 329.

Woodall, W. H. and Montgomery, D. C. (2014). Some Current Directions in the Theory and Application of Statistical Process Monitoring. *Journal of Quality Technology*. Taylor & Francis 46(1) pp. 78–94.

Palabras Clave: Control Estadístico de Procesos (CEP) – Mejora Continua – Gráficos de control.

Área de Aplicación: Industria y Mejoramiento de la calidad.

Datos de duración. Datos faltantes.

A green arrow graphic pointing to the right, with a white outline of a second arrow nested inside it, creating a double-arrow effect.

ANÁLISIS DE SOBREVIDA MEDIANTE EL SOFTWARE R, APLICACIÓN A PACIENTES CON CÁNCER DE PÁNCREAS

Claudia Castro Kuriss

Instituto Tecnológico de Buenos Aires (ITBA)

ccastro@itba.edu.ar

Resumen

El análisis de supervivencia es un tópico de interés en Ciencias de la Salud e Ingeniería donde se estudia el tiempo de especímenes puestos a prueba o expuestos a un factor o enfermedad hasta que sobreviene un evento de interés. Los tiempos de aquellos objetos que no tienen el evento o no fallan durante el estudio se denominan censurados. El análisis de los datos censurados requiere de técnicas específicas desarrolladas para ello que reciben el nombre general de Análisis de Supervivencia (Confiablez en Ingeniería). El modelo de Cox es el que habitualmente se emplea para modelizar la variable Tiempo hasta la falla. Ver Cox y Oakes (1984). Este modelo requiere de suposiciones muy fuertes que muchas veces no son cumplidas por las co-variables elegidas para explicar la supervivencia de los ejemplares, ya que se trata de un modelo de riesgos proporcionales. El modelo debe ser significativo globalmente y según cada co-variables, y además es necesario que se satisfagan los criterios del modelo de Cox. Las formas de estudiar si se satisfacen o no estos criterios son: a) mediante el empleo de gráficos, b) mediante del estudio de los residuos del modelo y c) a través de tests formales. Cuando se determina que una o varias variables no satisfacen las hipótesis requeridas entonces se pueden utilizar extensiones al modelo de Cox como son la incorporación de co-variables dependientes del tiempo o el modelo de Cox estratificado. Ver Therneau y Grambsch (2000). Estos diagnósticos y modelizaciones se pueden realizar con el lenguaje R que tiene una enorme versatilidad pero es necesario conocer las distintas librerías donde el tema se encuentra desarrollado ya que para poder abarcar todos los tópicos se requieren por lo menos diez de ellas además del paquete básico denominado *survival* (Therneau, 2015). El objetivo del presente trabajo es el de proponer una forma práctica de extender el modelo de Cox a co-variables dependientes del tiempo en lenguaje R que requiere que la base de datos sea adaptada al proceso estocástico de conteo que está involucrado, además de presentar las posibilidades en R para abordar el tema de supervivencia a través de las distintas librerías. Específicamente se

emplearán las técnicas propuestas para el estudio de un conjunto de datos reales provenientes de un ensayo clínico de tiempos hasta la progresión de pacientes con cáncer de páncreas. Ver Moore (2016). Para cada paciente se registró el tipo de tumor, su estadio, la fecha de muerte, y la de progresión si se observó antes de la muerte. Se aplicaron las técnicas de sobrevida al conjunto de datos. No se encontró diferencia significativa para la sobrevida según el tipo de tumor (p-valor 0.134). Se consideró a *Estadio* como predictor de la progresión. Se encontraron los residuos de Score, Martingala y de Schoenfeld para el modelo. Mediante gráficos y tests formales para el modelo de Cox propuesto, se determinó que la variable *Estadio* no cumplía la suposición de riesgo proporcional (Klein y Moeschberger, 2003). Se propusieron varias alternativas para superar este problema, una fue considerar la variable *Estadio* dependiente del tiempo. Dentro de las propuestas se eligió el mejor modelo de acuerdo al criterio de verosimilitud máxima y al de Akaike (Sakamoto et. al, 1986). Todos los gráficos, análisis, tests y validaciones se hicieron manejando rutinas desarrolladas por el autor en R utilizando varios paquetes sobre el tema de interés.

Referencias bibliográficas

- Cox, D.R., Oakes, D. (1984) Analysis of Survival Data. Chapman and Hall, New York.
- Klein, J.P., Moeschberger, M.L. (2003) Survival Analysis. Techniques for Censored and Truncated Data. Springer, New York.
- Moore D. F. (2016) Applied Survival Analysis Using R. Springer, New York.
- Therneau, T., Grambsch P. (2000) Modeling Survival Data: Extending the Cox Model. Springer, New York.
- Therneau T (2015). A Package for Survival Analysis in S, version 2.38, <https://CRAN.R-project.org/package=survival>
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Sakamoto, Y., Ishiguro, M., and Kitagawa G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publishing Company.

Palabras Clave: Análisis de Sobrevida-Co-variables dependientes del tiempo-Lenguaje R.

Área de Aplicación: Biología-Salud Humana-Otras Ciencias de la Salud-Industria y Mejoramiento de la Calidad.

Estadística espacial.

A green arrow graphic pointing to the right, with a white outline of a second arrow nested inside it.

COMPARACION DE MODELOS GEOESTADISTICOS: UNA APLICACIÓN AL ESTUDIO DE CONTAMINACION MAGNETICA

Mauro A. E. Chaparro¹⁻³, Marcela Natal¹, Lila Ricci¹, Marcos A. E.
Chaparro²⁻³

¹Centro Marplatense de Investigaciones Matemáticas- Universidad Nacional de Mar del
Plata (CEMIM-UNMDP)

²Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos
Aires (CIFICEN, CONICET-UNCPBA)

³Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
chaparromauro76@gmail.com

Resumen

El monitoreo de la calidad del aire a través de la utilización de técnicas del magnetismo ambiental se presenta como una herramienta alternativa a los métodos clásicos desde hace tiempo. El material particulado (PM) generado por el parque automotor (vehículos particulares y de servicio público) es producido por combustión de gasolina y diesel y por el desgaste de los frenos. Un parámetro muy utilizado en los ensayos orientados a los estudios de magnetismo ambiental es la susceptibilidad magnética (específica y volumétrica). Es posible hacerla determinación de este parámetro magnético de forma directa con mediciones *in-situ* (κ) o bien a través de muestras en laboratorio (χ).

Si consideramos un sitio de referencia como limpio, se puede suponer que un “aumento magnético” en la concentración es debido a la emisión y depositación de partículas contaminantes y magnetizadas de origen antropogénico.

Los valores κ y χ se distribuyen en un área, no pueden ser considerados como muestras independientes, se aplicaron entonces métodos de estadística espacial. Las determinaciones a través de mediciones *in-situ* ofrecen un nivel de simplificación considerable con respecto a la determinación de la concentración en laboratorio, lo cual motiva la comparación de mapas, por esto nos propusimos construir modelos geoestadísticos para cada parámetro y posteriormente comparar sus mapas de predicción espacial. Aplicamos esta propuesta en el marco de un estudio sobre la distribución de la concentración de material magnético en la ciudad de Mar del Plata, Argentina. El área

de muestreo comprende aproximadamente 780 manzanas en la que se encuentran parte de la zona costera, microcentro, uno de los macrocentros comercial de la ciudad y zonas residenciales.

Se estudiaron variogramas empíricos con diferentes números de clases con tendencia de primer, de segundo orden y con la covariable “tránsito”. Los modelos de variogramas teóricos fueron: exponencial-circular-gaussianos y wave. Para la selección del modelo que mejor ajusta se aplicó el método “*leave one out*”. Para la etapa de predicción se utilizó para cada uno de los modelos construidos el método de Kriging ordinario. Para comparar analíticamente los mapas se utilizó un estadístico alternativo al coeficiente de Pearson que involucra la media y varianza Kriging de cada uno de los mapas. El coeficiente correlación de Pearson resultó significativo y de 0.63, mostrando un valor aceptable de correlación para el modelo que utiliza como covariable “tránsito”. Las mayores diferencias se observan en lugares alejados del microcentro, en zonas residenciales con registro en ambos métodos bajos. Para los cálculos y gráficos se utilizó el paquete geoR del software libre R.

Referencias bibliográficas

- Chaparro, Marcos A.E. ; Chaparro, Mauro, P. Rajkumar, V. Ramasamy , Ana M. Sinito. (2011). Magnetic parameters, trace elements and multivariate statistical studies of river sediments from the south eastern India: a case study from Vellar River. *Environmental Earth Sciences* . DOI: 10.1007/s12665-010-0704-2., 63(2):297-310. ISSN: 1866-6280.
- A.E. Chaparro, (2014). Radiation impact assessment of naturally occurring radionuclides and magnetic mineral studies of Bharathapuzha river sediments, South India. *Environmental Earth Sciences, Ed. Springer. DOI:10.1007/s12665-013-2751-y, 71 (8): 3593-3604. ISSN:1866-6280.*
- Gargiulo, J.; R. Senthil Kumar, Chaparro, M.; Chaparro, M.A; Natal, M. (2016). Magnetic properties of air suspended particles in thirty eight cities from south India. *Atmospheric Pollution Research, 7,626-637. ISSN: 1309-1042. Doi:10.1016/j.apr.2016.02.008.*
- Bivand, R. S.; Pebesma E, Gómez Rubio, V. (2013). *Applied Spatial Data Analysis with R*. Springer.
- Diggle, Ribeiro. (2007). *Model-based Geostatistics*. Springer.

Banerjee, S.; Carlin, B.; Gelfand, A. (2004). Hierarchical Modeling and Analysis for Spatial Data. ChapmanHall.

Palabras Clave: Geoestadística – Susceptibilidad magnética- Comparación de mapas.

Área de Aplicación Ecología y Medio Ambiente.

Inferencia Estadística

A green arrow graphic pointing to the right, with a white outline of the arrowhead.

EVALUANDO A LOS ALUMNOS A TRAVES DE INDICADORES DE PENSAMIENTO ESTADISTICO

Noemí M. Ferreri, Graciela H. Carnevali, Melina Pascaner

Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de

Rosario

nferreri@fceia.unr.edu.ar

Resumen

El pensamiento estadístico puede considerarse una filosofía de aprendizaje y acción que destaca entre sus principios la omnipresencia de la variabilidad y se apoya en tres pilares fundamentales: una apropiada comprensión de la teoría de la Probabilidad y la Estadística; el dominio de los procesos involucrados en la resolución de problemas de naturaleza estadística, los cuales se resumen en ciclos como, por ejemplo, el PPDAC (Wild & Pfannkuch, 1999) que van desde la formulación del problema hasta la elaboración de conclusiones, pasando por la obtención de datos de calidad y por la realización de los análisis estadísticos correspondientes y, por último, un conjunto de otras dimensiones como el pensamiento crítico y las actitudes que se requieren para la resolución de este tipo de problemas.

El desarrollo de este pensamiento complejo debe ser el principio director de los cursos de Estadística, lo que implica que las clases, los trabajos prácticos, las evaluaciones, etc. deben estar orientadas hacia el mismo. Para ello se han definido y presentado en trabajos anteriores un conjunto de indicadores tanto de las etapas del ciclo PPDAC, como de los conceptos estadísticos involucrados y de las otras dimensiones de interés.

El objetivo del presente trabajo es evaluar la resolución de dos problemas por parte de los alumnos de Ingeniería Industrial (FCEIA, UNR), aplicando indicadores de pensamiento estadístico y registrar sus aciertos y errores para luego reorientar la práctica docente.

Los problemas propuestos a los alumnos de Ingeniería Industrial a comienzos del segundo curso de Estadística son problemas de decisión estadística diseñados a partir de un conjunto de indicadores de pensamiento estadístico, que luego fueron utilizados para evaluar el desempeño de los mismos.

Los trabajos de los alumnos fueron analizados indicador por indicador y luego, con una visión general.

En los dos problemas propuestos se hacía una breve descripción de la situación, de modo que el alumno sólo debía reconocer todos los elementos “estadísticos” que allí se encontraban (población, variables, parámetros de interés, hipótesis, etc.). También se les brindaban datos de contexto para utilizar en la etapa de “Conclusiones”. La diferencia fundamental es que en el primero, los alumnos recibían una salida de software que debían utilizar para el análisis de los datos y las conclusiones; mientras que en el segundo, los alumnos recibían las conclusiones en contexto y debían proponer datos cuyo análisis resultara acorde con las conclusiones obtenidas. Este cambio fue pensado para que se consideraran ambas etapas en correspondencia.

En general, los alumnos transitaron adecuadamente la etapa de planteo del problema, especialmente para el primero de los propuestos y el análisis de los datos en ambos, aunque en esa etapa mayormente descuidaron chequear el cumplimiento de los requerimientos de las herramientas inferenciales aplicadas. A la hora de elaborar las conclusiones, sólo tuvieron en cuenta la información de contexto en el segundo problema, ya que en las consignas se les sugería aplicarla.

Algunos alumnos presentaron dificultades al confundir valores de la variable de interés, con los estadísticos y con los parámetros, lo que trajo aparejado problemas en la definición de la regla de decisión. En relación a este punto, también se observaron dificultades en la definición del estadístico de prueba del primer problema y de su distribución bajo la hipótesis nula. Por último cabe señalar que otra de las dificultades observadas fue mezclar tareas de las etapas de análisis con las de planificación del estudio estadístico.

El desarrollo del pensamiento estadístico es una competencia que se adquiere en el largo plazo y que trasciende un curso o dos de Estadística. El análisis de lo realizado por los alumnos implica que hay que seguir insistiendo en la resolución de este tipo de problemas, profundizando en los tres pilares.

Referencias bibliográficas

Carnevali, G.H. y otros (2017) Uso de Indicadores para el trabajo con casos en el primer curso de Estadística para Ingeniería Industrial. *XX EMCI Nacional y XII EMCI Internacional, Santiago del Estero*

- Carnevali, G.H. y Ferreri, N.M. (2015) Resolución de problemas de naturaleza estadística: indicadores para su evaluación en alumnos de ingeniería industrial, *XIX EMCI Nacional y XI EMCI Internacional, San Nicolás*
- Ribeiro Campos, C. (2016). La educación estadística y la educación crítica. *2do. Encuentro Colombiano de Educación Estocástica, Bogotá*. Recuperado de https://www.researchgate.net/publication/307895353_LA_EDUCACION_ESTADISTICA_Y_LA_EDUCACION_CRITICA
- Toapanta-Toapanta, G. M. y otros (2018). Las competencias para el aprendizaje de la Estadística en los estudiantes de Educación. *ROCA, Revista Científico Educativa de la Provincia Granma, Cuba, volumen 14* (1), pp. 253-266. Recuperado de <http://revistas.udg.co.cu/index.php/roca/article/view/976>
- Wild, C. y Pfannkuch, M. (1999) Statistical Thinking in Empirical Enquiry (with discussion). *International Statistical Review, volumen 67*, (3), pp. 223-265

Palabras Clave: Pensamiento Estadístico – Indicadores – Evaluación

Área de Aplicación: Enseñanza de la Estadística.

Métodos multivariados.



ANÁLISIS PREDICTIVO Y EXPLORATORIO DE ABANDONO EN SECUNDARIA

Natalia da Silva, Juan José Goyeneche

Instituto de Estadística, Facultad de Ciencias Económicas y de Administración

Universidad de la República

natalia@iesta.edu.uy

Resumen

El sistema de educación formal en Uruguay se divide en cuatro niveles: educación preprimaria e inicial (entre 3 y 5 años), educación primaria (entre 6 y 11 años), educación secundaria (entre 12 y 17 años) y finalmente educación terciaria o superior (18 años o más). Si bien se ha logrado tener un nivel importante de universalización de la educación primaria, se encuentran serios problemas en la educación secundaria para lograr retener a los estudiantes en el sistema.

El objetivo de este trabajo es explorar y predecir el abandono de los alumnos pertenecientes a primer año de educación secundaria pública en Uruguay. El abandono en secundaria es un problema importante en Uruguay siendo el primer año de dicho ciclo el que presenta mayores niveles de abandono. Analizar la población en riesgo de abandono es fundamental para entender el problema y obtener información relevante para el desarrollo de políticas públicas.

La población que se va a analizar son los estudiantes de primero de educación secundaria que cursan primero en el año 2016. Dicha población se compone de una matrícula de 40.233 alumnos, que pertenecen a 254 centros educativos de educación media pública.

Se dispone de las transiciones de dichos alumnos entre los años 2016 y 2017, esto es, el grado en que estaban en 2016 y en 2017.

La exploración de los datos se realiza utilizando resúmenes y visualización estadística con ggplot2 (Wickham, 2016) que permite visualización de calidad, flexible y con una teoría basada en la gramática gráfica (Wilkinson, 2006) que lo sustenta.

Para predecir el abandono escolar se utiliza una variable de respuesta categórica con dos niveles (abandona, no abandona). Los modelos predictivos que se presentarán entran dentro de los métodos de aprendizaje supervisado. En aprendizaje supervisado el objetivo es predecir la variable de respuesta basado en un conjunto de predictores. Si la

variable de respuesta es cualitativa, el problema se llama de clasificación, mientras que si es cuantitativa el problema es de regresión (Haste y Tibshirani 2009).

Se presentarán resultados basados en árboles de clasificación (Breiman, et., al, 1984) y bosques aleatorios (Brieman, 2001). Los resultados indican que si se usan los modelos sin considerar el desbalance entre las dos clases se obtiene una tasa de error aceptable (del orden del 5%) pero el error dentro de los que abandonan es muy alto, siendo justamente este el peor error (predecir que no va a abandonar cuando realmente sí lo hace). Si se usan los modelos considerando el desbalance se reduce la tasa de error dentro de los que abandonan a costa de aumentar el error global. Como el número absoluto de alumnos que no abandonan es mucho más grande que el número que abandona la estrategia de usar modelos ponderados reduce el peor error a costa de aumentar mucho el error global (pasa de 5% a cerca de 39%).

Referencias bibliográficas

Breiman, L, Friedman, J, H., Olshen, R. A. Stone, C. J. (1984). Classification and regression trees, Monterey, Calif., USA: Wadsworth, Inc.

Brieman, L. (2001). Random forests, Springer.

Hastie, T & Tibshirani, R., (2009). The elements of statistical learning (2nded), Springer

Wickham, H. (2016). ggplot2: elegant graphics for data analysis, Springer

Wilkinson, L. (2006). The grammar of graphics. Springer Science & Business Media

Palabras Clave: Análisis predictivo, métodos supervisados, estadísticas educativas

Área de Aplicación Educación, Ciencia y Cultura

CARACTERIZACIÓN SOCIO ECONÓMICAS Y ACADÉMICAS DE LOS ESTUDIANTES DE ANÁLISIS MATEMÁTICO I EN LA MODALIDAD A DISTANCIA-AÑO 2017- EN LA FCE¹ DE LA UNRC²

Nancy Scattolini y Silvia Cabrera

Facultad de Ciencias Económicas de la Universidad Nacional de Río Cuarto

nancyscattolini@yahoo.com.ar

Resumen

Este trabajo se enmarca dentro del proyecto de investigación, acreditado y financiado por la Secretaría de Ciencia y Técnica de la UNRC para el período 2016-

¹ Facultad de Ciencias Económicas.

² Universidad Nacional de Río Cuarto.

2018: Análisis de los factores asociados a la deserción y de las estrategias institucionales que promueven la permanencia en las asignaturas del área matemática en el primer año de los tres planes de estudio de la FCE. En esta oportunidad analizaremos los contextos socioeconómicos y los rendimientos académicos de los estudiantes inscriptos en la asignatura Análisis Matemático I (AMI) en la modalidad a distancia en el año 2017. AMI es la primera de las asignaturas en el área matemática del ciclo básico de las tres carreras que ofrece la FCE.

En este contexto institucional, nos preguntamos: ¿Cuál es el perfil socioeconómico y académico de los estudiantes de AMI en la modalidad a distancia? y ¿Cuál es la relación entre esas características individuales y sus rendimientos académicos en la asignatura?.

Estos interrogantes impulsaron a plantearnos los siguientes objetivos que orientan a este trabajo: Explorar las características socioeconómicas y académicas de los estudiantes inscriptos en el año 2017, en la asignatura AMI en la modalidad a distancia de la FCE y estudiar la relación entre esas características individuales y el rendimiento académico en la asignatura.

El enfoque metodológico optado en esta investigación es un estudio de caso cuantitativo, descriptivo, retrospectivo, transversal y observacional. En cuanto a las técnicas y métodos utilizados fueron, la revisión de fuentes documentales oficiales de la Facultad, un análisis descriptivo de las mismas y el Análisis Factorial de Correspondencias Múltiples (AFCM). Este último, es una herramienta idónea para obtener información de grandes volúmenes de datos que describen realidades, y permite encontrar tipología de individuos y asociación de algunas de las variables seleccionadas o características relacionadas con esos individuos. Con respecto a la exploración de las fuentes oficiales, nos permitió identificar aquellas variables y las modalidades asociadas que, según Díaz Peralta (2008), se consideran dentro de los factores individuales, académicos y socioeconómicos que inciden en la deserción estudiantil.

Por su parte, el AFCM se aplicó a los 178 estudiantes inscriptos, y los resultados en términos globales permiten concluir que, el eje factorial 1 relaciona sobre el semieje positivo, los estudiantes inscriptos que no cursaron (libres no comenzó) o no rindieron en ninguna instancia, teniendo la posibilidad de hacerlo (libre por faltas), versus sobre el semieje negativo los que regularizaron y promocionaron. Los libres trabajan entre 20 y 30 hs semanales, la relación de su trabajo con la carrera es parcial o ninguna y algunos

de ellos poseen estudios terciarios o universitarios completos. Los que aprobaron la materia AMI, entre un 50 y 75% no trabaja y algunos estudiantes poseen títulos terciario completo.

En cuanto al segundo eje factorial, relaciona en el semieje positivo los estudiantes que trabajan más de 50 hs semanales, que provienen de establecimientos secundarios con orientación en ciencias sociales y donde su trabajo está totalmente relacionado con la carrera, versus en el semieje negativo se encuentran los estudiantes que trabajan entre 30 y 40 hs semanales, provienen de establecimientos privados, quedaron libres por parcial, y otros estudiantes poseen estudios terciarios completos o universitarios incompletos.

Se sugiere la interpretación de por lo menos el tercer eje factorial, dado que éste y el segundo están asociados a valores propios próximos, representando el segundo eje un poco más de variabilidad. Los valores propios, siguientes al tercero, presentan un decrecimiento casi regular, lo que se traduciría en datos poco estructurados de los que los factores son poco sintéticos.

Finalmente sugerimos, realizar un estudio de Clusters entre las características socioeconómicas y la variable rendimiento académico de los estudiantes de AMI en la modalidad a distancia de la FCE.

Referencias bibliográficas

- Baronio, A. M., & Vianco, A. M. (2015). Análisis de la Información - PIE 5. En A. M. Baronio, & A. M. Vianco, Cuadernos de Econometría (Primera ed., págs. 1-193).
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2004). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de Economía*, 60(60), 39-65.
- Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2007). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de Economía*, 65(65), 9-36.
- Díaz Peralta, C. (2008). Modelo Conceptual Para La Deserción Estudiantil. *Estudios pedagógicos*, 34(2), 65-86.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista, L. (2008). *Metodología de la Investigación*. La Habana: Félix Varela.

- Méndez Ramírez, I. [et al.]. (1996). El Protocolo de Investigación: lineamientos para su elaboración y análisis (2ª ed.). México: Editorial Trillas.
- Montes Gutiérrez, I. C., Almonacid Hurtado, P. M., Gómez Cardona, S., Zuluaga Díaz, F. I., & Tamayo Zea, E. (2010). Análisis de la deserción estudiantil en los programas de pregrado de la Universidad EAFIT. (C. Dirección de Investigación y Docencia Universidad EAFIT Medellín, Ed.) Medellín: Serie Cuadernos de Investigación.
- Peña, D. (2002). Análisis de Datos Multivariantes. Madrid: McGraw-Hill.
- Rodríguez, S., Fita, S., Torrado, M. (2004). El rendimiento académico en la transición secundaria-universidad. Revista de Educación. Temas actuales de enseñanza. V.1 n 334. Recuperado el 07-08-05 de <http://www.revistaeducacion.mec.es/re334.htm>
- Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. Revista de Educación Superior, 71, 33-51.
- Vásquez Velásquez, J., Castaño, E., Gallón Gómez, S., & Gómez Portilla, K. (2003). Determinantes de la deserción estudiantil en la Universidad de Antioquia. (4), 1-38.
- Palabras Clave:** Educación Superior- Educación a Distancia- Estudiantes - Caracterización socio económica y académica- Análisis Factorial de Correspondencias Múltiples.
- Área de Aplicación:** Educación, Ciencia Y Cultura
-

VALIDACION DE INSTRUMENTOS DE MEDICION MULTIDIMENSIONALES

Norma Leonor Rodríguez ¹, Gladys del V. Rosales ², Jesús Alvarado Izquierdo ³.

1. Facultad de Ciencias Exactas y Naturales – UNCa. Catamarca - Argentina.

2. Facultad de Ciencias Económicas- UNCa. Catamarca - Argentina.

3. Facultad de Psicología. Universidad Complutense de Madrid- España

norleor@yahoo.com.ar

Resumen

En las últimas dos décadas se han evidenciado esfuerzos importantes en diferentes países y organismos internacionales por precisar la calidad de los cuestionarios

utilizados en investigación especialmente en Ciencias Sociales, Salud, Psicología, entre otras ciencias. Los test, escalas, pruebas, cuestionarios y otros instrumentos similares forman parte de la actividad cotidiana de investigadores; en ellos diferentes características de los individuos, los estados, la salud, entre otros se evalúan periódicamente.

Los cuestionarios pueden ser una herramienta potente y eficaz para llegar a conocer los distintos aspectos involucrados en la investigación, aunque para ello se precisa una metodología adecuada en su diseño y administración; para que la calidad métrica del instrumento de medición permita confiar en los resultados obtenidos (Valencia Martín, 2006). Es así que el uso de cuestionarios en las encuestas basan su información en la validez de los datos proporcionados por los entrevistados respecto a percepciones, sentimientos, actitudes o conductas que transmite el encuestado, siendo ella, en muchos casos, difícil de contrastar y traducir a un sistema de medidas - a una puntuación-. Es esta característica la que hace tan complejo establecer los criterios de calidad de ellos (Martín Arribas, 2004).

Cabe señalar que ante la necesidad de usar cuestionarios surgió una variedad de estudios de validación de escalas, que en ocasiones demasiado superficiales se califican como procesos de “validación”, conduciendo a evaluar deficientemente la validez y fiabilidad del instrumento (Batista Foguet, Coenders Gallart & Alonso, 2004). Pero en la actualidad, con los avances metodológicos y computacionales, el investigador está obligado a conocer y utilizar técnicas apropiadas para el estudio de las relaciones entre variables. Pues los fenómenos de interés son complejos, tienen varios aspectos, que obedecen a múltiples causas y están frecuentemente medidos con error; de allí que identificar el origen de su variabilidad requiere utilizar técnicas multivariadas adecuadas (Grace, 2006; Byrne, 2006).

Pero no hay métodos ni modelos estándar que se apliquen siempre y en todos los casos si realmente se quieren obtener evaluaciones y mediciones válidas, especialmente cuando se utiliza un cuestionario multidimensional. Por lo tanto, se plantea como interrogante: ¿cómo se determina si un cuestionario cumple criterios mínimos de calidad métrica, para ser utilizados como instrumentos científicos en la investigación?

Para dar respuesta a dicha pregunta se plantea como objetivo principal en el presente trabajo: Establecer una propuesta de criterios metodológicos de validación de cuestionarios multidimensionales. Cuyos objetivos específicos pretenden: indagar los

diferentes criterios metodológicos para validar instrumentos de medición multidimensional y/o multipoblacional; y analizar las técnicas psicométricas y estadísticas multivariadas aplicables a este tipo de cuestionarios

El enfoque de la investigación se circunscribe en mixto-descriptivo. Según Striner y Norman (2004) para evaluar la validez, con técnicas cualitativas y cuantitativas, se deben considerar diferentes tipos de validez: *contenido, criterio, constructo e interna o confiabilidad*. En ellas coexisten esencialmente dos planteamientos estadísticos: uno, enraizado en las Teorías Psicométricas -validez de contenido, criterios- y el otro en Teorías de Modelos Multivariados -validez constructo y fiabilidad- (Martínez Arias et al, 2006; Johnson Dallas, 2000).

De los resultados de la investigación se expresa la relevancia que tiene en la validación de instrumentos de medición los criterios de transculturalización, prueba piloto, clasificación de las variables, estructura del cuestionario -sumativo o no sumativo, unidimensional o multidimensional- y la cantidad de poblaciones involucradas en el estudio. Justamente estos aspectos complejizan las estrategias a seguir para determinar los criterios en el proceso de validación, con el fin de aplicar procedimientos de forma correcta y rigurosa. Se considera como ejemplo la validación de un cuestionario Socio-Sanitario en el Gran Catamarca (Rodríguez, 2018). Luego se presenta una propuesta de criterios metodológicos de validación de cuestionarios multidimensionales y/o multi-poblacionales aplicables en las áreas de Ciencias Sociales y Humanas, y transferibles a otros campos de aplicación.

Referencias bibliográficas

- Byrne, B. M. (2006). *Structural Equation Modeling with EQS. Basic concepts, applications, and programming* (2nded.). Mahwah, NJ, Erlbaum, 3:117.
- Batista Foguet, J. M., Coenders Gallart, G. & Alonso, J. (2004). Análisis Factorial Confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la Salud. *Medicina Clínica*, 122 (Supl. 1): 21-27.
- Grace, J.B.(2006). *Structural Equation Modeling and Natural Systems*. United States of America by Cambridge University Press. New York: 143-323.
- Johnson Dallas, E. (2000). *Métodos Multivariados aplicados al análisis de datos*. México: International Thomson Editores S.A.: 15-54; 147-213.

Martín Arribas, M.C. (2004). Diseño y validación de cuestionarios. *Matronas Profesión*. 5(17): 23-29. Madrid. Recuperado el 10 de marzo de 2016 de: http://www.enferpro.com/documentos/validacion_cuestionarios.pdf

Martínez Arias, R.; Hernández, M & Hernández, M. (2006). *Psicometría*. Madrid: Alianza: 15-100.

Rodríguez, N.L. (2018). Modelos de Ecuaciones Estructurales en la validación de un cuestionario Socio-Sanitario en el Gran Catamarca. Tesis Doctoral del Doctorado en Ciencias Mención Matemática. FACEN- UNCa. Catamarca. Argentina.

Striner, D.L. and Norman, G.R. (2002). *Health Measurement Scales. A Practical Guide to their Development and use*. Third Edition. Oxford University Press. Canada: 126-151; 172-192, 248-253.

Valencia Martín, J. (2006). Análisis y validación de un cuestionario para medir la satisfacción de los usuarios en Atención Primaria. Toledo. España.

Palabras Clave: Validación – Cuestionarios – Técnicas - Psicométricas – Estadísticas.

Área de Aplicación: Otras Ciencias Sociales y Humanas.

Métodos robustos.



ROBUSTEZ EN ENSAMBLES DE MODELOS LINEALES REGULARIZADOS

Jorge Adrover⁽¹⁾, Marcelo Ruiz⁽²⁾, Rubén Zamar⁽³⁾

(1) FaMAF (Universidad Nacional de Córdoba), CIEM y CONICET, Argentina

(2) Facultad de Ciencias Exactas, Físico-Químicas y Naturales, Universidad Nacional de Río Cuarto

(3) Department of Statistics. University of British Columbia, Canadá.

adrover@famaf.unc.edu.ar

Resumen

Christides et al (2017) construyen ensambles de modelos lineales regularizados, minimizando una función objetivo que incluye una suma de residuos más un término que trata de penalizar variables en modelos raros con escaso número de coeficientes no nulos más otro término que controla la determinación de los distintos grupos de variables que configuran los ensambles que mejoran la predicción de la variable respuesta de interés. Más precisamente, $y \in \mathbf{R}^n$ es el vector de la variable respuesta observada y $\chi^t = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbf{R}^{p \times n}$ es una matriz de p características o variables medidas asociadas a cada una de las n valores de las variables respuestas. Si se plantea que hay un número desconocido $G > 1$ de submodelos o ensambles de variables, entonces la función objetivo a minimizar está dada por

$$O(y, X, \beta^1, \dots, \beta^G) = \sum_{g=1}^G \left[\frac{1}{2n} \sum_{i=1}^n (y_i - \mathbf{x}_i \beta^g)^2 + p_{\lambda_s}(\beta^g) + q_{\lambda_D, g}(\beta^1, \dots, \beta^G) \right]$$

Donde p_{λ_s} es una función de penalización para producir modelos raros (o sea cuando hay un número escaso de parámetros no nulos) y $q_{\lambda_D, g}$ es una función de penalización para producir diversidad entre los G modelos subyacentes. Más precisamente, se toma

$$p_{\lambda_s}(\beta) = \lambda_s \left(\frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \right), \alpha \in [0, 1],$$

$$q_{\lambda_D, g}(\beta^1, \dots, \beta^G) = \frac{\lambda_D}{2} \sum_{h \neq g} |\beta_j^h \beta_j^g|,$$

donde α , λ_s , λ_D , G son parámetros que se van ajustando en el proceso de estimación.

El problema que consideramos es una versión robusta de este procedimiento y sus propiedades. La formulación sería la siguiente. Dado $b \in (0, 1)$ y $\rho : \mathbb{R} \rightarrow [0, 1]$ una función tal que

- (i) ρ es par, continua y acotada;
 - (ii) ρ es monótona creciente con $\rho(0) = 0$ y $\lim_{x \rightarrow \infty} \rho(x) = 1$
 - (iii) si $\rho(v) < 1$ y si $0 \leq u < v$ entonces $\rho(u) < \rho(v)$, entonces, si ρ_1 y ρ_2 satisfacen (i),
- (ii) y (iii), sea $s_n(\mathbf{r}(\beta^g))$ M -escala determinada por la resolución de la ecuación,

$$\frac{1}{n} \sum_{i=1}^n \rho_1 \left(\frac{(y_i - x_i \beta^g)^2}{s_n(\mathbf{r}(\beta^g))} \right) = b$$

Entonces se define la función objetivo a minimizar como

$$O^R(y, \chi, \beta^1, \dots, \beta^G) = \sum_{g=1}^G \left[\frac{s_n^2(\mathbf{r}(\beta^g))}{2n} \sum_{i=1}^n \rho_2 \left(\frac{(y_i - x_i \beta^g)^2}{s_n(\mathbf{r}(\beta^g))} \right) + p_{\lambda_s}(\beta^g) + q_{\lambda_{D,g}}(\beta^1, \dots, \beta^G) \right]$$

La selección de los parámetros $\alpha, \lambda_s, \lambda_D, G$ se realiza siguiendo los lineamientos de Christides et al (2017). En este trabajo se realiza un estudio por simulación para evaluar el comportamiento predictivo de este estimador y otros competidores naturales en determinados escenarios de contaminación y diferentes tamaños muestrales (ver Smucler and Yohai (2017) para una alternativa regularizada robusta y otros competidores tales como MM-and adaptive MM- sparse estimator, sparse LTS, robust elastic net, etc.). Se estudia la consistencia del estimador propuesto y los resultados de simulación son analizados y comparados.

Referencias bibliográficas

[1] Smucler, E. and Yohai, V.J. (2017). Robust and sparse estimators for linear regression models. Computational Statistics and Data Analysis. 111: 116-130.

[2] Christidis, A., Lakshmanan, L V.S., Smucler, E., and Zamar, R.H. (2018). Ensembles of regularized linear models. Available at <https://arxiv.org/pdf/1712.03561.pdf>.

Palabras Clave: Penalización - Regresión - Predicción - Ensembles.

Área de Aplicación: Métodos Robustos - Ciencias Exactas y Naturales.

SM-ESTIMADORES PARA ANÁLISIS CANÓNICO

Donato, Stella Maris ⁽¹⁾ y Adrover, Jorge Gabriel ⁽²⁾

(1) Universidad Nacional de Cuyo

(2) FaMAF, Universidad Nacional de Córdoba, CIEM y CONICET

stelladonato@yahoo.com.ar

Resumen

En este trabajo se estudian las propiedades de una metodología robusta para computar vectores y correlaciones canónicas. El análisis de correlaciones canónicas (CCA) es una técnica de reducción de la dimensión en la cual dos vectores aleatorios son reducidos a vectores de dimensión más baja tras aplicar transformaciones lineales a cada uno de ellos, siendo las componentes de los vectores transformados las llamadas variables canónicas. Es conocido en la literatura que, como los nuevos vectores transformados minimizan la distancia cuadrada esperada entre ellos, que es una medida altamente sensible a observaciones atípicas, entonces los vectores y correlaciones canónicas usuales son también afectados. Para subsanar este problema, Adrover and Donato (2015) propusieron S-estimadores robustos por minimización de una M-escala basada en la distancia de transformaciones lineales de los vectores. Esta aproximación predictiva fue utilizada por Maronna (2005) al introducir S-estimadores para componentes principales.

El procedimiento se define a continuación. Dados vectores $\mathbf{x} = (x_1, \dots, x_p)^t$ e $\mathbf{y} = (y_1, \dots, y_q)^t$, con matrices de covarianzas Σ_{xx} y Σ_{yy} respectivamente, los SM-vectores canónicos estandarizados robustos se definen como $(A_o^{(SM)}, B_o^{(SM)}, \mathbf{a}_{SM}) = \arg \min_{((A|B, \mathbf{a}) \in B_{r,m}} \sigma(A, B, \mathbf{a})$, donde $r \leq \min(p, q)$ denota la cantidad de vectores canónicos a estimar, $B_{r,m} = \{((A|B, \mathbf{a}) : \mathbf{a} \in \mathbb{R}^r, A \in \mathbb{R}^{r \times p}, B \in \mathbb{R}^{r \times q}, AA^t = I_r = BB^t\}$ y $\sigma(A, B, \mathbf{a})$ es una M-escala definida implícitamente a partir de la ecuación

$$\sigma(A, B, \mathbf{a}) = \min \left\{ \sigma > 0 : E_{\tilde{F}} \left(\chi \left(\frac{\left\| A(\Sigma_{xx}^{(R)})^{-1/2} \mathbf{x} - B(\Sigma_{yy}^{(R)})^{-1/2} \mathbf{y} - \mathbf{a} \right\|}{\sigma} \right) \right) \leq \delta \right\},$$

donde $\Sigma_{xx}^{(R)}$ y $\Sigma_{yy}^{(R)}$ son estimadores robustos de Σ_{xx} y Σ_{yy} , $0 < \delta < 1$, $\chi: [0, \infty) \rightarrow [0, 1]$ es una función monótona no decreciente, continua por izquierda en $(0, \infty)$ y continua en 0 , $\chi(0) = 0$, $\lim_{x \rightarrow \infty} \chi(x) = 1$ y $\chi(x) > 0$ si $x > 0$. \tilde{F} denota la distribución conjunta de $\tilde{x} = \left(\Sigma_{xx}^{(R)}\right)^{-1/2} x$ e $\tilde{y} = \left(\Sigma_{yy}^{(R)}\right)^{-1/2} y$. En consecuencia, los SM-vectores canónicos robustos se definen como $A_{SM} = A_0^{(SM)} \left(\Sigma_{xx}^{(R)}\right)^{-1/2}$ y $B_{SM} = B_0^{(SM)} \left(\Sigma_{yy}^{(R)}\right)^{-1/2}$. Adrover and Donato (2015) presentaron un algoritmo de cómputo y propiedades de convergencia del mismo, así como estimadores robustos para correlaciones canónicas, comparando el comportamiento de los estimadores respecto de otras propuestas en un estudio de simulación, en distintos escenarios de contaminación y tamaños de muestras.

El objetivo de este trabajo es estudiar las propiedades estadísticas del estimador SM. Para analizar su comportamiento asintótico se demuestran la consistencia en el sentido de Fisher, consistencia, distribución asintótica normal bajo distribución elíptica de la muestra aleatoria y eficiencia relativa asintótica. Se prueban propiedades de robustez del estimador tales como punto de ruptura del estimador de escala asociado al estimador, robustez cualitativa y función de influencia. Finalmente, se muestra el comportamiento del estimador sobre un conjunto de datos reales estudiado en Alam *et al.* (2010), exhibiendo el mismo el mejor comportamiento predictivo respecto de otros estimadores utilizados al ser evaluados a través de validación cruzada.

Referencias bibliográficas

- [1] Adrover, J. G. y Donato, S. M. (2015). A robust predictive approach for canonical correlation analysis. *Journal of Multivariate Analysis*. 133: 356-376.
- [2] Maronna, R. A. (2005). Principal components and orthogonal regression based on robust scales. *Technometrics* 47(3): 264-273.
- [3] Alam, A., M. Nasser and K. Fukumizu. (2010). A Comparative Study of Kernel and Robust Canonical Correlation Analysis. *Journal of Multimedia*. 5(1):3-11.

Palabras Clave: Análisis Multivariado – Análisis Canónico – Estimadores robustos.

Área de Aplicación: Ciencias Exactas y Naturales

Modelos de regresión.



CALIBRACIÓN COMPARATIVA ROBUSTA

Patricia Giménez⁽¹⁾, Lucas Guarracino⁽¹⁾, Manuel Galea⁽²⁾

(1) Departamento de Matemática, FCEyN, Universidad Nacional de Mar del Plata

(2) Departamento de Estadística, Pontificia Universidad Católica de Chile

pcgimene@mdp.edu.ar

Resumen

La comparación de instrumentos o métodos de medición, cuando cada uno de ellos es usado para medir una característica común en un grupo de unidades experimentales, se conoce en la literatura estadística como calibración comparativa. La necesidad de comparar instrumentos de medición que difieren en costo, velocidad u otros factores, aparece con frecuencia en diversas áreas de la ciencia, como química, biología y medicina, entre otras.

El modelo de calibración comparativa que consideramos en este trabajo puede verse como un caso especial de un modelo de regresión lineal con errores en las variables, funcional. La covariable no observada, unidimensional, corresponde a la verdadera característica de interés, fija, desconocida. Asumimos que esta característica es medida sin sesgo por un instrumento patrón y varios instrumentos alternativos, que poseen cada uno un sesgo aditivo y un sesgo multiplicativo, siendo estas cantidades los principales parámetros de interés o parámetros estructurales del modelo. La verdadera característica desconocida, en cada sujeto, constituye una sucesión de parámetros incidentales cuyo número aumenta con el tamaño muestral.

Diferentes metodologías de estimación, incluyendo máxima verosimilitud, score corregido, score condicional y método de momentos han sido considerados en la literatura para este modelo bajo el supuesto de normalidad (Kimura, 1992; Bolfarine y Galea-Rojas, 1995 y Giménez y Bolfarine, 2000). No obstante, es sabido que los estimadores obtenidos mediante estos enfoques son sensibles a la presencia de observaciones atípicas y/o desviaciones de los supuestos del modelo.

En este trabajo proponemos un procedimiento de estimación paramétrico robusto, basado en una extensión del enfoque de máxima L_q -verosimilitud (Ferrari y Yang, 2010, Ferrari y La Vecchia, 2012). La metodología está relacionada a través de una transformación del parámetro, con la minimización de una medida de divergencia cuasi logarítmica (dependiendo de una constante real q) entre la verdadera densidad de los datos y la asumida como modelo paramétrico. Para $0 < q < 1$, proporciona una

alternativa útil de estimación robusta de los parámetros estructurales del modelo funcional, caracterizado por la presencia de observaciones independientes pero no idénticamente distribuidas. El procedimiento propone reemplazar los parámetros incidentales en la L_q -verosimilitud por estimadores que dependen de los parámetros estructurales y requiere de una transformación apropiada para obtener un estimador consistente de los parámetros de interés. El estimador se adapta de acuerdo a la discrepancia entre los datos y el modelo asumido a través del valor de q , que controla el balance entre eficiencia y robustez. El estimador de máxima verosimilitud es recuperado como caso particular cuando $q=1$.

Para q fijo, bajo condiciones de regularidad apropiadas probamos consistencia y normalidad asintótica de los estimadores y estudiamos la robustez infinitesimal a partir del cálculo de la función de influencia. Los estimadores propuestos muestran buenas propiedades de robustez. El desempeño de los estimadores se analiza a partir de resultados asintóticos, estudios de simulación y aplicaciones a conjuntos de datos reales. En la práctica, proponemos la selección de un valor apropiado de la constante q , como aquel que minimiza el error cuadrático medio del estimador calculado mediante *bootstrap* paramétrico.

Referencias bibliográficas

- Bolfarine, H. y Galea-Rojas, M. (1995). Comments on “Functional comparative calibration” (by D. Kimura). *Biometrics*, 51, 1579-1580.
- Ferrari, D. y Yang, Y. (2010). Maximum L_q -likelihood estimation. *Annals of Statistics*, 38, 753-783.
- Ferrari, D. y La Vecchia, D. (2012). On robust estimation via pseudo-additive information. *Biometrika*, 99, 238-244.
- Giménez, P. y Bolfarine, H. (2000). Comparing consistent estimators in comparative calibration models. *Journal of Statistical Planning and Inference*, 86, 143-155.
- Kimura, D.K. (1992). Functional comparative calibration using an EM algorithm. *Biometrics*, 48, 1263-1271.

Palabras Clave: Calibración comparativa – Modelos de regresión con errores en las variables – Estimación de máxima L_q -verosimilitud – Robustez.

Área de Aplicación: Ciencias Exactas y Naturales.

Muestreo

A horizontal green arrow pointing to the right, with a white outline of a second arrow inside it, creating a double-arrow effect.

UNA PROPUESTA DE CURSO DE MUESTREO CON R

Mariela Gallardo, Augusto E. Hoszowski, Federico Ramirez

Universidad Nacional de Tres de Febrero

aehoszowski@gmail.com

Resumen

Problemática

Como sucede en otras disciplinas con una fuerte aplicación, la enseñanza del Muestreo conlleva la necesidad de transmitir a los alumnos los conceptos teóricos fundamentales y el manejo de las técnicas necesarias para el desempeño profesional. Dependiendo de la orientación de la carrera, será el volumen relativo de cada uno de estos dos ingredientes. Pero aún en una disciplina tan 'aplicada' como el Muestreo es necesario un mínimo de bagajes teóricos para un exitoso desempeño profesional. El desarrollo de la informática cambió en poco tiempo las exigencias hacia los estadísticos. Mientras que hasta hace pocos años se requería que los estadísticos supieran *interpretar* las salidas de los primeros software estadísticos (BMDP, SAS, SPSS) y programar simplemente algunos *test* o modelos, con el auge del 'análisis de datos' el estadístico debe sumar a sus conocimientos básicos un manejo informático fluido que le permita la lectura de datos en diferentes formatos, su depuración y tratamiento. Aunque estas competencias se adquieren muchas veces en el trabajo profesional, es un 'plus' a la hora de acceder a un puesto de analista de datos ya poseerlas. Esto obliga a las universidades a actualizar sus programas de estudio. R, de *standard* dentro del ámbito académico pasó a ocupar un lugar central en el análisis de datos en la actividad privada, por lo que su valoración ha aumentado significativamente en el mercado. Actualmente no solo son deseables las habilidades y conocimientos para escribir *scripts* sencillos y funciones sino conocer algunos de los paquetes más utilizados, tanto para la manipulación de datos como para la graficación y adquirir la práctica de buscar en la red soluciones a los problemas que se puedan plantear. La gran cantidad de paquetes que R posee, uno de sus puntos fuertes, es también fuente de desorientación para los principiantes.

Objetivo del Trabajo

El presente trabajo expondrá aspectos del Curso de Muestreo obligatorio para los estudiantes de la Licenciatura en Estadística de la Universidad Nacional de Tres de Febrero y algunos de los trabajos prácticos realizados por los alumnos. La presentación se focalizará en las prácticas donde se aplicó R con cierto nivel de complejidad para un

primer curso, o donde se utilizaron paquetes de R actualmente muy utilizados por los analistas de datos, omitiéndose los temas más clásicos de un curso de muestreo. Como es inevitable, el incorporar nuevas competencias obliga siempre a dejar en parte temas que se dictaban. Esto ocurre en la mayoría de las disciplinas. Una alternativa es dejar de demostrar en el curso ciertos puntos teóricos, pero ahondando en ejercicios de simulación, lo que puede redundar en una mejor comprensión de las ideas del muestreo, así como ir ejercitando a los estudiantes en la escritura de programas R y la utilización de algunos de sus paquetes (dplyr, ggplot y png).

Referencias bibliográficas

- G. Cochran. (1997) Técnicas de Muestreo (4ta edición). México: CECSA.
- Y. Tillé. Dunod (2001) Théorie des Sondages.
- C. Sarndal, B. Swenson, J. Wretman (1992). Model Assisted Survey Sampling.
- T. Lumley (2010). Complex Surveys. Wiley.
- Y. Aragon, D. Haziza A. Ruiz-Gazen (2005). Les simulations dans l'enseignement des sondages. Revue MODULAD,86- Nro 32
- C. Goga and A. Ruiz-Gazen (2010) The use of Monte Carlo Simulations in teaching survey sampling. Proceedings of the Eighth International Conference on Teaching Statistics. Ljubljana, Slovenia.

Palabras Clave: Enseñanza - Situaciones de aprendizaje – Estadística – Muestreo - Programación.

Área de Aplicación: Enseñanza de la Estadística.

DETERMINACIÓN DEL TAMAÑO DE UNA MUESTRA Y ASIGNACIÓN EN PRESENCIA DE ERRORES DE CLASIFICACIÓN

Gregorio García, Isabel Alegre, Mariana Mendiburu

Dirección Nacional de Metodología Estadística, INDEC

ggarcia@indec.gob.ar, ialegre@indec.gob.ar, mmendiburu@indec.gob.ar

Resumen

En el año 2017 el INDEC establece la necesidad de llevar adelante un programa de renovación integral de sus encuestas para el sector económico. El mismo debe reconstruir y/o actualizar los principales indicadores económicos a través del empleo combinado de encuestas a empresas o unidades económicas (UE) y de registros fiscales

o administrativos en los principales sectores de la economía (Industria, Servicio y Comercio).

Las UE se clasifican según su Rama de Actividad con distintos niveles de agregación según el Clasificador Industrial Internacional Uniforme (CIIU). El objetivo de las encuestas es lograr estimaciones con una precisión predefinida principalmente en dos dimensiones o dominios en forma simultánea: (a) a nivel país donde se la requiere para cada rama de actividad a un nivel desagregado de 4 o 5 dígitos del clasificador, y (b) a nivel provincial donde la precisión es requerida a un nivel menor de desagregación, por lo general a 2 o 3 dígitos según la provincia y en sus actividades principales dentro de las mismas.

La rama de actividad en la que se registra una UE proviene de la propia declaración de la UE hecha a la Agencia Federal de Ingresos Públicos (AFIP) cuando se inscriben al inicio de sus actividades. Este y otros motivos pueden llevar a que actualmente se disponga de una clasificación errónea de su rama de actividad, por ejemplo el cambio de su actividad principal posterior a la fecha de inscripción, interpretación errónea de la rama propuesta por el aplicativo de AFIP. Estos errores plantean una dificultad al momento de determinar un tamaño de muestra y de asignación para las encuestas, ya que los diseños suelen ser estratificados e involucran en la definición de sus estratos y/o dominios a las Ramas de Actividad.

Los métodos tradicionales para dimensionar una muestra en los dominios en cuestión no resultan satisfactorios al estar presentes estos errores de clasificación, que afectan y deterioran la precisión de los estimadores calculados sobre las muestras surgidas de dichos métodos.

Piesse y Lohr (2016) plantean una alternativa para obtener un tamaño de muestra que logra la precisión deseada en cada dominio a pesar de la presencia de errores de clasificación. Su método se basa un diseño muestral en dos fases que involucra a las probabilidades de clasificación errónea de las unidades y a la determinación de los tamaños de muestra como solución a un problema de optimización no lineal, con restricciones que involucran entre otros factores, la precisión deseada para los estimadores en cada dominio y/o dimensión.

Este método es motivo de estudio y evaluado como alternativa al tradicional dentro del programa de renovación señalado, buscando encontrar tamaños de muestra

que satisfagan las restricciones impuestas al diseño teniendo en cuenta los errores de clasificación para las encuestas económicas del Instituto.

La metodología propuesta fue implementada en los software estadísticos SAS v9.4. a través del PROC OPTMODEL y R logrando resultados satisfactorios. El tamaño de muestra asignado por la nueva metodología resulta un 8% mayor respecto al tamaño requerido de no haber existido errores de clasificación, logrando la precisión prefijada en el dominio país con un modesto impacto en el tamaño muestral. Al asignar esta muestra mediante asignación óptima se logra la precisión predefinida para las estimaciones en un 90% de las ramas y provincias en las variables evaluadas.

La metodología continúa en etapa de desarrollo e investigación, y se pretende incorporar otros aspectos como involucrar en los cálculos tasas de no respuesta diferenciales previstas por dominio y/o estrato en la determinación de los tamaños de muestra.

Referencias bibliográficas

- Bethel, J. (1989), Sample Allocation in Multivariate Surveys, *Survey Methodology*, 15(1), 47-57.
- Kish, L. (1992). Weighting for Unequal Pi, *Journal of Official Statistics*, 8(2), 183–200.
- Lohr, S., y Piesse, A. (2016). Sample Design for Longitudinal Multiphase Samples with Misclassification, *Joint Statistical Meetings 2016 Survey Research Methods Section, Chicago*.
- Rivest L. P., Baillargeon S. (2017). Package ‘stratification’. Recuperado de: <https://CRAN.R-project.org/package=stratification>
- Valliant, R., Dever, J. A., Kreuter, F. (2013), *Practical Tools for Designing and Weighting Survey Samples*, Nueva York, Estados Unidos: Springer.

Palabras Clave: Diseño óptimo – Muestreo estratificado - Optimización restringida.

Área de Aplicación: Estadísticas Oficiales.

IMPUTACIÓN DE LA NO RESPUESTA UTILIZANDO EL PROCEDIMIENTO RANDOM FOREST

**Javier Bussi ⁽¹⁾, Lucía Hernández ⁽¹⁾, Gonzalo Marí ⁽¹⁾, Fernanda Méndez ⁽¹⁾,
Gerardo Mitas ⁽²⁾**

(1) Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística,
Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

(2) Dirección Nacional de Metodología Estadística, Instituto Nacional de Estadística y
Censos

jbussi@fcecon.unr.edu.ar

Resumen

Una de las mayores dificultades que ocurren en las encuestas es la presencia de no respuesta, la cual puede ser total o parcial. En el primer caso, una de las soluciones posibles sería el ajuste de los ponderadores muestrales, existiendo múltiples alternativas tales como el uso de post-estratificación o de la calibración de los pesos.

En el segundo caso, la imputación constituye una alternativa posible para atenuar el problema de la no respuesta. Entre los procedimientos que usualmente se utilizan, se pueden mencionar los métodos probabilísticos, como el hot-deck, que producen valores imputados distintos en cada repetición del proceso de imputación, o los métodos determinísticos, como por ejemplo utilizando modelos de regresión, los cuales producen los mismos valores ante la repetición del proceso. En ambos casos, es usual que estos métodos sean aplicados en grupos de imputación compuestos por unidades con características similares. Los mismos pueden surgir a partir del uso de variables de clasificación, por ejemplo características demográficas en el caso de personas, o determinados por métodos de clasificación basados en procedimientos que dependen del tipo de variables que se incluyen en el análisis. Un ejemplo de estos métodos se da con el uso de árboles de clasificación o de regresión.

El objetivo es describir la aplicación de un método alternativo de imputación, que utiliza el procedimiento Random Forest. Este método consiste en un proceso iterativo que asigna valores iniciales a los datos perdidos, construye un Forest ajustado el cual permite predecir nuevos datos imputados para cada una de las variables, y repite este procedimiento hasta su convergencia. Entre las principales ventajas de esta metodología, se encuentra el hecho que la misma permite trabajar en forma simultánea tanto variables cualitativas como cuantitativas. Se presentará el método y se describirán

sus principales características a través de la ejemplificación en un conjunto de datos.

Referencias bibliográficas

Breiman, L. (2001). Random Forest. *Machine Learning*, 45, 5-32.

Loh, W.Y., Eltinge, J., Cho, M., Li, Y. (2017). Classification and regression trees and forests for incomplete data from sample surveys. *Statistica Sinica*, Preprint SS-2017-0225, 1-33.

Stekhoven, D.J., Bühlmann, P. (2012). MissForest non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28, 112-118.

Tang, F., Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining*, 10, 363-377.

Palabras Clave: Imputación - Random Forest - Clases de Imputación - Hot Deck

Área de Aplicación: Estadística Oficial

DISEÑO MUESTRAL PARA UNA MUESTRA MAESTRA DEL ÁREA RURAL: (MMR- INDEC)

Fernanda Bonifazi¹, Sara Novaro¹, Gabriela Barbará²

¹Dirección Nacional de Metodología Estadística, INDEC. ²

Coordinación de Muestreo, INDEC.

fbonifazi@indec.gob.ar, gbarbara@indec.gob.ar

Resumen

Actualmente el Instituto Nacional de Estadísticas y Censos (INDEC) cuenta con la Muestra Maestra Urbana de Viviendas de la República Argentina (MMUVRA), a partir de la cual es posible seleccionar muestras de viviendas para aglomerados de 2000 habitantes o más. Ante la necesidad de extender la Encuesta Permanente de Hogares al Ámbito Rural se plantea como objetivo el desarrollo de un marco muestral y cartográfico adecuado para la extensión de encuestas en áreas rurales. El mismo permitirá la selección de una Muestra Maestra para el área Rural (MMR) y así sostener la selección de viviendas para las encuestas a la población rural del país.

Las áreas rurales comprenden viviendas particulares dispersas o agrupadas en pequeñas localidades o poblados con menos de 2000 habitantes, según los datos del último Censo Nacional de Población y Vivienda 2010 (CNPV2010). Para el diseño al

territorio Nacional se lo dividió en dominios de análisis compuesto por las cinco regiones geográficas: Centro, Cuyo, Noreste, Noroeste y Patagonia.

El diseño muestral involucra dos etapas de selección probabilísticas. En la primera etapa, las unidades primarias de muestreo (UPM) están definidas en base a la unidad política territorial departamento o partido, o un grupo contiguo de estos. Los mismos fueron seleccionados con probabilidad proporcional a la cantidad de viviendas rurales ocupadas y en construcción dentro de su dominio o región. En esta etapa se empleó una selección controlada, balanceando las cantidades de UPM a seleccionar por clase, provincia y estrato, en cada región. Luego, se realizó una selección aleatoria de una UPM por estrato de acuerdo al resultado del balanceo planteado.

En la segunda etapa, se definieron las unidades secundarias de muestreo (USM) dentro de cada UPM seleccionadas (UPS), a base de uno o más radios censales rurales según el CNPyV2010, o partes de estos. Los mismos fueron seleccionados con probabilidad proporcional a la cantidad de viviendas ocupadas y en construcción en el área incluida en su configuración. Las USM fueron estratificadas y seleccionadas en cada provincia por separado. Para ello, se aplicó un procedimiento de selección controlada similar al utilizado para las UPM, en este caso controlando el balance por clase o estrato y UPS, dentro de cada provincia. Una vez establecidas las cantidades de USM a seleccionar por clase, una muestra de estas unidades es seleccionada aleatoriamente mediante el procedimiento de selección sistemática, dentro de cada UPS.

Las unidades de última etapa fueron definidas como segmentos o pequeñas áreas con viviendas próximas, dentro de cada US seleccionada (USS). En zonas rurales de población dispersa, cada segmento es un área que contiene en promedio 4 viviendas próximas; en las zonas rurales agrupadas, es una manzana o sector completo, o grupo de estos, o lados completos de manzana/sector con aproximadamente 6 viviendas (entre 4 y 8). El trabajo de segmentación se realiza en gabinete utilizando imágenes satelitales y el sistema de información geográfica del Instituto.

La totalidad de segmentos de cada USS y por lo tanto de cada UPS, configuran la MMR, a partir del cual se seleccionarán aleatoriamente segmentos con las viviendas para las encuestas a realizarse en áreas rurales según las necesidades INDEC.

Referencias bibliográficas

Hansen, Morris H.; Hurwitz, William N.; Madow, William G (1953), *Sample Survey Methods and Theory Volume I and II*, Nueva York, Londres: John Wiley & Sons

Kish, Leslie (1965), *Survey Sampling*, Nueva York, Londres, Sidney: John Wiley & Sons

Särndal, Carl Erik, Swensson, Bengt, Wretman, Jan (1992), *Model Assisted Survey Sampling*, Nueva York, Estados Unidos: Springer.

Tiwwari, Neeraj; Chilwal, Akhil (2014), *A Simplified Approach for Two-Dimensional Optimal Controlled Sampling Designs*, Hindawi, *Advances in Statistics*, Volume 2014, ArticleID 875352.

Valliant, R., Dever, J. A., Kreuter, F. (2013), *Practical Tools for Designing and Weighting Survey Samples*, Nueva York, Estados Unidos: Springer.

Palabras Clave: Diseños Muestrales Complejos - Selección Controlada - Muestreo Multietápico - Población Rural

Área de Aplicación: Estadísticas Oficiales

Probabilidad y Procesos estocásticos

A decorative graphic element consisting of a light green arrow pointing to the right, with a white outline and a white arrowhead.

METODOLOGÍAS ESTOCÁSTICAS PARA LA MEDICIÓN DEL RIESGO DE MERCADO EN ORGANIZACIONES BANCARIAS

Martín Ezequiel Masci, Leonardo Andrés Dufour, Javier Ignacio García Fronti,

María Teresa Casparri

Universidad de Buenos Aires – Facultad de Ciencias Económicas – Centro de
Investigación en Metodologías Básicas y Aplicadas a la Gestión (CIMBAGE-
IADCOM)

martinmasci@economicas.uba.ar

Resumen

En la última década se han suscitado una serie de acontecimientos que han puesto a prueba la estabilidad del sistema financiero a nivel mundial. Los mismos se presentaron como fenómenos de carácter intrínseco y no solo produjeron inmediata reacción –crisis- sino que vulneraron fuertemente la sustentabilidad de las instituciones a largo plazo. En particular, la última crisis financiera (2007/2008) ha mostrado un conjunto de falencias del sistema bancario que se derrama en la estructura de las organizaciones e impacta en su dinámica. En este contexto, cobra especial relevancia la gestión integral de riesgos bancarios que llevan a cabo las organizaciones bancarias. Organismos internacionales, como el Banco Internacional de Pago (BIS) en Basilea, se han preocupado en aglutinar a los países con voluntad de regular los mercados financieros y sanear al sistema. Dentro de estas regulaciones, las recomendaciones del Comité de Supervisión Bancaria han mostrado que el rol de la gestión integral de riesgos impacta en la dinámica de todo el sistema bancario, sobre todo luego del año 2007 (Comité de Supervisión Bancaria de Basilea, 2010).

A la luz de los aspectos generales de la gestión de riesgos, la discusión está centrada en la implementación de los modelos que utilizan variables macroeconómicas y describen la dinámica del sistema bancario en escenarios estresados o con eventos extremos. En la actualidad, claramente, existe un gran volumen de información y los modelos de gestión basados únicamente en cálculos matemáticos no dan cuenta de la complejidad de ciertos fenómenos financieros.

La gestión de los riesgos de mercado está estrechamente vinculada a los modelos matemáticos que miden la exposición o máxima pérdida contingente: alta intensidad, muy baja frecuencia. Por este motivo, es muy importante realizar una correcta gestión

en las organizaciones bancarias, que involucre a todos los agentes de manera responsable.

El presente trabajo tiene como objetivo específico estudiar críticamente las metodologías estocásticas para la evaluación de riesgo de mercado. Para ello es necesario analizar las principales medidas de riesgo tradicionales (Basak y Shapiro, 2001; Bessis, 2011; Duffie y Pan, 1997; Jorion, 1997) comenzando con el Valor a Riesgo (VaR). Luego desde una visión más moderna, se estudiará en detalle la Teoría de Valores Extremos (*Extreme Value Theory*, EVT) que hará foco en el riesgo de la cola de la distribución de ganancias y pérdidas. Finalmente, y de acuerdo a la regulación actual, se analizarán los modelos alternativos de Déficit Esperado (*Expected Shortfall*, ES).

Tras este planteo inicial, se expondrá la incidencia de los modelos en la gestión eficiente de riesgos de mercado en bancos. Para ello, este trabajo pondrá énfasis en dos hipótesis: el VaR tradicional en activos muy expuestos a riesgos de mercado es ineficiente en la predicción de pérdidas cuando existen colas pesadas; ES y EVT poseen coherencia como medidas de riesgo y son más eficiente que VaR en la predicción de riesgos extremos. Sin embargo, la regulación actual no da cuenta de su necesidad en la práctica diaria; se advierte al respecto que aun cuando los textos de Basilea III incorporan la necesidad de utilizar EVT al 97,5% de confianza, esta práctica todavía no es usual en las regulaciones locales de los países miembros del Comité.


Referencias bibliográficas

- Basak, S. y Shapiro, A. (2001). Value-at-risk-based risk management: optimal policies and asset prices. *Review of Financial studies*, vol. 14, no. 2, pp. 371-405.
- Bessis, J. (2011). *Risk management in banking*. John Wiley & Sons.
- Comité de Supervisión Bancaria de Basilea (2010). *Convergencia internacional de medidas y normas de capital. Marco revisado*. Visión integral. Suiza: Banco de Pagos Internacionales.
- Duffie, D. y Pan, J. (1997). An overview of value at risk. *The Journal of derivatives*, vol. 4, no. 3, pp. 7-49.
- Jorion, P. (1997). *Value at risk: the new benchmark for controlling market risk*. Irwin Professional Pub.

Palabras Clave: Riesgo de mercado – Organizaciones bancarias – VaR – Teoría de Valores Extremos – Déficit Esperado

Área de aplicación: Otras Ciencias Económicas, Administración y Negocios.

Teoría sobre distribuciones de probabilidad



APLICACIÓN DE UNA CÓPULA MIXTA PARA LA ESTIMACIÓN DE LA PÉRDIDA TOTAL DE UNA CARTERA DE SEGUROS VEHICULARES

María José Bianco, Yennyfer Feo, Lucas Barreda

Centro de Investigación en Métodos Cuantitativos Aplicados a la Economía y la
Gestión (CMA)

Facultad de Ciencias Económicas, Universidad de Buenos Aires

Av. Córdoba 2122, C1120AAQ CABA, Argentina

mariajosebianco@economicas.uba.ar, feceyen@gmail.com,

barredalucas@economicas.uba.ar

Resumen

En el área actuarial, el correcto modelado y estimación de la distribución de pérdida total de una cartera de seguros constituye un eje central para el proceso de administración del riesgo. Desde un punto de vista individual, por un lado, permite determinar la prima y las cargas correspondientes a cada asegurado. Desde un nivel agregado, por el otro, cuantificar el riesgo de la cartera significa establecer la capacidad total y de retención de la misma, lo que llevará a su correspondiente trato con contratos de reaseguro (Peng Shi, 2015).

Típicamente se ha utilizado modelos basados en el trabajo desarrollado por Lundberg (desde 1903), por medio del cual se configura, bajo el supuesto de independencia, el monto promedio de pagos y el número de siniestros, y luego se define la pérdida total como el producto de estas cuantías. Una de las principales contribuciones a este trabajo ha sido el uso de modelos lineales generalizados introducidos por Haberman y Renshaw (1996), donde se incluyen covariables en el análisis.

Sin embargo, el supuesto de independencia puede ser restrictivo y llevar a un desvío significativo en la estimación (Kramer, 2012). Por tal motivo, Gschlöbl & Czado (2007) relajan dicho supuesto y modelan el monto promedio incluyendo como covariable al número de siniestros, la cual resulta significativa. En adelante, diversos trabajos como en Czado (2013) desarrollaron modelos de dependencia mediante el uso de cópulas.

De este modo, el objetivo del trabajo es aplicar y comparar dos metodologías para el cálculo de pérdida esperada a una cartera real de seguros de automóviles en

Brasil. Para el primer método, denominado tradicional, se consideró un modelo estadístico independiente para la variable pérdida total, y para el segundo método se consideró un modelo de cópula mixta que incorpora la dependencia entre las variables número de siniestros y monto del pago.

La metodología del trabajo se estructura, en este sentido, en dos secciones principales. Como primera instancia, con el objetivo de definir completamente el modelo de pérdida total para el caso tradicional, se estimaron los parámetros para las distribuciones Lognormal, Weibull, Pareto y Gamma mediante métodos estadísticos como el de máxima verosimilitud, de momentos y el de percentiles. Posteriormente, se evaluó la razonabilidad de las distribuciones mediante una serie de pruebas de bondad de ajuste que, permiten comparar los datos de la muestra con los datos del modelo ajustado y además medir el desempeño de las distribuciones mencionadas para escoger el más apropiado, para lo cual se emplearon gráficos de probabilidad (Probability-plot) y pruebas de Kolmogorov-Smirnov. Una vez seleccionada la distribución de pérdida se estimó la esperanza.

En la segunda fase se utilizó un modelo de cópula mixta para el modelado conjunto del monto de los pagos y el número de siniestros. Para la distribución marginal del pago por siniestros se propone el testeo para distribución Gamma, Weibull y Lognormal. Para la distribución marginal para la frecuencia de siniestros se testeó un modelo Exponencial y Exponencial-Poisson. Luego se probaron familias de cópulas conocidas: Gaussiana, Frank, Clayton rotada. Una vez modelada la cópula mixta, se estimó la pérdida esperada.

Todo el proceso se trabajó con el lenguaje y entorno estadístico de libre acceso R.

De las fases mencionadas se encontró que la distribución con mejor ajuste para la distribución de la pérdida fue la Log-normal tanto en las pruebas gráficas, como en las numéricas, estimando la pérdida esperada con una diferencia de 4% de la cifra empírica de la base. En el caso de la segunda parte, las variables de frecuencia y severidad tienen comportamientos muy extremos, una correlación de Kendall negativa y baja (-0,24), pero que rechazan la hipótesis de independencia al 5% de confianza. La mejor combinación de marginales estimadas fue Exp-Poisson para la función de densidad de la frecuencia y Lognormal para la función de densidad de la severidad, con la familia bivariada Clayton rotada 270 grados la cual es asimétrica y da mayor enfoque

a la relación en las colas de las marginales, dicha familia presentó las menores estimaciones de pérdida llegando a una diferencia de 12% con la pérdida promedio empírica de la base, la familia que mayor sobrestimaciones arrojó fue la copula Frank. Para finalizar se detectó que asumir independencia entre la severidad y la frecuencia llevaría a sobrestimaciones de la pérdida esperada de hasta 194%.

Referencias bibliográficas

Boland, J. Philip. (2006). Statistical and probabilistic methods in Actuarial science.

Chapman & Hall/CRC. New York.

Czado, C., Kastenmeier, R., Brechmann, C., Min, A. (2012). A mixed copula model for insurance claims and claim sizes. Center for Mathematical Sciences, TU Munchen.

Gschlößl, S. & Czado, C. (2007). ‘Spatial modelling of claim frequency and claim size in non-life insurance’, Scandinavian Actuarial Journal.

Haberman, S., Renshaw, A.E., (1996). Generalized linear models and actuarial science.

The Statistician 45, 407–436.

Kramer, N., Brechmann, E., Silvestrini, D., Czado, C. (2012). Total loss estimation using copula-based regression models. TU Munchen, Department of mathematical Statistics.

Kellison, S, London. R. L. (2011). Risk Models and their Estimation. Actex Academic Series. Winsted CT.

Klugman, S. A., H. H. Panjer, and G. E. Willmot (2004). Loss Models: From Data to Decisions. 2nd ed. New York: John Wiley & Sons.

Lundberg, F. (1903). Approximerad framställning afsannolikhetsfunktioner. II. aterförsäkring af kollektivrisker. Almqvist & Wiksells Boktr, Uppsala.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Shi, P., Feng, X., Ivantsova, A (2015). Dependent frequency-severity modeling of insurance claims. Insurance: Mathematics and Economics.

Palabras Clave: Pago por siniestros, número de siniestros, pérdida esperada, cópula mixta.

Área de Aplicación: Otras Ciencias Ciencias Actuariales.

**ANÁLISIS DE CORRELACIONES CONDICIONALES DINÁMICAS
APLICADO AL ESTUDIO DE LA INTERRELACIÓN ENTRE MERCADOS
BURSÁTILES**

Sergio Martín Buzzi

Departamento de Estadística y Matemática, FCE -UNC

sergio.buzzi@eco.unc.edu.ar

Resumen

En este trabajo se utilizan un modelo DCC-EGARCH y un modelo ADCC-EGARCH para estimar las correlaciones condicionales de un grupo de índices bursátiles. Luego, a partir del análisis de la evolución de dichas correlaciones condicionales se obtiene una intuición sobre la dinámica de la interrelación entre los índices. Aquellos mercados que se encuentren más integrados, por ejemplo los correspondientes a socios comerciales, deberían por lo general mostrar correlaciones más elevadas entre sí que con otras plazas. Por otra parte, en caso de existir efecto contagio, las correlaciones estimadas entre los índices deberían ser más bajas en épocas de crisis que en épocas de expansión. Se consideran los 9 mercados más importantes del mundo y los mercados de valores de Brasil y Argentina.

El modelo DCC-EGARCH consiste en estimar modelos GARCH exponenciales (EGARCH) sobre cada serie de retornos por separado y luego usar los residuos estandarizados obtenidos para estimar un modelo de correlaciones condicionales dinámicas (DCC); mientras que el modelo ADCC-EGARCH es una generalización del anterior que permite incorporar comportamientos asimétricos, en el sentido de que las correlaciones entre los índices serían mayores en épocas de crisis (efecto contagio).

Luego de estimar las correlaciones dinámicas por medio de los modelos DCC-EGARCH y ADCC-EGARCH, se grafican algunas series de correlaciones condicionales dinámicas seleccionadas. En primer lugar se grafican las correlaciones condicionales dinámicas obtenidas usando ambos métodos para retorno del Merval contra los restantes mercados. Luego se hace lo mismo con el rendimiento del índice Standard and Poor's. Ambos modelos arrojan estimaciones similares en el caso de ambos mercados.

En base a las estimaciones obtenidas para el índice Merval se concluye que: el índice con el que existe mayor correlación es en general el Bovespa; las correlaciones más bajas se dan con los mercados asiáticos; las correlaciones se reducen marcadamente

en la “Crisis del 2001”, por tratarse de una crisis doméstica y en la crisis global iniciada en 2007 las correlaciones alcanzan su nivel más elevado, por ser una crisis global.

Las correlaciones dinámicas del índice Standard and Poor’s con los restantes mercados, en general se mantienen relativamente estables en el tiempo. Se ve claramente una correlación elevada con el índice Nasdaq 100; luego una correlación moderada con los mercados de Europa y parte de América Latina y correlación reducida con los mercados asiáticos. El caso del índice Merval es bastante particular, porque la correlación respecto al mismo es más fluctuante que las restantes correlaciones del índice Standard and Poor’s, lo que refuerza la existencia de factores domésticos propios de la economía Argentina que recurrentemente desacoplan su mercado bursátil de los otros.

Referencias bibliográficas

Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, 4 (4), 537–572.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20 (3), 339–350.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Palabras Clave: Series de Tiempo – DCC EGARCH – ADCC EGARCH – Mercados.

Área de Aplicación: Economía

VARIACIÓN REGULAR EN VARIAS VARIABLES: APLICACIÓN A TEOREMAS DE CONVERGENCIA

Lila Ricci¹, Gabriela Boggio²

¹ Centro Marplatense de Investigaciones Matemáticas, Universidad Nacional de Mar del Plata

² Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística.

Universidad Nacional de Rosario

lricci@mdp.edu.ar

Resumen

Aparte de la clásica convergencia tipo TCL, en la que la varianza tiende a cero mientras la media permanece constante, se han definido otros tipos de convergencia que

resultan útiles en diferentes escenarios. En particular trataremos la convergencia tipo Tauber que es usada por Jørgensen y Martínez [2] en teoremas de convergencia de Modelos Exponenciales con Dispersión univariados que tienen como límite distribuciones gamma. En este caso, el parámetro de dispersión queda fijo mientras que la media tiende a 0 o a 1 y es llamada también convergencia tipo variación regular dado que en las demostraciones de dichos teoremas se hace necesario aplicar propiedades de las funciones de variación regular. La teoría correspondiente fue desarrollada por Jovan Karamata a partir de 1930

[1] y su aplicación original fue sobre los teoremas Tauberianos. Éstos teoremas vinculan propiedades de funciones con propiedades de sus transformadas de Laplace y resultan de gran utilidad para aplicar a variables aleatorias que tienen expresiones complejas pero cuya transformada de Laplace toma una forma más simple. En Teoría de Probabilidades, las funciones de variación regular han sido aplicadas al estudio de las distribuciones extremas. Estas aparecen frecuentemente en aplicaciones a finanzas, redes sociales, series de tiempo, análisis de riesgos, seguros y últimamente a telecomunicaciones. Nuestro interés es generalizar los teoremas de convergencia de Jørgensen y Martínez a modelos multivariados [3] y para esto se hace necesario contar con generalizaciones de los teoremas tauberianos y de la teoría de variación regular; en particular del llamado Teorema de Representación de Karamata. Resnick por su parte extendió la noción de variación regular a funciones de varias variables, para su aplicación a la teoría multivariada de distribuciones de valores extremos. En cuanto al vínculo con los teoremas tauberianos, ahora en varias variables, se cuenta con extensiones de los teoremas tauberianos dadas por Omey [4], que resultaron imprescindibles para nuestro estudio sobre convergencia de modelos Exponenciales con Dispersión multivariados. Sin embargo, los resultados no siempre estuvieron formulados en los términos o para las situaciones para las que los teníamos que aplicar, y se hizo necesario adaptarlos. Presentaremos en particular la adaptación de teoremas de Karamata, que han enunciados y probados por Resnick, para su aplicación a funciones de variación regular con índice $\rho < 0$ en \mathbf{R}^n , a funciones de variación lenta, es decir con índice ($\rho = 0$), en \mathbf{R}^n .

Referencias bibliográficas


- [1] L de Haan, E Omey, and S Resnick. Domains of attraction and regular variation in \mathbf{R}^n . *Journal of Multivariate Analysis*, 14(1):17 – 33, 1984.

- [2] B. Jorgensen. *The theory of dispersion models*. Chapman and Hall, 1997.
- [3] B. Jorgensen. Construction of multivariate dispersion models. *Brazilian Journal of Probability and Statistics*, 27(3):285–309, 2013.
- [4] E. Omey and E. Willekens. Abelian and Tauberian theorems for the Laplace transform of functions in several variables. *Journal of Multivariate Analysis*, 30(2):292 – 306, 1989.

Palabras Clave: Variación regular; modelos exponenciales con dispersión; convergencia tipo Tauber.

Área de Aplicación: Ciencias Exactas y Naturales.

Otras Categorías Metodológicas

A green arrow graphic pointing to the right, with a white outline of a second arrow inside it, creating a double-arrow effect.

LA EVALUACIÓN EN ESTADÍSTICA EN CARRERAS DE OTRAS DISCIPLINAS

Ernesto A ROSA

Universidad Nacional de Tres de Febrero (UNTREF)
Departamento de Metodología, Estadística y Matemática
ear@fibertel.com.ar

Resumen

Desde el punto de vista de sus contenidos, la Estadística no es una disciplina convencional, ya que en ella se entremezclan aspectos matemáticos, nociones filosóficas, y fuertes contenidos conceptuales, cuya combinación no se encuentra en otras ciencias. Tampoco lo es para su enseñanza o transmisión, ya que dependiendo de la formación básica de los docentes y de su capacidad didáctica, puede llegar a ser una materia incomprensible para la mayoría de los alumnos, o una asignatura entretenida que se convierte en herramienta de suma utilidad para su aplicación a innumerables disciplinas con las que se entrecruza (del “*Programa de Enseñanza de la Estadística y Ciencias Afines - PEECA*”, que está vigente en la UNTREF desde el año 2003).

Presentación

Partiendo de ese concepto, las formas de evaluar los conocimientos que se transmiten en Estadística, tienen aún un mayor nivel de complejidad, ya que para hacerlo se deben considerar:

- La disciplina en que se enmarca la Carrera.
- Los objetivos de la materia dentro del Plan de Estudios.
- Los contenidos desarrollados en el Programa y conocimientos efectivamente adquiridos.
- El enfoque adoptado para desarrollar la asignatura.
- El nivel de profundidad con que los alumnos deberían ser capacitados en Estadística.
- Los conocimientos de Estadística necesarios para el correcto cursado de asignaturas posteriores correlativas en el Plan.
- El tipo de problemas de Estadística que deberán enfrentar y resolver los alumnos al finalizar sus estudios.

En la presente Ponencia (que corresponde a una derivación del Trabajo Final presentado en el “*Seminario de Evaluación de los Aprendizajes Universitarios*” realizado en el marco del “*Posgrado en Docencia Universitaria*” de la UNTREF), se analizan en detalle los contenidos de los cursos de Estadística de acuerdo a los Planes de Estudio en que se enmarcan, los enfoques con que deben desarrollarse esos cursos, y particularmente las formas en que deben evaluarse los conocimientos transmitidos y adquiridos en ellos. También constituye un detalle de la Comunicación presentada en la Conferencia Satélite del IASE del Congreso del ISI 2015, denominada “*Métodos de Evaluación en un Curso de Estadística*”.

Desarrollo del Tema

El trabajo original fue presentado partiendo de las formas de Evaluación, padecidas inicialmente por el autor como alumno en los diferentes niveles educativos, y experimentadas posteriormente, desde los inicios en las actividades docentes y su evolución hasta la actualidad. Para llegar a las “*formas de evaluar*”, se hacen referencias a las “*formas de enseñar*”, experimentadas inicialmente como alumno, pasando posteriormente a las experiencias iniciales como docente en múltiples carreras de otras disciplinas, las que permitieron en definitiva, con la experiencia allí adquirida, elaborar esta propuesta.

Posteriormente se analizan las formas de evaluación de los conocimientos adquiridos en los diversos niveles educativos, que usualmente implicaban tener que memorizar teoremas y demostraciones, para posteriormente recitarlos o escribirlos. ¿Qué sentido tenía forzar a la realización de esos esfuerzos mentales? Peor aún era tener que memorizar las fórmulas que usualmente remataban esos desarrollos teóricos, sabiendo que en pocos días pasarían a conformar una masa de recuerdos fusiformes, integrada por remembranzas de los estudios primarios o secundarios con los más recientes y complejos desarrollos del nivel universitario.

En el trabajo se pasa luego a analizar las causas del porqué resultaba tan molesto que periódicamente evaluaran los conocimientos adquiridos en las materias cursadas. La sensación de incomodidad o desagrado, no provenía de ser evaluados, sino de la forma o esquema utilizado para la evaluación, y que obligaba a la situación de tener que escoger entre *estudiar para aprender o estudiar para aprobar* (paradoja central en la problemática de Evaluación para los especialistas en Educación). También se analizan

los factores negativos detectados en la Enseñanza de la Estadística (primero como estudiante y después como colaborador docente).

Finalmente, se expone la evolución realizada en los procedimientos aplicados para enseñar primero y evaluar después los cursos de Estadística desarrollados en diversas carreras de otras disciplinas (ingeniería, ciencias económicas, sociales, biomédicas, exactas, etc.), intentando no replicar las malas experiencias recibidas y observadas.

Conclusiones

En lo referente específicamente a la **Evaluación**, el énfasis fue puesto en el enfoque que debería aplicarse, replanteando todo el sistema de medición de la captación de los conocimientos adquiridos por parte de los alumnos, redefiniendo tanto la forma de hacerlo, como los temas a incluir, las condiciones previas al momento de realizarla, la forma de devolver a los alumnos los resultados que habían logrado, etc., impulsando, dentro de lo posible, que los alumnos *estudiaran para aprender* y no solo *para aprobar*.

En resumen, estas modificaciones consistían en que los alumnos pudiesen durante la evaluación:

Consultar cualquier bibliografía que se tenga al alcance.

Recurrir a ejemplificaciones o consulta de experiencias o casos similares al enfrentado.

Aclarar con quien evalúa, cualquier duda del planteo o problema a resolver.

Debatir el problema con otros “profesionales” en un grupo interdisciplinario (lo más difícil).

Además, había que resolver el problema en un tiempo limitadísimo.

Naturalmente, para la resolución de esos problemas, no se requiere de ninguna manera demostrar que se manejan artilugios matemáticos que fundamenten lo que se utiliza.

Por último, en el trabajo se analizan detalladamente los efectos de esos cambios en la enseñanza y evaluación, con los beneficios conexos en el cumplimiento de los objetivos de los cursos.

Referencias bibliográficas

Aburto Luyo, H., (2000). Evaluación de los Aprendizajes. Min. Educación, Perú.

Alvarez Mendez, J. M., (2001). Evaluar para Conocer, Examinar para Excluir. Morata, Madrid, España.

- Carballo, R., (1990). Evolución del Concepto de Evaluación. Bordón, 42, (Pág. 423-431, España.
- DINIECE, (2003). La Evaluación de la Calidad Educativa en Argentina: Experiencias Provinciales. Min. Educación, R. Argentina.
- García Martínez, F. A., (2005). La Evaluación en el Complejo Mundo de la Educación. Etic@net, Año II, N° 5, Granada, España.
- García Ramos, J.M., (1989). Bases Pedagógicas de la Evaluación: Síntesis. DialNet, Madrid, España.
- Lafi, P., (2003). La Paradoja de la Evaluación Educativa: Evaluar para Aprobar vs. Evaluar para Aprender. Mendoza, R. Argentina.
- Litwin, E., (1998). La evaluación de los Aprendizajes en el Debate Didáctico Contemporáneo. Ed. Paidós Educador, Buenos Aires, R. Argentina.
- OCDE, (1995). Measuring What Students Learn, París.
- Popham, W. J., (2008). La Evaluación Referida a Criterios. Cultura Evaluativa, México.
- Santos Guerra, M. A., (1999). 20 Paradojas de la Evaluación del Alumnado en la Universidad Española. Revista Electrónica Interuniversitaria. Univ. de Granada, España. www.evaluacion.edusanluis.com.ar. La Evaluación Educativa como Derecho Humano. San Luis, R. Argentina.

Palabras Clave: Estadística, Enseñanza, Métodos de Evaluación.

Área de Aplicación: Enseñanza de la Estadística.

**ASOCIACION ESTADISTICA: ERRORES EN CONTENIDOS
CONCEPTUALES Y PROCEDIMENTALES EN ESTUDIANTES
UNIVERSITARIOS**

Andrea Fabiana Righetti, Silvia Joeques

Facultad de Ciencias Económicas .UNC.

analizamos@yahoo.com.ar

Resumen

Uno de los contenidos más importantes de cualquier curso introductorio de Estadística es el de asociación estadística, entendiendo como tal a la extensión del concepto de correlación, a variables numéricas y no numéricas. Dicho contenido se desarrolla en base a tres ejes temáticos: análisis de regresión, análisis de correlación y

tablas de contingencia. En el proceso de adquisición y consolidación de estos conocimientos es posible cometer errores que pueden llegar a formar parte del conocimiento científico (Rico, 1995). El análisis de los errores que los alumnos manifiestan en su proceso de aprendizaje, constituyen una excelente herramienta para revelar el estado del conocimiento, imprescindible a la hora de realimentar los procesos de enseñanza y de aprendizaje con el fin de mejorar los resultados. Actualmente es posible encontrar una amplia bibliografía relacionada con el estudio de los errores en la temática de la estadística descriptiva en general. Pero las investigaciones sobre lo que ocurre en relación al tema de la asociación estadística, cuando se aplica inferencia estadística, ya sea mediante procedimientos de intervalo de confianza o pruebas de hipótesis, no se han desarrollado con igual magnitud. Entre las investigaciones pioneras se encuentra Vallecillos y Batanero (1997a, 1997b), Estepa (2004). Batanero, Estepa y Godino (1997,1998), y en nuestro país las de: Álvarez, Gonzales García y Terán (2009); Rodríguez, Agnelli y Huerta (2010); Lavalle, Micheli y Rubio (2006) y Korin, (2010), entre otros. Por este motivo se llevó a cabo un estudio de los errores más frecuentes en asociación estadística, desde una concepción cognoscitivista–constructivista del aprendizaje y la enseñanza, basada principalmente en la teoría del aprendizaje significativo de Ausubel (1976) y en la enseñanza para la comprensión de Perkins (1995). Los objetivos específicos de la investigación consistieron en: a) tipificar los errores conceptuales y procedimentales, b) agrupar a los estudiantes en función del tipo de errores detectados y de un conjunto de variables socio-demográficas de desempeño académico y de dedicación al estudio y c) indagar las opiniones de los alumnos acerca de los factores que pueden influir en los errores conceptuales y procedimentales.

Metodología: Se llevó a cabo un estudio exploratorio descriptivo a estudiantes de segundo año de la carrera de contador (FCE-UNC). En una primer etapa se aplicó un cuestionario con el que se evaluaron las respuestas de los estudiantes en el segundo parcial de la asignatura y en una segunda etapa se efectuaron entrevistas, tomando una muestra de estudiantes seleccionados por muestreo aleatorio sistemático, para una evaluación cualitativa. El cuestionario, elaborado para esta investigación sobre contenidos conceptuales y procedimentales referidos a los bloques temáticos de asociación estadística, se tomó como base las investigaciones referidas a las pruebas de hipótesis en general de Vallecillos y Batanero (1997a) y las referidas a la asociación estadística en particular de Estepa (2004). Se analizaron estadísticamente las respuestas

de los estudiantes en los cuestionarios, tanto en sus aspectos socio demográficos como de desempeño académico. A partir de un análisis multivariado se determinaron perfiles de estudiantes y mediante el análisis de contenido de entrevistas, se indagó sobre la opinión de los estudiantes acerca de los factores que pueden influir en los errores.

Resultados y Conclusiones: El estudio de los errores en contenidos conceptuales y procedimentales en asociación estadística, permitió detectar la confusión entre conceptos claves tales como: parámetro y estimador o regresión y correlación en las pruebas de hipótesis; dificultad para interpretar tanto las hipótesis como las conclusiones en el contexto de los problemas planteados y en el reconocimiento de los supuestos implícitos en cada metodología estadística. Además alertó sobre la complejidad que significa para los estudiantes el tema de asociación estadística no sólo debido a la cantidad de conceptos y fórmulas que involucra, sino a la relación entre los conceptos en cada una de las partes que conforma una prueba de hipótesis. A partir de la descripción de estos errores se indagó sobre sus posibles causas. Los estudiantes a menudo adquieren un conocimiento frágil (Perkins, 1995), que puede manifestarse en problemas de retención, comprensión o transferencia. El perfil de los estudiantes mostró que el grado de desempeño académico y de asistencia (previo al aprendizaje de la asociación estadística), está asociado con algunos tipos de errores.

Referencias bibliográficas

- AUSUBEL, D. (1976). *Psicología educativa: un punto de vista cognitivo*. México: Trillas.
- ALVAREZ, M.E.; M.E. GONZALES GARCIA y T. TERÁN. (2009). “Análisis de los errores en los conocimientos previos al proceso de enseñanza y aprendizaje del análisis de regresión bivariante lineal de los alumnos de la carrera de contador público”. *Decimocuartas Jornadas, Investigaciones en la Facultad de Ciencias Económicas y Estadística*. Rosario. Noviembre 2009. Universidad Nacional de Rosario. Argentina.
- BATANERO, C.; A. ESTEPA y J. D. GODINO. (1997). “Evolution of students’ understanding of statistical association in a computer based teaching environment”. En Garfield, J.B. y Burrill, G. (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics*. IASE Round Table Conference Papers, (pp. 191-205). Voorburg: International Statistical Institute.

- BATANERO, C.; A. ESTEPA. y J. D. GODINO, (1998). “La construcción del significado de la asociación mediante actividades de análisis de datos: reflexiones sobre el papel del ordenador en la enseñanza de la estadística”. *II Seminario de la Sociedad Española en Educación Matemática*. Pamplona.
- ESTEPA, A. (2004). “Investigación en Educación Estadística. La asociación estadística”. En R. Luengo (Ed.). *Líneas de investigación en Educación Matemática*, (pp. 227-255). Badajoz . Tecnigraf, S. A.
- LAVALLE, A.; E. MICHELI y N. RUBIO. (2006). “Análisis didáctico de la regresión y correlación para la enseñanza media”. *Relime*, 9 (3), 383-406.
- KORIN, C. (2010). “La comprensión de los test de hipótesis estadísticos. Un estudio con alumnos universitarios”. *FaMAF - Revista de Educación Matemática*. 25. Disponible en www2.famaf.unc.edu.ar/rev_edu/documents/vol_25/prop_08.pdf
- PERKINS, D. (1995). *La escuela inteligente*. Barcelona: Gedisa.
- RICO, L (1997). “Reivindicación del error en el aprendizaje de las matemáticas” *Revista Epsilon* 38, 185- 198.
- RODRÍGUEZ, M.I.; H. AGNELLI y A. HUERTA. (2010). “Pruebas de Hipótesis Estadísticas: Algunas consideraciones para la práctica docente”. *Revista de educación matemática*. Disponible en: http://www2.famaf.unc.edu.ar/rev_edu/documents/vol_25/prop_04.pdf.
- VALLECILLOS, A. (1994). *Estudio teórico - experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral Departamento de Didáctica de la Matemática. Universidad de Granada.-
- VALLECILLOS, A. y C. BATANERO. (1997a). “Aprendizaje y Enseñanza del Contraste de Hipótesis: Concepciones y errores”. *Enseñanza de la Ciencia* 15(2), 189-197.
- VALLECILLOS, A. y C. BATANERO. (1997b). “Conceptos activados en el Contraste de Hipótesis Estadísticas y su comprensión por estudiantes universitarios”. *Recherches en didactique des mathématiques* 17(1), 29-48.

Palabras Clave: Asociación Estadística – Contenidos conceptuales y procedimentales
– Regresión y correlación.

Área de Aplicación: Enseñanza de la Estadística.

ALGUNAS ACTIVIDADES PARA ENSEÑAR CONCEPTOS ESTADÍSTICOS

Hugo Darío Echevarría

Universidad Nacional de Río Cuarto, Universidad Nacional de Villa María

hechevarria2007@hotmail.com

Resumen

Se ha señalado insistentemente que el aprendizaje de la Estadística genera dificultades en los alumnos en todos los niveles educativos, por lo que se han realizado diversas propuestas tendientes a solucionarlas. Por ejemplo, Batanero (sin datos) sostiene: “como en cualquier otra rama de las matemáticas, el material manipulativo debe desempeñar un papel básico en los primeros niveles de enseñanza, por la necesidad que tienen los niños de contar con referentes concretos de los conceptos abstractos que tratamos de enseñarles”. Aunque autora se refiere a la probabilidad y a los primeros niveles de enseñanza, ofrecer los contenidos en forma concreta también puede ser importante en el nivel universitario, sobre todo si los alumnos tienen escasa motivación en relación a ellos.

El objetivo de esta ponencia es presentar algunos ejercicios que pueden implementarse en el aula sin contar con materiales sofisticados, desarrollados con la finalidad de exponer conceptos de estadística, para estudiantes que tienen escasa o nula formación matemática y que deben asistir a las clases desde las 18 horas luego de haber trabajado a doble turno en instituciones educativas. Ante esta circunstancia, se decidió utilizar situaciones cotidianas y lo más concretas posibles, a su vez, tratando de que no permanezcan todo el tiempo sentados sino que realicen algunas acciones que los impulsen a mantener la atención en los contenidos dados.

No se trata de una propuesta completa (como por ejemplo, la enseñanza de la estadística con programas informáticos o mediante proyectos), sino de algunos ejercicios que pueden integrarse en cualquiera de los planteos ya realizados, sobre todo, en el momento en que el docente explica los conceptos teóricos involucrados. Además, requiere de un grupo que sea lo suficientemente numeroso como para realizar análisis estadísticos, aunque si éste fuera demasiado grande podría seleccionarse una muestra para implementar las actividades.

Sin seguir fielmente sus ideas, estos ejercicios se inspiraron en algunos conceptos de Gardner. La forma de aplicación en el aula de su teoría ha sido por demás debatida, aquí tomo lo que sugieren Pérez Sánchez y Beltrán Llera (2006). Según ellos,

con esta propuesta, el docente cambia constantemente la forma de presentación y vías de acceso a los contenidos (a través de la inteligencia lingüística, corporal-cinestésica, lógico-matemática, etc.), favorece el aprendizaje cooperativo, siempre tiene en cuenta tanto el ritmo como las condiciones de cada alumno y ofrece experiencias directas (los alumnos podrían levantarse y desplazarse dentro del aula, el docente les presenta objetos para manipularlos o les solicita que construyan algo). Aunque puede parecer contradictorio, también tomo ideas de Amos et al. (1969), quienes escribieron su libro en un momento de auge del conductismo y basándose en esta corriente psicológica, pero con una forma de exposición que bien puede aceptarse desde un paradigma constructivista.

A modo de ejemplo, seguidamente muestro algunos ejercicios que pueden llevarse a cabo en el aula, algunos responden a nociones de estadística descriptiva y otros a estadística inferencial:

Estadística descriptiva

1. Considerar la estatura como variable para describir el grupo.
2. Formar en orden por estatura.
3. Obtener la moda, la mediana, la media, el primero y el tercer cuartil, algún decil o percentil.
4. Construir un gráfico concreto: escribir en el pizarrón los valores ordenados ubicándose cada alumno frente a su valor, de modo que se formen filas con el largo del número de casos.
5. Hacer un gráfico de barras, observar si el gráfico de barras es similar en cuanto a forma a la figura que se dio cuando los alumnos formaron frente a su valor de estatura.
6. Analizar la forma de la distribución: ¿es simétrica?, ¿es alta y delgada o aplanada (curtosis)?, ¿es unimodal, bimodal, trimodal, multimodal?
7. Contestar ¿dónde se halla la media?, ¿coincide o se parece a algún caso?, ¿a quién se parece más?, ¿este caso, representa bien al grupo?, ¿coinciden o se parecen la media, la mediana y la moda?
8. Calcular los desvíos respecto de la media de cada integrante, aplicarles el módulo y sumarlos.
9. Calcular el promedio de los desvíos.
10. Recategorizar la variable con 5 categorías iguales y repetir el gráfico de barras,

luego con algunas más. Analizar las diferencias y semejanzas con un histograma construido por un programa estadístico (por ejemplo, el PSPP).

11. Seleccionar otra variable al menos ordinal, preferentemente con mucha dispersión. Cada alumno debe escribir en una hoja su valor en esa variable.
12. Repetir los puntos anteriores (en todos los casos luego de obtener un estadístico se reflexiona sobre el significado del mismo para describir la población o la muestra y se compara la información que ofrece cada estadístico).

Estadística inferencial

13. Elegir tres variables, todos los alumnos escriben en un papel de 10 x 10 cm sus valores.
14. Hacer bollos con ellos e introducirlos en una caja. Responder: si se forman dos grupos al azar, ¿se parecerán?, ¿por qué?
15. Formar al azar dos grupos (por ejemplo, un alumno va extrayendo bollos, asignándose el primero a un grupo, el segundo al otro y así sucesivamente).
16. Si un grupo se toma como muestra de la totalidad del curso, ¿serán similares las distribuciones de la muestra a las del curso?, si se tratara de hacer un estudio experimental riguroso (en el sentido usado en Echevarría, 2016), ¿los grupos se parecerán entre sí?
17. Tomar algunas variables y comparar los grupos sobre la base de ellas.
18. Contestar: ¿los grupos se parecen?, ¿por qué?

Estos ejercicios sólo se han implementado parcialmente hasta el momento de escribir esta ponencia, por lo que no dispongo datos en relación al efecto que pueda tener en el aprendizaje o en la atención de los alumnos y, si bien da la impresión de que en el grupo implementado produjo una mayor atención y participación en clase, será conveniente realizar algunos estudios sistemáticos tendientes a evaluarla.

Referencias bibliográficas

Amos, J.; Lloyd Brown, F. y Mink, O. (1969) *Introducción a la estadística: un texto programado*. Madrid: Ediciones del Castillo.

Batanero, C. (sindatos). ¿Hacia dónde va la educación estadística? <http://www.ugr.es/~batanero/pages/ARTICULOS/BLAIX.pdf> (consultada: 12-05-2018).

Echevarría, H. (2016) *Diseños de investigación cuantitativa en psicología y educación*.
Río Cuarto: UniRío editora. [https://www.unrc.edu.ar/unrc/comunicación/
/editorial/repositorio/978-987-688-166-1.pdf](https://www.unrc.edu.ar/unrc/comunicación/editorial/repositorio/978-987-688-166-1.pdf) (consultada: 05-04-2016).

Pérez Sánchez, L. y J. Beltrán Llera. (2006) Dos décadas de “inteligencias múltiples”:
implicaciones para la psicología de la educación. *Papeles del Psicólogo*. Vol
(27)3: 147-164. [http://redalyc.uaemex.mx/src
/inicio
/ArtPdfRed.jsp?iCve=77827304&iCveNum=0](http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=77827304&iCveNum=0). Consultada 30-05-2007.

Palabras Clave: enseñanza de la estadística - materiales concretos - alumnos
universitarios.

Área de Aplicación: Enseñanza de la Estadística

LA RELEVANCIA VALORATIVA DE LOS INTANGIBLES EN EL MERCADO DE CAPITALES ARGENTINO. ESTUDIO EMPIRICO APLICANDO UN MODELO PARA DATOS DE PANEL DE EFECTOS FIJOS

Cecilia R. Ficco

Facultad de Ciencias Económicas – Universidad Nacional de Río Cuarto
ceciliaficco@yahoo.com.ar

Resumen

La relevancia de la información contable para la valoración de las empresas en los mercados de capitales ha sido motivo de debate, fundamentalmente, desde mediados de los noventa del siglo pasado, cuando comenzaron a realizarse diferentes estudios que han mostrado una disminución de la relevancia valorativa de la información contable a lo largo del tiempo. La falta de reflejo, en los estados financieros, de gran parte de las inversiones en intangibles, ha sido identificada como una de las principales causas de la debilidad en la relación entre variables de mercado e información contable (Lev & Zarowin, 1999; Lev, 2001; Beattie & Thomson, 2005).

Así, los sistemas contables en vigor presentan una importante limitación para brindar información relevante para valoración externa de las empresas. Esta pérdida de relevancia valorativa de la información contable revela, además, la existencia de otras fuentes de información, representativas de la creación de valor de las empresas, que son tomadas en cuenta en la toma de decisiones de inversión, estando esas fuentes de

información directamente vinculadas a los intangibles y, en particular, al capital intelectual.

Los intangibles se convierten, de este modo, en un objeto de estudio que adquiere una importancia central en la investigación contable orientada al mercado de capitales. Surge, así, la necesidad de evaluar su relevancia valorativa, es decir, de estudiar si los inversores los tienen en cuenta en la fijación de los precios de mercado de los títulos, como lo han hecho los estudios previos de Ritter & Wells (2006), Swartz, Swartz & Firer (2006), Wang (2008), Damash, Durand & Watson (2009), García, Kimura, de Barros & Cruz (2010), Oliveira, Rodrigues y Craig (2010), Ferraro & Veltri (2011), Veltri & Silvestri (2011), Berzkalne & Zelgalve (2014), Kimouche & Rouabhi (2016a) y Kimouche & Rouabhi (2016b), entre otros.

Y este es, precisamente, el problema que aborda la presente investigación, a partir del cual se formula la siguiente pregunta general que recoge su esencia: ¿es la información sobre intangibles, reconocidos y no reconocidos por la contabilidad, relevante para la valoración de las empresas en el mercado de capitales?

En este marco, el presente trabajo tiene como objetivo fundamental estudiar la relevancia valorativa de los intangibles en el mercado de capitales argentino, utilizando un modelo lineal de valoración construido sobre la base del modelo de Ohlson (1995) bajo una especificación de efectos fijos.

En lo que respecta al método, la investigación realizada responde a un enfoque cuantitativo, en cuyo marco se realiza un estudio de tipo correlacional, en tanto se pretende examinar la asociación entre variables, específicamente, variables referidas a intangibles y al valor de mercado de las empresas cotizantes. El diseño de investigación es no experimental y, en relación a la dimensión temporal, se aplica un diseño longitudinal, ya que los datos se recolectan en distintos momentos del tiempo pero durante un período determinado: 2009-2015. Más específicamente, se aplica un diseño longitudinal de panel, en tanto las variables son medidas para el mismo grupo de sujetos (empresas cotizantes), en los diferentes años que integran el período mencionado.

Para la medición de las variables se utilizan datos contables de las empresas que se obtienen de los estados financieros anuales que presentan en la Bolsa de Comercio de Buenos Aires. También se recolecta información procedente del mercado, particularmente, la referida al precio de las acciones, la que se obtiene de los registros referidos a la negociación de las mismas en el Mercado de Valores de Buenos Aires.

A partir de la estimación del modelo construido, empleando datos correspondientes a una muestra de 46 empresas cotizantes en el mercado de capitales argentino en el período 2009-2015, se obtienen resultados que revelan una asociación significativa entre la variable referida a “otros activos intangibles” y dos variables proxy correspondientes al capital humano y estructural con el precio de mercado de las acciones, lo que brinda evidencias en favor de la relevancia valorativa de estas dos dimensiones del capital intelectual y de los referidos activos intangibles. Esto indica que son tomados en cuenta por los inversores a la hora de valorar a las empresas cotizantes argentinas y tomar sus decisiones de inversión en relación a las mismas.

Referencias bibliográficas

- Beattie, V. & Thomson, S. J. (2005). Intangibles and the OFR. *Financial Management*, 29-30.
- Berzkalne, I. & Zelgalve, E. (2014). Intellectual capital and company value. *Procedia-Social and Behavioral Sciences*, 110, 887-896.
- Dahmash, F. N., Durand, R. B. & Watson, J. (2009). The value relevance and reliability of reported goodwill and identifiable intangible assets. *The British Accounting Review*, 41(2), 120-137.
- Ferraro, O. & Veltri, S. (2011). The value relevance of intellectual capital on the firm's market value: an empirical survey on the Italian listed firms. *International Journal of Knowledge-Based Development*, 2(1), 66-84.
- Garcia, C., Kimura, H., de Barros, L. & Cruz, L. (2010) The Impact of Intellectual Capital on Value Added for Brazilian Companies Traded at the BMF-BOVESPA. Disponible en: SSRN: <https://ssrn.com/abstract=1571576>.
- Kimouche, B. & Rouabhi, A. (2016a). The impact of intangibles on the value relevance of accounting information: Evidence from French companies. *Intangible Capital*, 12(2), 506-529.
- Kimouche, B. & Rouabhi, A. (2016b). Intangibles and Value Relevance of Accounting Information: Evidence from UK Companies. *Jordan Journal of Business Administration*, 12(2).
- Lev, B. (2001). *Intangibles: Management, Measurement and Reporting*. The Brookings Institution Press.
- Lev, B. & Zarowin, P. (1999). The Boundaries of Financial Reporting and How to Extend Them. *Journal of Accounting Research*, Autumn, 353-385.

- Ohlson, J. (1995). Earnings, Book Values and Dividends in Equity Valuation. *Contemporary Accounting Research*, 11(2), 661-687.
- Oliveira, L., Rodrigues, L. & Craig, R. (2010). Intangible assets and value relevance: Evidence from the Portuguese stock exchange. *The British Accounting Review*, 42(4), 241-252.
- Ritter, A. & Wells, P. (2006). Identifiable intangible asset disclosures, stock prices and future earnings. *Accounting & Finance*, 46(5), 843-863.
- Swartz, G. E., Swartz, N. P. & Firer, S. (2006). An empirical examination of the value relevance of intellectual capital using the Ohlson (1995) valuation model. *Meditari Accountancy Research*, 14(2), 67-81.
- Veltri, S. & Silvestri, A. (2011). Direct and indirect effects of human capital on firm value: evidence from Italian companies. *Journal of Human Resource Costing & Accounting*, 15(3), 232-254.
- Wang, J. C. (2008). Investigating market value and intellectual capital for S&P 500. *Journal of intellectual capital*, 9(4), 546-563.

Palabras Clave: Modelos lineales – Datos de panel – Efectos fijos – Valoración de intangibles – Mercado de capitales - Argentina

Área de Aplicación: Ciencias Económicas, Administración y Negocios

COMPARACIÓN DE LOS ESTILOS DE APRENDIZAJE DE LOS ALUMNOS DE LAS CARRERAS DE BIOLOGÍA Y DE MATEMÁTICA

Vanessa Edith Figueroa, Adriana Ayelén Quiroga Hahn, Norma Leonor

Rodríguez, María Cecilia del Valle Marchetti

Centro de Investigación de Estadística Aplicada.

Facultad de Ciencias Exactas y Naturales. Universidad Nacional de Catamarca.

vfiguefi10.vf@gmail.com

Resumen

Los estilos de aprendizaje se constituyen en un amplio campo de estudio y han llevado a considerar las diferencias personales de los alumnos en el proceso de aprendizaje, las que se relacionan con sus resultados académicos. En términos generales, los estilos de aprendizaje se definen como los rasgos cognitivos, afectivos y fisiológicos que sirven como indicadores relativamente estables, de cómo los

estudiantes perciben, interaccionan y responden a sus ambientes de aprendizaje (Juárez Lugo et al, 2012). Por ello, el conocimiento del estilo predominante en los alumnos se considera una herramienta docente relevante, permitiendo adaptar las estrategias de enseñanza a los diferentes estilos y lograr así elevar el rendimiento académico de ellos.

Por ello, el presente trabajo tiene como objetivo principal comparar los estilos de aprendizaje de los estudiantes de las carreras de Licenciatura en Ciencias Biológicas y Profesorado y Licenciatura en Matemática de la FaCEN – UNCa.

Por lo cual se plantearon los siguientes objetivos específicos:

- Identificar los estilos de aprendizaje de los alumnos de, Bioestadística de la Licenciatura en Ciencias Biológicas y de Estadística del Profesorado y Licenciatura en Matemática.

- Establecer estrategias de enseñanza apropiadas a los estilos de aprendizaje de los alumnos en estudio.

Se parte de la hipótesis que los estilos predominantes de los alumnos de las carreras de Biología son diferentes de los alumnos de las carreras de Matemática.

La investigación es cuantitativa, descriptiva e inferencial de enfoque transversal, pues se analizaron las puntuaciones obtenidas al aplicar un cuestionario propuesto por Pulido (2011) a los alumnos de segundo y tercer año de las carreras mencionadas en el año 2016, tales puntuaciones permitieron determinar los estilos de aprendizaje de los mismos. De las 4 etapas que conforman el cuestionario original se considera en la presente investigación solo los aspectos externos del alumno que corresponden a la primera y cuarta etapa del proceso de enseñanza y aprendizaje. Del análisis estadístico se verifica el cumplimiento de la hipótesis planteada. Se observó que en los alumnos de Bioestadística predominan el *estilo visual e introvertido* por lo tanto los estudiantes tienen memoria visual y más facilidad para incorporar y retener grandes cantidades de información con rapidez, además tienen más desarrollada la inteligencia intrapersonal, estudian solos, necesitan tiempo para pensar y reflexionar, les agrada escuchar más que hablar, optan por exámenes escritos, son reflexivos, se sienten cómodos con métodos de enseñanza donde no deben interactuar y prefieren que el docente sea el protagonista. Estos resultados se corresponden con los obtenidos por Acevedo Pierart et al (2009) pero se contraponen con los obtenidos por Moreno Colín y Sánchez Fabila (2013).

En el caso de los alumnos de Matemática se observó que la mayoría de los alumnos predomina el *estilo visual* y utilizan el *estilo auditivo combinado con otro*

estilo, además son *extrovertidos* por lo que predomina el pensamiento verbal o simbólico, de manera secuencial y ordenada, los mismos necesitan escuchar lo estudiado paso por paso además de tener la característica de que aprenden mejor cuando pueden compartir con sus compañeros, se relacionan para procesar la información, realizan trabajos grupales, discusiones, debates y trabajos en equipos, además interactúan con el docente y responden muy bien cuando el docente utiliza métodos de enseñanza interactivos. Tales resultados no coinciden con los obtenidos por Amado Moreno (2007).

De acuerdo a los resultados obtenidos se sugiere a los docentes de las carreras de Biología que utilicen como método de enseñanza el expositivo con material didáctico visual. Mientras que en las carreras de Matemática el método de enseñanza que los favorece es el expositivo interactivo con material didáctico visual.

Referencias bibliográficas

- Acevedo Pierart, C., et al (2009). Estrategias de aprendizaje en alumnos universitarios y de enseñanza media. *Revista Estilos de Aprendizaje*, 4 (2).
- Amado Moreno, et al (2007). Estilos de aprendizaje de estudiantes de Matemáticas en educación superior. *Memorias de la XVII Semana Regional de Investigación y Docencia en Matemáticas*, Departamento de Matemáticas, Universidad de Sonora México. *Mosaicos Matemáticos*, 20: 13-21.
- Juárez Lugo, C., et al (2012). El cuestionario de estilos de aprendizaje CHAEA y la escala de estrategias de aprendizaje ACRA como herramienta potencial para la tutoría académica. *Revista Estilos de Aprendizaje*, 10 (10).
- Moreno Colín, R. y Sánchez Fabila, G. (2013). Estilos de aprendizaje de la anatomía e historia de la Biología en estudiantes de la carrera de Biología en la FES IZTACALA, UNAM. *Revista Estilos de Aprendizaje*, 11 (11)
- Pulido, M. (2011). Estilos de aprendizaje de los estudiantes, un aspecto importante a tener en cuenta. *Tesis Doctoral en Docencia Universitaria de Disciplinas Tecnológicas*. Facultad de Ciencias Agrarias. UNCa. Argentina.

Palabras Clave: Estilos de aprendizaje – Bioestadística – Estadística – Estrategias de enseñanza y aprendizaje

Área de Aplicación: Enseñanza de la Estadística

MUESTREO SIMPLE Y DOBLE PARA EL MONITOREO DE UNIDADES NO CONFORMES EN PROCESOS DE ALTA CALIDAD

Silvia Joeques¹, Andrea Righetti¹, Marcelo Smrekar²

1- Instituto de Estadística y Demografía - Facultad de Ciencias Económicas - U.N.C.

2- Laboratorio de Ingeniería y Mantenimiento Industrial, Facultad de Ciencias Exactas, Físicas y Naturales - U.N.C.,
joeques@gmail.com

Resumen

En este trabajo se propone un nuevo procedimiento, el gráfico de control con muestreo doble (MD), cuyo objetivo es mejorar la capacidad para detectar cualquier condición de fuera de control mediante la observación de una segunda muestra. La eficacia de este procedimiento se basa principalmente en el hecho de que permite observar cambios pequeños o moderados en la proporción de unidades no conformes del proceso, sin aumentar el muestreo. El muestreo doble es especialmente útil cuando se lo utiliza en procesos de alta calidad, en los cuales la proporción de unidades no conformes y el tamaño de la muestra son reducidos. En esta situación, una alternativa posible consiste en emplear el gráfico np mejorado en la primera etapa del muestreo, dado que este procedimiento ha mostrado una gran ventaja sobre el gráfico np habitual para atributos. La utilización del muestreo doble en una segunda etapa del muestreo, tiene, además, dos posibles ventajas. En primer lugar, permite reducir la cantidad total de inspección y en segundo lugar, le da al proceso una segunda oportunidad antes de tomar una decisión. Finalmente se muestran los beneficios del gráfico np mejorado con muestreo doble evaluados en términos de eficacia estadística (longitud media de corrida, ARL) y se lo compara con el mismo gráfico con muestreo simple. Además, se presenta el procedimiento para la elección adecuada del plan de muestreo doble en función de los cinco parámetros requeridos. El trabajo incluye una aplicación con datos reales que permite observar los beneficios del muestreo doble en el monitoreo y seguimiento de procesos de alta calidad.

Objetivo general: Mostrar las ventajas de la utilización del muestreo doble MD para procesos de alta calidad, basada en la utilización del gráfico np mejorado en lugar del gráfico tradicional de Shewhart.

Objetivos específicos:

- a) Determinar el comportamiento del gráfico np mejorado, empleado en la primera etapa del MD, que valida la aproximación binomial-normal para muestras chicas.
- b) Establecer el desempeño comparativo del gráfico np mejorado con muestreo doble respecto al mismo gráfico empleando el muestreo simple.
- c) Efectuar la comparación entre los dos procedimientos mediante la evaluación del ARL_0 que indica la cantidad de puntos a graficar antes de dar una señal de alarma, cuando el proceso está bajo control.

Metodología: La metodología consiste en determinar en primer lugar el gráfico de control con muestreo simple para luego, de ser necesario, pasar al muestreo doble. El procedimiento del gráfico de control con muestreo simple se determina a partir del cálculo de p_0 y p_1 , siendo p_0 la proporción de ítems no conformes y p_1 la de ítems conformes. Posteriormente se establecen los valores de α y β , a efectos de definir el plan de control que consiste en obtener el Límite Superior de Control (LSC) con parámetro p_0 . Finalmente se toma una muestra de n unidades del proceso en intervalos de tiempo fijo y se calcula la proporción de unidades no conformes, p . Este valor se compara con el LSC y en función del resultado se toma la decisión de aceptar o rechazar el lote. En una etapa siguiente, se define el gráfico np con MD. Este gráfico utiliza cinco parámetros: el tamaño de la primera muestra (n_1), el número de aceptación para la primera muestra (límite de advertencia, LA), el número de rechazo para la primera muestra LSC_1 , el tamaño de la segunda muestra (n_2) y el número de aceptación (no rechazo) para la segunda etapa de muestreo ($LSCL_2$). A partir de la determinación de estos parámetros se define el plan de control. Por último, se muestra una aplicación con datos reales que permite observar los beneficios del muestreo doble en el monitoreo y seguimiento de procesos de alta calidad.

Resultados y conclusiones: En este trabajo se propone un nuevo procedimiento, el gráfico de control con muestreo doble (MD) cuyo objetivo es mejorar la capacidad de detectar cualquier condición de fuera de control mediante la observación de una segunda muestra. La comparación con el MS reveló que el procedimiento de MD puede ser considerado más apropiado para la detección de los incrementos en la proporción de ítems no conformes en el proceso, debido a que ofrece mejor eficacia estadística (en términos de longitud promedio de corrida), sin aumentar el muestreo. Finalmente, quien

desea resolver el problema de optimización del gráfico np con muestreo doble puede descargar fácilmente el programa en R desde la web el cual se encuentra disponible en el sitio: <https://sites.google.com/site/marcelosmrekar/home>

Referencias bibliográficas

Chakraborti S., Human S.W, Graham M.A. (2008). Phase I Statistical Process Control Charts: An overview and some results. *Quality. Engineering.* 21 (1). 52–62.

Cornish–Fisher expansion. (2013). Section 3.14 in Holton, A. Glyn. Value-at-Risk: Theory and Practice, second ed. e-book at <http://value-at-risk.net>.

Costa A.F.B., Claro F.A.E. (2007). Double sampling X control chart for a first-order autoregressive moving average process model. *Int. J. Adv. Manuf. Technol.* on-line edition.

He D., Grigoryan A., Sigh M. (2002). Design of double- and triple-sampling X-bar control chart using genetic algorithms. *Int. J.Prod. Res.* 40 (6) 1387–1404.

Joekes S., Pimentel Barbosa E. (2013). An improved attribute control chart for monitoring non-conforming proportion in high quality processes. *Control Engineering Practice.* vol. 21, Issue 4. Elsevier, pp. 407–412. <http://dx.doi.org/10.1016/j.conengprac.2012.12.005>.

Montgomery D.C. (2008). Introduction to Statistical Quality Control. Sixth ed. Wiley.

R Development Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna Austria, ISBN: 3-900051-07-0, 2012. URL <http://www.R-project.org>.

Rodrigues A.A., Epprecht E.K., Magalhães M.S. (2011). Double-sampling control charts for attributes. *J. Appl. Stat.* 38 (1) 87–112.

Wu Z., Wang Q. (2007). An np control chart using double inspections. *J. Appl. Stat.* 34 (7) 843–855.

Xie M., Goh T.N., Kuralmani V. (2002). *Statistical Models and Control Charts for High Quality Processes.* Kluwer Academic Publ.,

Palabras Clave: Gráficos de control – Gráfico np mejorado – Muestreo doble – Procesos de alta calidad – Proporción de unidades no conformes

Área de Aplicación: Industria y mejoramiento de la calidad

A horizontal orange arrow pointing to the right, with a white outline of the arrowhead. The word "POSTERS" is centered within the arrow.

POSTERS

Biología

A horizontal orange arrow pointing to the right, with a white outline of a second arrow inside it, creating a double-arrow effect.

UN MODELO GEOESTADÍSTICO PARA ANALIZAR LA DISTRIBUCIÓN ESPACIAL DE LA FRACCIÓN COMERCIAL DE CENTOLLA (*Lithodes santolla*) EN LA PATAGONIA CENTRAL, ARGENTINA.

Mauro A. E. Chaparro^{1,3}, Carla Firpo², Cecilia Mauna²

¹Centro Marplatense de Investigaciones Matemáticas- Universidad Nacional de Mar del Plata (CEMIM-UNMDP).

² Instituto Nacional de Investigación y Desarrollo Pesquero – Sub Programa Pesquerías de Crustáceos Bentónicos (INIDEP).

³Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
mchaparro@unmdp.edu.ar

Resumen

La estructura espacial de una población sometida a explotación y el esfuerzo pesquero aplicado sobre ella, es muy importante al momento de tomar decisiones con respecto a su manejo. Los denominados *king crabs* se caracterizan por ser stocks pesqueros muy importantes desde el punto de vista económico cuya población adulta presentan fuerte y persistente estructura espacial, permaneciendo en áreas restringidas. La pesquería de centolla (*Lithodes santolla*) en el Sector Patagónico Central se inició en 2004 y, a partir de 2009 se fueron incorporando gradualmente nuevos buques. Actualmente son 5 los buques congeladores con permisos para pescar centolla, entre las latitudes 44 y 48° S y desde la longitud 62°W hasta la costa, en la denominada Área II. Esto constituye un importante esfuerzo pesquero, para un recurso que aún no ha sido cuantificado y no cuenta con campañas de investigación específicas.

Este trabajo tiene por objetivo, analizar la distribución espacial de la fracción comercial de centolla utilizando la captura por unidad de esfuerzo de cada lance estandarizado (CPUEST), mediante técnicas geoestadísticas.

La campaña de investigación para la recolección de datos se realizó a mediados de octubre de 2016, con una duración aproximada de 20 días de trabajo efectivo y un diseño cuasi-regular de 209 lances con los que se cubrió la zona mencionada. La CPUE por lance se estimó a partir del número de machos de talla comercial (machos ≥ 110 mm de largo de caparazón), considerando como esfuerzo pesquero el número de trampas de la línea estandarizados a 3 días de pesca.

Para la detección de puntos atípicos se utilizó el *índice de Moran Local (IML)* en vecindades con distancias máximas de 0,3° (≈ 18 millas náuticas). El resultado de este

análisis mostró 10 puntos como potencialmente atípicos. Se realizó un análisis de los variogramas direccionales concluyéndose que no existe una dirección destacada.

En base a la representación del variograma empírico se utilizó una distancia de 2° (≈ 120 mn). Se analizaron 2 modelos, uno exponencial sin tendencia y otro esférico considerando que existe una tendencia en el valor medio de grado 1. De acuerdo con los resultados obtenidos, no se observaron diferencias destacadas entre ambos modelos sin embargo, dada la complejidad que se incorpora al considerar una tendencia en la media (modelo esférico) y en base al principio de parsimonia, se seleccionó el modelo exponencial.

El análisis de la validación cruzada mostró que los errores estandarizados presentan una distribución normal (test de Shapiro, $p < 0.01$), y en la representación espacial no se observó un patrón sistemático en los errores. Con el modelo de variograma definido se realizaron las predicciones en los puntos no muestrales a través de un kriging ordinario. La grilla de predicción se dividió cada $0,01^\circ$ (0,6 mn), resultando de esta manera una grilla de 155.526 puntos de predicción.

A partir de los resultados de la simulación condicionada a este modelo, se determinó que el valor medio CPUEST es de 3,14 con un intervalo de confianza del 95% de 0,45-5,85 para las medias simuladas. Es de destacar que el promedio global (3,34) sin considerar la dependencia espacial es mayor que cuando la misma es considerada. Los mapas obtenidos mediante kriging ordinario evidenciaron la estructura espacial en parches de la fracción comercial de esta especie.

Por otra parte, se construyeron mapas de probabilidades de superar umbrales de CPUEST de 3 y 4 individuos comerciales por trampa los cuales, de acuerdo a la información histórica de la pesquería y la distribución del esfuerzo pesquero durante las últimas temporadas de pesca, pueden considerarse como límites para la rentabilidad comercial. De esta forma se delimitaron los sectores con mayor rendimiento comercial coincidentes con las áreas históricas de pesca, mostrando la estabilidad espacial del recurso en los últimos años. Finalmente los resultados obtenidos en este trabajo podrían ser utilizados para realizar estimaciones de abundancia y definir medidas de manejo pesquero.

Referencias bibliográficas

Firpo, C; Wyngaard, J.; Iorio, M.I. 2013. Estandarización de las capturas de centollón (*Paralomis granulosa*) por trampa con diferentes tiempos de pesca en el Canal

Beagle, Argentina. Informe Técnico INIDEP N°91. 20 pp.

Firpo, C.; Mauna, C.; Wyngaard, J.; Lertora, P. 2015. Evolución de los desembarques, esfuerzo pesquero y captura por unidad de esfuerzo (CPUE), en la pesquería patagónica de centolla (*Lithodes santolla*), en el período 2011-2015. Inf. Téc. Of. INIDEP N° 22/2015.

Orensanz, J.M.; Jamieson, G.S. 1998. The assessment and management of spatially structured stocks: an overview of the North Pacific Symposium on Invertebrate Stock Assessment and Management. En: Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management. Editado por: G.S. Jamieson and A. Campbell. Can. Spec. Publ. Fish Aquat.Sci. 125. pp. 441-459.

Sudipto Banerjee, Bradley P. Carlin, Alan E. Gelfand. 2015. Hierarchical Modeling and Analysis for Spatial Data. Second Edition. Taylor & Francis Group. ISBN: 13: 978-1-4398-1918-0 (eBook - PDF). pp 583.

Palabras Clave: Estadística espacial – Pesquería –Crustáceos.

Área de Aplicación: Biología

Ecología y Medio Ambiente



ESTUDIO DE LOS ÍNDICES DE VEGETACIÓN EN EL NORTE DE LA PROVINCIA DE SANTA CRUZ

Marisa Sandoval , Paula Paredes, Dora Maglione , Julio Soto

Universidad Nacional de la Patagonia Austral- Unidad Académica Río Gallegos.

msandoval@uarg.unpa.edu.ar

Resumen

En la Provincia de Santa Cruz encontramos una gran diversidad de ambientes determinados fundamentalmente por el clima y el suelo. A escala regional existe un fuerte gradiente climático, donde las precipitaciones disminuyen de Oeste a Este y de Sur a Norte. Estas características condicionan la actividad ganadera predominante, que se realiza en forma extensiva, adaptándose a las restricciones del ambiente. En el caso del suelo las diferencias se encuentran asociadas al origen geológico del sustrato que le dio el lugar. Las áreas ecológicas intentan agrupar territorios que pueden considerarse, a determinada escala, una combinación particular de suelos, clima y vegetación. Para la provincia de Santa Cruz fueron definidas por Cuadra y Oliva (1994).

Para analizar la vegetación en las diferentes áreas ecológicas es posible utilizar los Índices de Vegetación (IV's), como estimador de sus características o condición. En diversos estudios se ha encontrado una asociación significativa positiva entre los IV's con productividad primaria neta aérea, biomasa, índice de área foliar, cobertura vegetal o evapotranspiración (Goward *et al.*, 1985) o variables climáticas como precipitación y temperatura (Jobbágy *et al.*, 2002). Estas relaciones permiten estudiar temporal y espacialmente la fenología y dinámica de la cobertura vegetal (White *et al.*, 1997).

El objetivo de este trabajo fue comparar y observar el comportamiento de diferentes índices de vegetación para las áreas ecológicas del Norte de la provincia de Santa Cruz y el de establecer asociaciones con los distintos elementos climáticos y los datos de vegetación de los Monitores Ambientales de Regiones Áridas y Semiáridas (MARAS), instaladas en el área (Oliva *et al.* 2011).

El área de estudio se localiza en la franja comprendida desde 45°58'46.28" S y 71°57'46.96" O hasta 47°31'51.82" S y 65°45'56.68" O, conteniendo 4 áreas ecológicas: Pastizal Subandino, Sierras y Mesetas Occidentales, Meseta Central y Estepa Arbustiva del Golfo San Jorge. Se calcularon diferentes índices de vegetación de una colección de imágenes Landsat 5TM, con un porcentaje de cobertura de nubes inferior al 30%, para el período junio de 2010 a mayo de 2011, seleccionado debido al avance de la sequía

registrada entre 2012 y 2016. Para todos los procesos en relación con las imágenes y para la extracción de valores de píxeles se empleó el programa Google Earth Engine Code Editor. A partir de la información obtenida se modelaron temporalmente los distintos índices en el periodo considerado y se compararon sus comportamientos en las distintas áreas ecológicas. Además, se trabajó con los datos de temperatura del aire y precipitación, obtenido por interpolación de estaciones meteorológicas de la provincia (Maglione et al. 2016), para estudiar las asociaciones de estas variables climatológicas, y de las variables asociadas al comportamiento de la vegetación.

La evolución anual de los índices de vegetación presentó un comportamiento diferencial en cada región ecológica analizada, debido principalmente a características estructurales asociada a variables climáticas.

Referencias bibliográficas:

- Cuadra D., Oliva G. (1994). Ambientes naturales de la Provincia de Santa Cruz. Revista Espacios. Universidad Nacional de la Patagonia Austral. UARG.
- Goward S.N., Tucker C.I., Dye D.G. (1985). North American vegetation patterns observed with the NOAA-7 AVHRR. *Vegetation*, 64, 3-14.
- Jobbágy E.G., Sala O.E., Paruelo J.M. (2002). Patterns and controls of primary production in the patagonian steppe: a remote sensing approach. *Ecology* 83 (2), 307-319.
- Maglione D., Soto J., Sáenz J.L., Talay C., Sandoval M., Bonfilio O., Llancahuen M. (2016). Una metodología para la construcción de mapas predictivos para variables climatológicas. Bienales del IV Encuentro de Investigadores de la Patagonia Austral. ISBN 978-987-3714-37-5 4° Ed.
- Oliva G., Gaitán J. Bran D., Nakamatsu V., Salomone J., Buono G., Escobar J., Ferrante D., Humano G., Ciari G., Suarez D., Opazo W., Adema, E. Celdrán, D. (2011). *Manual para la Instalación y Lectura de Monitores MARAS*. PNUD. Buenos Aires, Argentina.
- White M., Thornton P., Running S. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, 11(2), 217-234.

Palabras Clave: Áreas ecológicas – Maras - Índices de vegetación – Tipos fisonómicos – Asociación

Área de Aplicación: Ecología y Medio Ambiente

Economía

A horizontal orange arrow pointing to the right, with the word "Economía" written in blue text inside it.

COTIZADOR INMOBILIARIO CON RANDOM FOREST EN R

Patricia GUARDIA ⁽¹⁾, María Pía CAUMONT ⁽²⁾

1 Ministerio de Hacienda. República Argentina

2 Dirección General de Estadística y Censos del Gobierno de la Ciudad de Buenos Aires

patricia_pa@hotmail.com , mariapiacaumont109@gmail.com

Resumen:

Se presenta la UTILIZACION de una técnica RANDOM FOREST en Lenguaje R aplicado al Mercado Inmobiliario de la Ciudad de Buenos Aires, para producir una cotización monetaria de departamentos.

Objetivo: Describir los recursos estadísticos utilizados para esta técnica.

Materiales y métodos: Se utilizó una base de datos CUANTITATIVA, que se encuentra a disposición en la página data.buenosaires.gob.ar (datos de 2014), para utilizar el algoritmo basado en los arboles de decisión, randomForest (packaging) para obtener un precio de venta optativo frente a los que se presentan en el mercado inmobiliario. El dataset cuenta con 12.724 casos y 17 variables.

En principio se realiza una limpieza de la base, se identifican las variables útiles para el proceso, y se eliminan o imputan los datos faltantes por casos, ya que el algoritmo no corre bajo la presencia de missing. Se divide el set de datos en training y test (70/30 dividido aleatoriamente) para generar el espacio necesario para probar la función y aplicarla.

RF es muy similar a un Bagging de Arboles de decisión, además de generar variabilidad sobre los registros, genera variabilidad sobre los predictores. RF logra descorrelacionar los arboles a través de un pequeño ajuste: muestrear un subconjunto (m) de los M predictores. Cada vez que el algoritmo evalúa un Split de un árbol, solamente considera una muestra aleatoria simple de m y M. Considera las m variables elegidas y selecciona la que mejor Split genera.

Resultados: Dado que RF aproxima casi cualquier relación entre X e Y, resulta útil para detectar relaciones no lineales sin necesidad de parametrizarlas previamente. De este modo la estimación ofrece un elemento de debate, frente a un precio base que hasta el momento era indiscutible, ya que no se encontraba un argumento sustentable para su contienda.

Conclusiones: Con el apoyo de técnicas Estadísticas, se puede hacer un aporte al campo de la Urbanización; relacionando dos instrumentos que están a disposición y generando un resultado de fácil aplicación que permite un argumento de discusión con una base teórica y un error tangible, puede aplicarse en muchos otros ámbitos económicos, sociales y de investigación, observando el cuidado pertinente de las bases de datos.

Referencias bibliográficas

Anish Singh Walia. (july 24, 2017). Random Forests in R. [ADVANCED MODELING]
Recuperado de: <https://datascienceplus.com/random-forests-in-r/>
Emanuel Santana. (november 26, 2014). Data mining con R. [EJEMPLO EN
RANDOM FOREST] Recuperado de: <http://apuntes-r.blogspot.com/2014/11/ejemplo-de-random-forest.html>

Palabras Clave: RANDOM FOREST - ARBOLES DE DECISION - COTIZADOR
INMOBILIARIO - DATA MINING - MINERIA DE DATOS

Área de Aplicación: Economía, Estudios de Mercado

CONSTRUCCIÓN DE UN ÍNDICE ESPACIAL DE BANCARIZACIÓN. UNA PROPUESTA PARA LA REGIÓN CENTRO

Fernando García, María Inés Stímolo

Facultad de Ciencias Económicas Universidad Nacional de Córdoba

fgarcia.unc@gmail.com

Resumen

La bancarización resulta importante en tanto constituye un motor para el desarrollo económico y social de un país o región al favorecer la disponibilidad de servicios financieros para la población y las empresas y del nivel de acceso y utilización de tales servicios por parte de los distintos agentes económicos.

En este sentido, una medición adecuada del proceso de bancarización a través de un Índice de Bancarización (IB) resulta clave en tanto facilita la comprensión de la bancarización y contribuye a que exista un reconocimiento respecto a su trascendencia como elemento que puede apoyar al crecimiento y desarrollo económico.

Este trabajo propone la construcción de un IB para la región Centro utilizando metodologías estadísticas que incorporan la dimensión espacial de los datos.

El índice considera aspectos referentes a las diferentes dimensiones de análisis: Magnitudes Agregadas, Disponibilidad y Cobertura Geográfica y Utilización y Acceso, utilizando los indicadores parciales propuestos por el GMM (2011), los que se construyen a partir de información proporcionada por el Banco Central de la República Argentina (BCRA) y el Instituto Nacional de Estadística y Censos (INDEC) para el año 2010.

Como los datos están georreferenciados, es decir asociados con una localización, el Análisis de Componentes Principales (PCA, son sus siglas en inglés) puede verse afectado por los patrones espaciales subyacentes en los datos. En tal sentido, resulta adecuado aplicar PCA Geográficamente Ponderadas (GWPCA, son sus siglas en inglés) que permite incorporar la heterogeneidad espacial de los datos, es decir considerar situaciones donde los datos espaciales no son bien descritos por un modelo global (Lloyd, C. 2010; Harris, P., Brunson, C. & Charlton, M., 2011). Los resultados muestran que las Componentes Principales (CPs) seleccionadas permitirían una mejor visualización de la heterogeneidad espacial y corroborar que GWPCA constituye un aporte metodológico significativo en relación a PCA. De esta manera, se recomienda avanzar en la construcción del IB a partir de las CPs obtenidas a través de esta metodología.

Un índice de este tipo es importante en tanto permite la medición de manera global y sintética al incorporar la dimensión espacial de los datos. Así, facilita una visión general de la bancarización en la región Centro en su conjunto y en cada una de las provincias que la componen. Se destaca la provincia de Santa Fe, exhibiendo un mayor nivel de bancarización. Sigue en importancia la provincia de Córdoba, pero con un comportamiento más heterogéneo que contrasta con el de la provincia de Entre Ríos, que aunque presenta un nivel de bancarización menor exhibe un comportamiento más homogéneo.

Pese a las limitaciones del índice, vinculadas fundamentalmente a la imposibilidad de contar con datos de indicadores parciales que midan en forma más adecuada las dimensiones Magnitudes Agregadas y Acceso y Utilización, este indicador debería ser considerado como una aproximación a la medición de la bancarización regional y, acaso, como un trabajo en proceso.

Referencias bibliográficas

Gollini, I, Lu, B, Charlton, M, Brunson, C. & Harris, P. (2015). GW model: an R Package for exploring Spatial Heterogeneity using Geographically Weighted Models. *Journal of Statistical Software*, 63 (17), 1–50.

Grupo de Monitoreo Macroeconómico (2011). Indicadores de bancarización, Buenos Aires.

Harris, P., Brunson, C. & Charlton, M. (2011). Geographically weighted principal components analysis». *International Journal of Geographical Information Science*, 25(10), 1717-1736.

Lloyd, C. D. (2010). Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems*, 34 (5), 389-399.

Palabras Clave: Componentes Principales Geográficamente Ponderadas – Bancarización - Índice.

Área de Aplicación: Economía

ANÁLISIS DE SERIES DE LAS FINANZAS PÚBLICAS DE MENDOZA

Graciela Gei, Franco Bignone

Facultad de Ciencias Económicas – Universidad Nacional de Cuyo

graciela.gei@fce.uncu.edu.ar

Resumen

En este trabajo se presenta la modelación de las principales variables de recursos y gastos de la Administración Pública de la provincia de Mendoza, mediante el análisis estadístico univariado de series de tiempo.

Es un tema recurrente en el análisis fiscal el desarrollo de metodologías que permitan proyectar series temporales de los gastos e ingresos del Estado. Es de particular relevancia la obtención de estimaciones que permitan cerrar el ejercicio en curso. Esto implica pronosticar con datos actualizados el comportamiento de las series bajo estudio en los próximos meses, ya que de esta manera se puede obtener el total de ingresos que la Administración Pública podría disponer en los meses restantes del año y, en consecuencia, determinar y especificar en conjunto con el total de gastos a ejecutar. Además de proyecciones para series agregadas (ingresos y gastos totales, resultado operativo) es necesaria la obtención de predicciones sobre la composición de

estos recursos y erogaciones, lo que permite profundizar el análisis que se puede realizar desde el punto de vista económico y fiscal.

El objetivo general del trabajo consistió en la modelación de las principales series de gastos e ingresos de la Administración Pública de Mendoza, siendo el objetivo particular la obtención de proyecciones puntuales y por intervalos que permitieron cerrar el ejercicio fiscal 2017. Asimismo, se destaca que la modelación realizada se puede utilizar como base para la proyección en años posteriores.

Se trabajó con 39 series de tiempo de frecuencia mensual, que se extienden entre enero de 2007 y junio de 2017. El primer paso en la metodología aplicada consistió en la obtención de series estacionarias, para luego ajustar a los datos transformados el mejor modelo $ARMA(p,q)$ posible, tal que el mismo tenga el menor Criterio de Información de Akaike Corregido (AICC) dentro de los modelos con coeficiente principal significativo y buen comportamiento en los residuos. También se analiza el Criterio de Información Bayesiana (BIC).

Se comparan las funciones muestrales de autocorrelación y autocorrelación parcial correspondientes a los datos estacionarios con las del modelo teórico ajustado, a fin de comprobar en forma gráfica la bondad del ajuste del modelo $ARMA(p,q)$ seleccionado.

A partir del modelo ajustado a los datos estacionarios y teniendo en cuenta la transformación aplicada y las diferenciaciones realizadas, se obtiene el modelo para los datos originales. Con este modelo se determinan las predicciones para los siguientes 12 meses, junto con sus intervalos de predicción. Se muestran además los valores estimados para el acumulado correspondiente a diciembre de 2017, así como los intervalos aproximados del 95% de confianza para estos valores.

A los efectos de comprobar que el modelo seleccionado ajusta bien a los datos originales se analiza el comportamiento de los residuos del modelo obtenido para los datos estacionarios. Si el ajuste realizado es bueno, estos residuos deben comportarse como un Ruido Blanco. Para realizar este análisis se utilizan diferentes test en los cuales se chequea si los residuos tienen un comportamiento que puede considerarse de Ruido Blanco y, además, se comprueba en forma gráfica ese comportamiento.

Entre los resultados de la investigación, se obtuvieron estimaciones puntuales y por intervalos para las 39 series seleccionadas y se comprueba que los modelos predictivos obtenidos ajustan adecuadamente a los datos.

Ya que se modelan series que se obtienen como suma y resta de otras, se obtienen distintas estimaciones alternativas para la serie de resultado operativo, que es la de mayor relevancia en lo que al análisis económico refiere.

Se propone como extensión a la investigación, la modelación mediante metodologías alternativas de series de tiempo que se aplican particularmente a casos en los cuales algunas series se obtienen como agregación de otras.

Referencias bibliográficas

Box, G.E.P., Jenkins, G.M. (1976). "Time Series Analysis, Forecasting and Control", San Francisco, Estados Unidos: Holden Day.

Brockwell, P., Davis, R.A. (1987). "Time Series: Theory and Methods", Nueva York, Estados Unidos: Springer-Verlag.

Leiva, R. (1995). "Introducción al análisis de series de tiempo", Mendoza, Argentina: F.C.E. de la U.N.C.

Peña, D. (2005) "Análisis de Series temporales" Madrid, España: Alianza Editorial.

Palabras Clave: Finanzas Públicas – Proyecciones – Intervalos de Predicción - Modelos ARMA

Área de Aplicación: Economía

MEDICIÓN DE LA PROPENSIÓN AL EMPRENDEDURISMO EN ESTUDIANTES DE CIENCIAS ECONÓMICAS

Mariana Guardiola¹ y Norma Patricia Caro²

Facultad de Ciencias Económicas – Universidad Nacional de Córdoba

¹marianaguardiola@eco.unc.edu.ar ; ²pacaro@eco.unc.edu.ar

Resumen

El presente trabajo tiene por objetivo medir la propensión de un grupo de estudiantes de Ciencias Económicas a emprender una nueva idea o proyecto a partir de dos escalas propuestas en la literatura consultada. Una de ellas referida a la autoeficacia emprendedora, entendida como la convicción de ser un emprendedor y poder llevar a cabo el desarrollo de dicha idea o proyecto, y la otra, a aspectos inherentes a la personalidad proactiva, abarcando cualidades propias de las personas que identifican oportunidades y actúan sobre ellas.

En el año 2016 se diseñó y administró una encuesta piloto a 360 estudiantes avanzados de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba (UNC) a fin de evaluar, entre otros aspectos, su propensión a emprender proyectos de calidad. El cuestionario fue sometido a revisión de otros docentes y de estudiantes de grado quienes a través de sus aportes y sugerencias permitieron hacer modificaciones para una mejor comprensión del instrumento. Se utilizaron escalas *Likert* como instrumento de recolección de datos, partiendo de un conjunto de afirmaciones cuyo objeto es considerar las valoraciones personales que van desde lo menos valorado (1) a lo más valorado (5), además de la consistencia de las respuestas actitudinales.

En un primer avance se validó una escala referida a la propensión al uso de financiamiento colectivo para nuevas ideas o proyectos, lo que se completa en este trabajo con la aptitud al emprendedurismo en estudiantes de grado.

A través de un modelo de ecuaciones estructurales, se estimaron las relaciones entre las variables latentes y los indicadores que se explican a través de estas variables.

Los modelos de ecuaciones estructurales combinan modelos de regresión con métodos multivariados a fin de establecer múltiples relaciones entre las variables observadas y no observadas.

Los resultados obtenidos de las estimaciones y las medidas de bondad de ajuste evidencian un buena *performance* del modelo propuesto identificando los principales indicadores que son explicados por la autoeficacia emprendedora y la personalidad proactiva (variables latentes).

A su vez, ambas variables latentes presentan una relación de covariación significativa lo que estaría indicando que los estudiantes con mayor convicción a emprender y desarrollar proyectos o ideas, son aquellos con mayor proactividad.

En este sentido, se espera poder continuar con esta línea de investigación para completar este estudio que aún no ha sido realizado en nuestro país al tiempo que poder extenderlo a un universo mayor de estudiantes con diferentes especialidades.

Referencias bibliográficas

- Acock, A (2013). *Discovering Structural Equation Modeling Using Stata*. Stata Press Publication Stata Corp LP. College Station. Texas. 2.
- Cupani, M. (2012). Análisis de ecuaciones estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación. *Revista Tesis* N° 1 pp 186 - 199.

- Kline, R. (2011). Principles and Practice of Structural Equation Modeling. 3^o ed. The Guilford Press- New York - London.
- Salvador Ferrer, C. M., & Morales Jiménez, J. (2009). Fundamentos psicológicos de la autoeficacia emprendedora en jóvenes mexicanos. *Alternativas en Psicología*, 14(20), 35-47.
- Salvador Ferrer, C. M., & Morales Jiménez, J. (2009). Autoeficacia Emprendedora en Jóvenes Mexicanos. ¿Cómo Afecta la Satisfacción Vital y la Inteligencia Emocional Percibida?. *Interamerican Journal of Psychology*, 43(2).
- Seibert, S. E., Crant, J. M., & Kraimer, M. L. (1999). Proactive personality and career success. *Journal of applied psychology*, 84(3), 416.

Palabras Clave: Ecuaciones Estructurales – Emprendedurismo – Variables Latentes.

Área de Aplicación: Área de Aplicación: Economía

VARIABLES EXPLICATIVAS DE LA POBREZA EN LOS HOGARES ARGENTINOS: APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA

Laura Agustina López, Martín Saino

Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

martinsaino@gmail.com

Resumen

Problemática y/o antecedentes del trabajo

Indudablemente uno de los fenómenos que caracterizan el escenario económico-político-social de la Argentina actual es la pobreza. Dicha problemática posee un carácter acuciante y urgente dada la importancia que tienen sus efectos sobre el bienestar y las condiciones de vida de los hogares argentinos. Para paliar esta dificultad estructural, resulta clave que cuestiones relacionadas a dicho fenómeno ocupen un lugar central en el debate económico y social, en todos los niveles de gobierno, en aras de orientar el diseño y la aplicación de políticas públicas y sociales.

En este sentido, el abordaje de la investigación tiene como eje central la identificación de las variables explicativas de la pobreza, en las distintas regiones geográficas argentinas, bajo una óptica multidimensional. En otros términos, los factores de pobreza se definen bajo un enfoque planteado como complemento de la perspectiva unidimensional (ampliamente tratada por la literatura) la cual se basa

principalmente en un estudio meramente monetario. Dicho enfoque multidimensional propone una cohesión de determinantes de diferente naturaleza y no sólo monetaria, buscando avanzar en un examen que incorpore factores intervinientes referidos a las personas e instituciones, los cuales no quedan reflejados cuando solo se trabaja con indicadores estrictamente monetarios. De este modo, lo que se plantea es un análisis más integral y complejo de la pobreza, contemplando y considerando otras dimensiones de la vida humana, tales como obstáculos socioeconómicos o circunstancias personales.

Objetivos

Se plantea como objetivo general de la investigación determinar las variables que mayor influencia tienen en la vulnerabilidad de los hogares argentinos de ser clasificados como pobres; siendo el móvil esencial identificar aquellas problemáticas o circunstancias de los hogares en situación de pobreza donde deben focalizarse los esfuerzos de políticas públicas y sociales.

Ahora bien, en lo que hace al planteo de objetivos específicos, se introducen como tales:

- Caracterizar la situación de los hogares argentinos por regiones geográficas, a través de la construcción de perfiles de pobreza, con el fin de presentar una primera aproximación descriptiva de las disparidades regionales, en cuanto a condiciones de privación, de los hogares pobres respecto a los no pobres.
- Realizar una selección a priori de variables que se consideran relevantes en la explicación de la pobreza, de acuerdo al consenso académico, y reflejar con ellas atributos socioeconómicos, tanto individuales del jefe de hogar como referidos al hogar en su conjunto.
- Aplicar la regresión logística para determinar qué variables deben incluirse en el análisis, de acuerdo al grado de significatividad estadística que reporten. De este modo, se procura dilucidar qué tipo de relación o influencia tienen sobre la pobreza y cuantificar dicho peso sobre la chance de ser clasificado como pobre.

Metodología

La fuente de datos utilizada es la Encuesta Permanente de Hogares (EPH), provista por el INDEC. Asimismo, el análisis recae en las distintas regiones geográficas, definidas por la EPH a través de la agregación de los aglomerados que las componen (Gran Buenos Aires, NOA, NEA, Cuyo, Pampeana y Patagonia), utilizando particularmente los datos correspondientes al segundo trimestre de 2016. Se tienen por

un lado, variables referidas exclusivamente al hogar (base Hogares de la EPH) y por otro, variables referidas a características personales de cada uno de sus miembros (base Personas de la EPH). A su vez, si bien la mayoría de las variables propuestas en el modelo de regresión se toman tal cual como fueron diseñadas en la encuesta; otras variables se construyeron a partir de fórmulas utilizadas en otros trabajos de investigación y a partir de un diseño propio que permita reflejar la característica en cuestión.

Particularmente, para la “construcción” de la variable dependiente se utilizó el procedimiento detallado por el INDEC: el método de la línea de pobreza. Por otro lado, para las variables independientes, se definieron a priori como las más relevantes para el análisis: tamaño del hogar, edad del jefe de hogar, estado civil del jefe, estatus laboral del jefe, tipo de acceso al agua, calidad de pisos interiores, calidad de cubierta de techo, calidad desagüe del baño, combustible usado para cocinar, régimen de tenencia, tasa de participación esperada, años de educación promedio de las personas del hogar en edad de trabajar, máximo nivel educativo alcanzado por el hogar, nivel educativo del jefe, región geográfica.

La elección de esta lista de variables obedece a una revisión de los diferentes trabajos y aplicaciones de carácter empírico, tanto para Argentina como para otros países, en los cuales se toman como referencia variables similares. Se tienen tanto investigaciones enfocadas particularmente a la construcción de indicadores multidimensionales, los cuales implican definir dimensiones y variables dentro de ellas; como aquellas que se detienen particularmente en la caracterización del fenómeno a través de modelos econométricos (probit o logit).

Por último, se procede a correr la regresión logística. Para dicha estimación, se recurrió a la utilización del software estadístico SPSS, a través del cual se aplicarán distintas técnicas, conocidas como enter (introducir) y forward stepwise (por pasos hacia adelante). A su vez, cabe aclarar, que dentro de este último método, SPSS da las opciones de elegir entre tres criterios alternativos para adoptar decisiones estadísticas: razón de verosimilitud (RV), condicional y Wald. De este modo, se obtienen diferentes modelos que serán comparados bajo distintos criterios, a saber: estadísticos de bondad de ajuste (-2loglikelihood, Cox & Snell, Nagelkake, Hosmer and Lemeshow), capacidad de predicción y el criterio de parsimonia.

Resultados o conclusiones

En base a los resultados arrojados por el modelo, se concluye que la variable referida a la condición de actividad del jefe de hogar es la que mayor influencia tiene en la chance de ser clasificado como un hogar pobre. Se demostró, además, a partir de un análisis de indicadores de empleabilidad tales como la tasa de empleo, de desempleo y de inactividad, que las disparidades entre pobres y no pobres en las distintas regiones son notables. Queda en evidencia que destinar esfuerzos en políticas dirigidas a proteger el estatus laboral de un jefe de hogar argentino llevará a reducir el riesgo de que su hogar sea pobre.

Asimismo, aspectos de infraestructura y características habitacionales de las viviendas también resultan ser factores cuyo peso es determinante en la chance de ser un hogar pobre; y en definitiva donde debe recaer la política a implementar. Particularmente, el modelo señala como más significativas a la “Calidad del desagüe del baño”, “Calidad de la cubierta del techo” y “Régimen de tenencia”: cuanto más deficiente sea la infraestructura y precario el régimen, más chances habrá de que repercuta a nivel pobreza en los hogares bajo estudio.

Con respecto al análisis de la tasa de participación esperada, planteada como la capacidad que tiene un hogar de generar ingresos, se arribó a la conclusión de que generalmente los hogares pobres tienen mayor cantidad de miembros dependientes que los hogares no pobres. Dicho resultado se encuentra íntimamente relacionado a otra de las variables propuestas como explicativas: el tamaño familiar. En resumidas cuentas, el mayor número de miembros dependientes, una menor proporción de generadores de ingresos y el tamaño relativamente grande de los grupos familiares pobres respecto a los hogares no pobres determinan su capacidad para generar propias condiciones de sustentabilidad y progreso.

Por último, en lo que hace al análisis educativo, se comprobó que los hogares pobres se caracterizan por tener menor cantidad de años en promedio destinados a la educación formal respecto a los no pobres. En este sentido, resulta clave tomar como referencia el resultado de la regresión: aumentar la cantidad de años en formación educativa lleva claramente a una reducción del factor de riesgo de pobreza. Es indudable la importancia que tiene la esfera educativa dentro del estudio de la pobreza ya que deja en evidencia la necesidad de políticas destinadas a mejorar no sólo la retención escolar sino también la calidad de enseñanza; para en conjunto paliar esta deficiencia y con ello, a su vez, disminuir el riesgo de caer en situación de pobreza.

En síntesis, la determinación de las variables que más influencia ejercen sobre el fenómeno de la pobreza resulta clave, no solo a modo explicativo de determinantes profundos; sino para intentar dar un paso más allá metodológica, teórica y políticamente: ampliar el estudio a una visión multidimensional, reconociendo que son muchas las dimensiones que deben analizarse y no solo una referida a los ingresos, permitirá entender mejor la problemática estructural de la pobreza para luego llevar a cabo una implementación práctica e integral de políticas públicas y sociales.

Referencias bibliográficas

- Altimir, O. (1979). Dimensión de la pobreza en América Latina.
- Angarita, M. J. U. LA TEORÍA DE LAS CAPACIDADES EN AMARTYA SEN.
- Bazán Ojeda, A., Quintero Soto, M. L., & Hernández Espitia, A. L. (2011). Evolución del concepto de pobreza y el enfoque multidimensional para su estudio. *Quivera*, 13(1).
- Bustelo, M., & Lucchetti, L. (2004). La pobreza en Argentina: perfil, evolución y determinantes profundos (1996, 1998 y 2001). Documentos de Trabajo del CEDLAS.
- Conconi, A. (2009). Pobreza multidimensional en Argentina: ampliando las medidas tradicionales de pobreza por ingresos y NBI (Doctoral dissertation, Facultad de Ciencias Económicas).
- de Río, G. (2007). Grupo de Expertos en Estadísticas de Pobreza. Compendio de mejores prácticas en la medición de la pobreza. Santiago de Chile, Recuperado de: http://www.eclac.cl/deype/publicaciones/sinsigla/xml/9/34409/rio_group_compendium_es.pdf.
- Feres, J. C., & Mancero, X. (2001). El método de las necesidades básicas insatisfechas (NBI) y sus aplicaciones en América Latina. Cepal.
- Feres, J. C., & Mancero, X. (2001). Enfoques para la medición de la pobreza: breve revisión de la literatura. CEPAL.
- Gasparini, L., Cicowiez, M., & Escudero, W. S. (2013). Pobreza y desigualdad en América Latina. Conceptos, herramientas y aplicaciones. Buenos Aires: Temas.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.

- Instituto Nacional de Estadística y Censos. (2016). Incidencia de la pobreza y de la indigencia en 31 aglomerados urbanos - Resultados segundo trimestre de 2016. Sitio web: <https://www.indec.gob.ar>
- Instituto Nacional de Estadística y Censos. (2016). Valorización mensual de la Canasta Básica Alimentaria y de la Canasta Básica Total - Gran Buenos Aires. Sitio web: <https://www.indec.gob.ar>
- Laderchi, C. R., Saith, R., & Stewart, F. (2003). Does it matter that we do not agree on the definition of poverty? A comparison of four approaches. *Oxford development studies*, 31(3), 243-274.
- Lopez, C., & Safojan, R. (2014). Un análisis multidimensional de la pobreza: evidencia reciente de las regiones de Argentina. *Revista de Economía Política de Buenos Aires*, (12), 36.
- Ojeda, S., Pereyra, L. E., & Gontero, S. (2005). La pobreza en los hogares del Gran Córdoba: aplicación del modelo de regresión logística. *Revista de Economía y Estadística*, 43(1), 99-121.
- Ordaz, J. L. (2009). México: impacto de la educación en la pobreza rural. Cepal.
- Pérez López, C. (2005). Métodos estadísticos avanzados con SPSS. Thompson. Madrid.
- Ravallion, M. (1998). Poverty lines in theory and practice (Vol. 133). World Bank Publications.
- Rowntree, B. S. (1941). Poverty and progress. A second social survey of York. *Poverty and progress. A second social survey of York*.
- Saino, M. "Una aplicación de métodos estadísticos: caracterización de los ingresos de los residentes en el aglomerado de Posadas, segundo trimestre de 2011". Universidad Nacional de Córdoba, Facultad de Ciencias Económicas, Córdoba. No publicado.
- Sen, A. (1983). Poor, relatively speaking. *Oxford economic papers*, 35(2), 153-169.
- Sen, A. (1985). A sociological approach to the measurement of poverty: a reply to Professor Peter Townsend. *Oxford Economic Papers*, 37(4), 669-676.
- Sen, A. (1992). Sobre conceptos y medidas de pobreza. *Comercio exterior*, 42(4), 310-322.
- Verba, S., Schlozman, K. L., Brady, H., & Nie, N. H. (1993). Citizen activity: Who participates? What do they say?. *American Political Science Review*, 87(2), 303-318.

Verner, D. (2004). Education and its poverty-reducing effects: The case of Paraíba, Brazil (Vol. 3321). World Bank Publications.

Vivas Pacheco, H. (1996). Perfiles de Bienestar de los Grupos Vulnerables en Cali-Yumbo. Centro de Investigaciones y Documentación Socioeconómica CIDSE.

Wagle, U. (2000). The policy science of democracy: The issues of methodology and citizen participation. *Policy Sciences*, 33(2), 207-223.

Wagle, U. (2002). Volver a pensar la pobreza: definición y mediciones. *Revista Internacional de Ciencias Sociales*, 171, 18-33.

Palabras Clave: pobreza multidimensional, perfiles de pobreza, modelo de regresión logística, EPH.

Área de Aplicación: Economía.

ACTITUD DE UNA COMUNIDAD LOCAL HACIA EL DESARROLLO DEL TURISMO MEDIANTE EL USO DE MODELOS DE ECUACIONES ESTRUCTURALES

Laura Isabel Luna y Norma Patricia Caro

Facultad de Ciencias Económicas – Universidad Nacional de Córdoba

lauraisabel.luna@gmail.com; pacaro@eco.unc.edu.ar

Resumen

El turismo es una actividad económica de fundamental importancia en el desarrollo económico de una región, en este caso en la provincia de Córdoba. El apoyo de los residentes locales es esencial para su desarrollo. Una vez que una comunidad se convierte en un destino, la calidad de vida de los residentes locales se ve afectada por las consecuencias de este desarrollo. La teoría del intercambio social (Blau, 1964) postula que las actitudes de un individuo como integrante de la comunidad hacia la industria del turismo, y el apoyo que muestre para su desarrollo están influenciadas por la evaluación que realice de las repercusiones que tiene para la comunidad tal industria. De esta forma, la manera en que los residentes perciben los elementos económicos, socioculturales y medioambientales del intercambio, afecta la manera en que reaccionan al turismo. Por tanto, se torna esencial el estudio de la percepción de los impactos por parte de la población local en aquellos lugares que son destinos turísticos. Comprender la reacción local y los factores que influyen en estas actitudes es esencial para lograr el objetivo de un apoyo favorable para un turismo sustentable (Cardona, 2012). Esta

investigación se enmarca en un proyecto más amplio que pretende profundizar y avanzar en el estudio de los factores e interrelaciones que influyen en las actitudes de los residentes y por tanto en el desarrollo turístico en las localidades turísticas de la provincia de Córdoba. Se realizó un trabajo empírico en la ciudad de Cosquín durante la realización del Festival Nacional de Folklore durante el mes de enero de 2018 con un cuestionario de 15 ítems basado en el propuesto por Gursoy (2002), cuyo procesamiento fue realizado con modelos de ecuaciones estructurales.

Estos modelos estadísticos permiten estimar el efecto y las relaciones entre múltiples variables. Nacen de la necesidad de dotar de mayor flexibilidad a los modelos de regresión. Son menos restrictivos por el hecho de permitir incluir errores de medida tanto en las variables criterio (dependientes) como en las variables predictoras (independientes).

Se definieron dos constructos: *Beneficios generales* y *Costes generales* que permiten medir la valoración que realizan los residentes de diversos tipos de efectos del turismo. Estas variables se ven afectadas por diversos elementos (apego, preocupación económica y por la comunidad, ecocentrismo, valoración del grado de dependencia económica, etc.) y afectan a la valoración general del turismo que realizan los individuos.

Para cada uno de estas variables latentes se construyó un modelo de medida que poseen coeficientes de fiabilidad aceptables (alphas de Cronbach de 0,722 y 0,754). El constructo *Beneficios Generales* está compuesto por seis indicadores formativos que miden los principales beneficios económicos asignados a la actividad económica: oportunidades de empleo, mayores inversiones empresariales, mayores oportunidades de negocio para la población local, mejora en la infraestructura de la localidad, mejora en la calidad de servicios, además de beneficios culturales. Por otro lado, el constructo *Costos Generales* está compuesto por ocho indicadores formativos que miden diversos costos percibidos por los residentes: pérdida cultural, aumento del costo de vida, aumento de robos, problema de convivencia entre residentes y turistas, entre otros.

Se calcularon medidas de bondad de ajuste que muestran que ambas variables latentes explican el conjunto de indicadores definidos.

Referencias bibliográficas

- Allen, L. R.; Hafer, H. R.; Long, P. T. & Perdue, R. R. (1993). Rural residents' attitude estoward recreation and tourism development. *Journal of Travel Research* 32(1): 27-33
- Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.
- Befu, Harumi (1977). *Social Exchange*. *Annual Review of Anthropology*, 6, 225-281.
- Butler, R. W. (1980). The concept of a tourist area cycle of evolution: Implications forthe management of resources. *The Canadian Geographer* 24(1): 5-12
- Cardona, J. R. (2012). *Actitudes de los residentes hacia el turismo en destinos turísticos consolidados: el caso de Ibiza*. Tesis Doctoral . Universitat de les Illes Balears.
- Cardona, J., Serra Cantallops, A. (2014). Inicios del turismo y actitudes de los residentes: El caso de Ibiza – España. *Estudios y perspectivas en turismo*, 23(1), 01-22.
- Fredline, E. & Faulkner, B. (2000). Host community reactions: A cluster analysis". *Annals of Tourism Research* 27(3): 763-784
- García F., Balbuena Vázquez A., Cortés Macías R. (2015). Resident's attitudes towards the impacts of tourism, *Tourism Management Perspectives*, Volume 13, Pages 33-40.
- Gutierrez Taño, D. (2010). *Las actitudes de los residentes ante el turismo*. Tesis Doctoral. Universidad de La Laguna. España.
- Gursoy, D., J. S. Chen, & Y. Yoon. (2000). Using Structural Equation Modeling to Assessthe Affects of Tourism Impact Factors and Local Residents Support for Tourism Development. In annual conference proceedings of the Travel and Tourism. Research Association, pp. 243-250. Boise ID: TTRA.
- Gursoy, D.J., Jurowski, C. & Uysal, M. (2002). Resident Attitudes: A Structural Modeling Approach. *Annals of Tourism Research*, 29(1):79-105.
- Jurowski, C. (1994). *The Interplay of elements Affecting Host Community Resident Attitudes toward Tourism: A Path Analytic Approach*. PhD dissertation. Virginia Polytechnic Institute and State University.
- Jurowski, C., Uysal, M. & Williams, R.D. (1997). A Theoretical Analysis of Host community Resident Reactions to Tourism. *Journal of Travel Research*, 36(2):3-11

González García1 R; Sanz V.(2018) Perception of residents about the impact of sports tourism on the community: Analysis and scale-validation. Journal of Physical Education and Sport, 18(1), Art 19, pp. 149 – 156.

Palabras Clave: Ecuaciones estructurales- Variable latente – Turismo - Comunidad local

Área de Aplicación: Economía

Educación, Ciencia y Cultura

A horizontal orange arrow pointing to the right, with the text "Educación, Ciencia y Cultura" written in blue inside it.

ADAPTACIÓN Y ANÁLISIS DE UNA ESCALA DE MEDIDA PARA ESTRATEGIAS DE APRENDIZAJE AUTÓNOMO EN EDUCACIÓN A DISTANCIA

Adrián M. Moneta Pizarro y María Alejandra Juárez

Universidad Nacional de Córdoba

amoneta@eco.uncor.edu / juarezmaale@hotmail.com

Resumen

La Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de Córdoba (UNC) cuenta con un Ciclo Básico a Distancia (CBD) para sus carreras de grado. El objetivo de este CBD es ofrecer una alternativa de cursado que resuelva con mayor eficacia problemas de acceso, deserción y retraso académico de los alumnos, especialmente en asignaturas de los primeros años (Moneta Pizarro et al., 2017). Sin embargo, estudios exploratorios advierten que la proporción de alumnos que logran la regularidad en estos cursos no llega a 40 % y que entre las posibles causas se encontrarían problemas relacionados con deficiencias en las habilidades de los estudiantes para el aprendizaje autónomo (Blanch et al., 2013 octubre).

Desde los comienzos de la educación a distancia, los estudios sobre esta opción pedagógica destacan la importancia que tiene desplazar el énfasis de las propuestas formativas desde la enseñanza hacia el aprendizaje. El uso de estrategias de aprendizaje autónomo es un factor diferencial para explicar el rendimiento académico de los alumnos (De Miguel, 2006). Resulta entonces fundamental contar con escalas de medición de las habilidades de aprendizaje autorregulado y que éstas sean válidas para identificar su asociación con el desempeño académico de los estudiantes.

El objetivo general de este trabajo es proponer un modelo de medida válido para el aprendizaje autónomo en educación a distancia. Los objetivos específicos son adaptar el Cuestionario de Estrategias de Trabajo Autónomo (CETA) para estudiantes universitarios (López Aguado, 2010) y validarlo estadísticamente sobre una muestra de 290 alumnos del curso de Microeconomía I del CBD de la FCE-UNC con métodos de análisis factorial y modelación de ecuaciones estructurales.

Los resultados demuestran consistencia interna de la escala total adaptada ($\alpha=0.8633$) y de cada una de las estrategias de aprendizaje identificadas como sub-escalas (α colaboración=0.8909, α síntesis=0.8365, α ampliación=0.8848 y α planificación=0.7552). Las estimaciones del modelo de medida resultaron satisfactorias

reportando índices con bondad de ajuste satisfactoria (RMSEA=0.048; CFI= 0.952; SRMR=0.052). Esto evidencia que el instrumento propuesto es confiable y útil para obtener indicadores de estrategias de aprendizaje autónomo en cursos con características similares a los del CBD de la FCE-UNC y para ser aplicado en investigaciones que relacionen estos constructos con otras variables relevantes.

Referencias bibliográficas

Blanch, N.; Andreone, A.; Bravino, L.; Margaría, O.; Montero, L.; y Juárez, M. A. (2013, octubre). Evaluación del sistema de educación a distancia en la Facultad de Ciencias Económicas de Córdoba. Póster presentado en XLI Coloquio Argentino de Estadística, Mendoza, Argentina.

De Miguel, M. (2006). Metodologías de enseñanza y aprendizaje para el desarrollo de competencias. Orientaciones para el profesorado universitario ante el Espacio Europeo de Educación Superior. Madrid: Alianza.

López Aguado, M. (2010). Diseño y análisis del Cuestionario de Estrategias de Trabajo Autónomo (CETA) para estudiantes universitarios. Revista de Psicodidáctica, 15(1), 77-99.

Moneta Pizarro, A.; Montero, L.; Juárez, M.; Depetris, J.; y Fagnola, B. (2017). Adaptación y validación de un instrumento de medida para la interacción en b-learning. Virtualidad, Educación y Ciencia, 14 (8), 27-41.

Palabras clave: Aprendizaje autónomo - Educación a distancia - Análisis factorial - Ecuaciones estructurales.

Área de aplicación: Educación, ciencia y cultura.

ESTUDIO DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES UNIVERSITARIOS EN EL CICLO BÁSICO. EL CASO DE LAS ASIGNATURAS DEL DEPARTAMENTO DE MATEMÁTICA Y ESTADÍSTICA DE LA FACULTAD DE CIENCIAS ECONÓMICAS/UNRC.

María del Carmen Regolini, Sonia Curti, Daniel Ivars

Facultad de Ciencias Económicas. Universidad Nacional de Río Cuarto
maregolini@fce.unrc.edu.ar – soniacurti@yahoo.com.ar – divars@yahoo.com.ar

Resumen

Desde el año 2003, la Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de Río Cuarto (UNRC) ofrece el dictado de sus carreras de grado, Licenciatura

en Economía, Licenciatura en Administración y Contador Público, en dos modalidades: presencial y a distancia, otorgándole a esta Unidad Académica su impronta *bimodal*. Los Planes de Estudios de estas carreras tienen un Ciclo Básico común, con una duración de dos años y son idénticos para ambas modalidades de cursado.

El rendimiento académico es una preocupación para las instituciones académicas ya que el desempeño de sus estudiantes es un reflejo del proceso de formación institucional y brinda un indicador que permite una aproximación a la realidad educativa (Díaz et al., 2002 en Garbanzo Vargas, 2007).

En el Proyecto de Investigación “*De la Educación Superior Bimodal: Análisis del proceso de enseñanza-aprendizaje en asignaturas del Ciclo Básico del Departamento de Matemática y Estadística de la Facultad de Ciencias Económicas de la UNRC*”, aprobado y subsidiado por la Secretaría de Ciencia y Técnica de la UNRC, las preguntas que orientan este estudio aluden a: ¿Cuáles son los elementos característicos del proceso de enseñanza-aprendizaje en las asignaturas del Ciclo Básico del Departamento de Matemática y Estadística de la FCE de la UNRC? ¿Qué distinguen a los procesos de enseñanza-aprendizaje en las asignaturas del Ciclo Básico del Departamento de Matemática y Estadística de la FCE de la UNRC, en ambas modalidades de cursado? Fijando entre sus objetivos específicos, establecer similitudes y diferencias entre los procesos de enseñanza-aprendizaje de las asignaturas correspondientes al Ciclo Básico del Departamento de Matemática y Estadística de la FCE de la UNRC por modalidad de cursado.

Para dar respuesta a estos interrogantes se vienen desarrollando diversos análisis de tipo exploratorio y descriptivo, referidos a examinar tanto las condiciones finales de cursado obtenidas por los estudiantes de ambas modalidades como su rendimiento académico, a través del retraso en la aprobación de los exámenes finales en las cuatro asignaturas del Departamento de Matemática y Estadística (DMyE) correspondientes al Ciclo Básico (CB). La información utilizada proviene del Sistema Integral de Alumnos para Docentes (SIAL) de la UNRC y tiene un alcance temporal comprendido entre los años 2003 y 2016.

Las condiciones finales de cursado se estudian siguiendo a Rodríguez, Fita y Torrado (2004) los cuales proponen un indicador preciso y accesible para valorar el rendimiento académico, asumiendo el logro en diferentes componentes del aprendizaje, donde se incluyen aspectos personales y sociales, que se incorporan a los académicos.

Sin embargo, las calificaciones obtenidas por los estudiantes en los exámenes finales, si bien representan un indicador de rendimiento considerado inmediato, han suscitado diversas críticas entre los investigadores educativos debido a que no logra diferenciar entre un aprendizaje profundo de otro considerado superficial (Valle et al., 2003).

Los estudios e investigaciones referidos a logros académicos comprenden numerosas variables que inciden en su desempeño académico. Dentro de éstas, Rodríguez, Fita y Torrado (2004), asumen que las calificaciones obtenidas reflejan los logros académicos en los diferentes componentes del aprendizaje, donde incluyen aspectos personales, académicos y sociales. Manifestando que el desempeño académico es un fenómeno multicausal, pues envuelve una enorme capacidad explicativa de los distintos factores y espacios temporales que intervienen en el proceso de aprendizaje.

El equipo de investigación de este Proyecto, definió una nueva variable para determinar el tiempo que transcurre entre el año en que un estudiante se inscribe para cursar una asignatura y el año en que logra aprobar su examen final; el rezago en la aprobación de las materias, se determina mediante la distribución de frecuencia de esta variable. Los resultados obtenidos han permitido avanzar sobre la problemática abordada desde diferentes perspectivas complementarias y dejan al descubierto una situación muy diferenciada en ambas modalidades de cursado. A saber, entre los años 2003-2015, el 50% de los estudiantes presenciales logra la condición final de cursado *Regular* en las materias de primer año elevándose a valores alrededor del 70% en las correspondientes asignaturas de segundo año. En cambio, en la modalidad a distancia dichos porcentajes oscilan alrededor del 20% y del 30% respectivamente. En el período comprendido entre los años 2008 y 2016, el estudio de las condiciones finales de cursado *Libre*, en sus diferentes categorías, arrojó que en la modalidad presencial, la mayor proporción corresponde a los estudiantes *Libre por parcial* mientras que en la modalidad a distancia, la proporción de estudiantes *Libre* es muy elevada con respecto a la condición final de cursado *Regular*, en las cuatro materias del Ciclo Básico; destacándose la asignatura Estadística y Probabilidades, por ser la que presenta la menor diferencia entre los porcentajes de estas dos condiciones finales de cursado. En general, en la modalidad a distancia, dentro de la condición final de cursado Libre la categoría que ejerce mayor influencia es *Libre por faltas*.

Por su parte, el análisis del retraso en la aprobación de las cuatro asignaturas del DMyE comprendido entre los años 2004 y 2014 ha revelado que durante el primer año,

desde que se efectiviza la inscripción para el cursado de Análisis Matemático I, Análisis Matemático II, Álgebra Lineal y, Estadística y Probabilidad sólo el 33% logra su aprobación, en la modalidad a distancia mientras que el 55% en la modalidad presencial. En cambio, las proporciones de aprobación en estas materias con un año de retraso respecto al período de inscripción de su cursado, son del 35% y el 33%, respectivamente (Regolini et al., 2016 y Curti et al., 2017).

Las conclusiones de estas tareas realizadas podrían estar reflejando diversas particularidades, entre las cuales cabe mencionar que: a medida que transitan los estudiantes por las diferentes asignaturas del DMyE, correspondientes al Ciclo Básico, la condición final de cursado Libre, disminuye. La identificación de estudiantes con distintos rendimientos académicos, vislumbraría qué factores dificultan lograr con éxito del proceso de enseñanza-aprendizaje en estas asignaturas que cursan del DMyE. Así como, la diferenciación temporal que transcurre en la aprobación de las asignaturas del Ciclo Básico del Departamento de Matemática y Estadística dentro del proceso de enseñanza-aprendizaje, respecto a la modalidad de cursado.

Finalmente, se menciona que los resultados obtenidos de los trabajos efectuados quedan a disposición de las Autoridades de la FCE de la UNRC, en virtud de que se aprecia su potencial aporte de utilidad en la toma de decisiones de gestión académica.

Referencias bibliográficas

- Curti, S., Regolini, M.C. e Ivars D. (2017). Una mirada del abandono en la FCE de la UNRC: análisis en el Área Matemática del Ciclo Básico. VII Conferencia Latinoamericana sobre el Abandono en la Educación Superior. UNC. Argentina. Publicado en Actas de la Conferencia: <http://viiclables-alfaguia.eco.unc.edu.ar/trabajos-presentados#linea-3-practicas-curriculares>
- Garbanzo Vargas, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación* 31(1), 43-63. ISSN: 0379-7082. Costa Rica.
- Latiesa, M. (1992). La deserción universitaria. Madrid, siglo xxi-cis, nº 124.
- Moncada Mora, Luis F. y Rubio Gómez, María José. (2011) Determinantes inmediatos del rendimiento académico en los nuevos estudiantes matriculados en el sistema de educación superior a distancia del Ecuador: caso Universidad Técnica Particular de Loja. *RIED: Revista Iberoamericana de Educación a Distancia*, (14,

n.2), 2011, p. 77-95. ISSN: 1138-2783. [Fecha de consulta: 31 de agosto de 2017]

Montero Rojas, E., Villalobos Palma, J. y Valverde Bermudez, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. *Revista Electrónica de Investigación y Evaluación Educativa*, 215-234.

Regolini, M.C., Curti, S. e Ivars D. (2016). Estudio comparativo de las condiciones finales de cursado en asignaturas del Área Matemática. Facultad de Ciencias Económicas Bimodal de la UNRC. 7º Seminario Internacional de Educación a Distancia Enseñar en la virtualidad: Nuevas Presencialidades en la Educación Superior. Organizado por la Red Universitaria de Educación a Distancia de Argentina (RUEDA). Sede UNL y la UNER. ISBN: en trámite.

Regolini, M.C. y Curti S. (2016). Trayectoria académica de los estudiantes presenciales en el Ciclo Básico de la FCE/UNRC: Asignaturas del Dpto. de Matemática y Estadística en el período 2003-2005. XXIII Jornadas de Intercambio de Conocimientos Científicos y Técnicos. SCyT. FCE. UNRC. Río Cuarto. Argentina. Trabajo publicado completo en compilación formato Ebook ISBN: 978-987-688-197-5.

Regolini, M.C., Ivars D. y Curti, S. (2017). Rendimiento académico en el Ciclo Básico: una medida del retraso en la aprobación de las asignaturas en la Facultad Bimodal de la UNRC. XXIV Jornadas de Intercambio de Conocimientos Científicos y Técnicos. Secretaría de Ciencia y Técnica. FCE. UNRC. Río Cuarto. Argentina. Publicación en trámite.

Rodríguez, S., Fita, S., Torrado, M. (2003). El rendimiento académico en la transición secundaria-universidad. *Revista de Educación. Temas actuales de enseñanza*, 334, Mayo-Agosto.

Tejedor Tejedor, F.J. y García-Valcárcel Muñoz-Repiso, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, 342, 443-473.

Palabras Clave: Condición final de cursado- Estudiantes – Facultad bimodal – Ciclo Básico – Área matemática

Área de Aplicación: Educación, Ciencia y Cultura.

EVALUACIÓN DEL PERFIL DE LOS ESTUDIANTES DE UN CURSO DE GRADO DE ESTADÍSTICA

María Inés Stimolo¹-Mariana Guardiola²

Facultad de Ciencias Económicas – Universidad Nacional de Córdoba

¹mstimolo@eco.unc.edu.ar; ²marianaguardiola@eco.unc.edu.ar

Resumen

En un primer curso de estadística de las carreras de grado de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, se administra a los estudiantes un cuestionario donde se les consulta sobre aspectos personales, académicos y algunos hábitos de vida. Esta encuesta está diseñada con fines didácticos puesto que es utilizada en diferentes actividades de aprendizaje durante el dictado del curso.

Particularmente, la base de datos resultante es utilizada como herramienta de aprendizaje del manejo de un *software* estadístico y para la revisión de algunos contenidos del curso; en algunas cohortes se solicitó su análisis a partir de un trabajo práctico grupal, mientras que en los últimos años se está proponiendo una evaluación parcial optativa en gabinete informático en la cual los estudiantes deben, a partir del procesamiento de los datos y análisis de los reportes obtenidos, dar respuesta a una serie de consignas a través de una evaluación virtual del aula *Moodle* del curso. Cabe aclarar que la base de datos es puesta a disposición de los estudiantes con anterioridad a dicha instancia de evaluación a los efectos de que puedan utilizarla a modo de ejercitación y revisión de los contenidos a evaluar.

En este trabajo se tuvo en cuenta la información de los estudiantes de las cohortes 2015-2018 y las distintas escalas de medición de las variables involucradas, considerando por un lado las referidas a aspectos personales y de estilo de vida, y por otro, las relacionadas con su desempeño académico.

Un primer análisis consistió en representar la información en un espacio multidimensional reducido a partir de la aplicación de técnicas multivariadas, con el objetivo de identificar las variables más relevantes al mismo tiempo que lograr reducir la dimensión. Luego, en una segunda instancia, se identificaron grupos de estudiantes para cada una de las cohortes utilizando distintas metodologías estadísticas de agrupamiento; algunas que consideran sólo variables numéricas (*k-means*), otras que

permiten sólo variables categóricas -en cuyo caso se procedió a categorizar las variables numérica (*k-modas*) y las que permiten combinar distintas escalas de medición de las variables (*k-prototipos*). De la comparación de los resultados obtenidos de la aplicación de estas metodologías, se propone un abordaje de análisis del perfil de los estudiantes a partir de esta encuesta.

De esta manera, un instrumento diseñado con fines académicos resultó ser una herramienta útil para analizar el perfil de los estudiantes en grupos masivos en los cuales resulta complejo alcanzar un adecuado conocimiento de sus características. En este sentido, se presenta el diseño de una metodología que permita utilizar esta encuesta para definir las características del curso y así proponer estrategias didácticas diferenciadas.

Referencias bibliográficas

Aldas Manzano, Joaquin y Uriel Jimenez, Ezequiel (2017) Análisis multivariante aplicado con R. 2ª ed. Paraninfo.

Huang, Z. (1997) A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. in KDD: Techniques and Applications (H. Lu, H. Motoda and H. Luu, Eds.), pp. 21-34, World Scientific, Singapore.

Huang Z. (1998): Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Variables, *Data Mining and Knowledge Discovery* 2, 283-304.

Husson, F., Le, S. and Pages, J. (2010). *Exploratory Multivariate Analysis by Example Using R*, Chapman and Hall.

Le, S., Josse, J. & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*.25(1). pp. 1-18.<http://www.jstatsoft.org/v25/i01/>

Pinto Javier Rivera (2014) Clustering de series de tiempo con datos Categóricos. Trabajo de tesis para el Máster en técnicas estadísticas. Universidad de Coruña. Universidad de Santiago de Compostela. Universidad de Vigo

Palabras Clave: Clasificación no Supervisada – Análisis Multidimensional- Análisis de Perfiles.

Área de Aplicación: Educación, Ciencia y cultura

Enseñanza de la Estadística



**COMPRESION DE INFORMACIÓN CUALI-CUANTITATIVA
ASIGNATURA DEL INGRESO UNIVERSITARIO. UNA MATERIA
INTERDISCIPLINARIA**

Diana Rosana Kucukbeyaz - Hernan Antonio Barrios

UNTREF

ranush2003@yahoo.com.ar; hbarrios@untref.edu.ar

Resumen

Introducción:

Los cambios producidos en la estructura social durante las últimas décadas se han traducido al interior de las universidades bajo la forma de un conjunto de nuevas tensiones que impactan directamente en la dinámica de su funcionamiento. Por tanto, en el caso de los estudiantes, el desafío que se les presenta es doble: afrontar los retos de la nueva experiencia en un momento en el que al sistema universitario se encuentra en un proceso de transformación. En tal sentido, el presente trabajo busca indagar y dar cuenta de algunos de los impactos generados por un sistema universitario “en expansión-consolidación” en locaciones geográficas inéditas, junto a la llegada masiva a estas nuevas casas de estudio de los llamados estudiantes de nuevo ingreso.

Para este fin se analizará un caso en particular, la Universidad Nacional de Tres de Febrero y la constitución de políticas académicas que buscan analizar, trabajar y ofrecer alternativas a algunas de las tensiones que se presentan como la adquisición de herramientas para afrontar esta nueva etapa de estudio y la tensión entre masividad y calidad.

En este contexto se plantea como necesario pensar las estructuras universitarias como espacios democráticos, democratizantes y democratizadores en donde se puedan desarrollar y consolidar prácticas institucionales, pedagógicas y sociales. El análisis de estas prácticas para constituir un espacio democratizador lo constituye “comprensión de la información cuanti-cualitativa” que se dicta en el ingreso de las carreras de grado. Modelo de asignatura interdisciplinaria donde el dictado se realiza en la actualidad por “una pareja pedagógica”, un docente del área de la estadística y un docente del área de historia.

Objetivos:

Que los participantes

Reconozcan la relación existente entre la problematización del campo disciplinar y la articulación de los contenidos de la escuela media y/o universidad a partir de situaciones áulicas concretas.

Adquieran las capacidades y destrezas necesarias para aplicar técnicas de análisis estadístico a la resolución de problemas de diversa índole.

Reflexionen en Incrementar la capacidad crítica, respecto de las posibilidades que la Estadística tiene para abordar y resolver problemas que se plantean en los distintos campos del saber tanto desde el análisis cuantitativo como el análisis cualitativo.

Adquirir habilidades para elaborar proyectos curriculares que impliquen cambios en la organización del aula y en la adecuación del desempeño docente en contextos diversos.

Referencias bibliográficas

Arquina, Mónica, compositor/a *La universidad entre la autonomía y la planificación :*

Tres ensayos en diálogo. -- Buenos Aires (Argentina) : Federación Nacional de Docentes Universitarios. Instituto de Estudios y Capacitación. ;

Universidad Nacional de General Sarmiento, 2014.

Arzoumanian de Kalayci, Ana *La Universidad posmoderna.* -- Buenos Aires

(Argentina) : Grupo Editor Latinoamericano, 199.

Fernández Lamarra N. La educación superior argentina en debate. Situación, problemas y perspectivas, Bs. As, IESALC- UNESCO y Eudeba, 2003.

Gentili, P. “Una vergüenza menos, una libertad más”. En “La reforma universitaria, desafíos y perspectivas noventa años después”. Sader, Emil; Gentili, Pablo; Aboites, Hugo; editores. CLACSO. Bs. As. 2008.

Gentili, Pablo. “Pedagogía de la igualdad. Ensayos sobre la educación excluyente”.

CLACSO. Siglo XXI Bs. As. 2011

Imbernón, Francisco, coordinador/a Formación del profesorado universitario para el fortalecimiento institucional : Análisis de necesidades de la Universidad Nacional de Tres de Febrero (Argentina). Caseros (Argentina): Eduntref, 2004.

Palabras Clave: Interdisciplinariedad – Pareja Pedagógica – Tensiones/espacios democratizantes.

Área de Aplicación: Enseñanza de la Estadística

ESTADÍSTICA DESCRIPTIVA. IMPLEMENTACIÓN DE R-COMMANDER

Alvarez E., Cámpora J.J., Cuesta C., Meroi N.

Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

normameroi@outlook.com

Resumen

Las nuevas tecnologías de la información y de la comunicación están produciendo un gran impacto y transformación en la sociedad, en la cultura y, por extensión, en la educación. Estas tecnologías abren nuevas concepciones para la enseñanza, favorecen el cambio y la mejora en nuestras aulas, especialmente en campos específicos de la educación.

Nuestras instituciones de formación superior, para responder a estos desafíos, deben revisar sus referentes actuales y promover experiencias innovadoras en los procesos de enseñanza-aprendizaje apoyados en las nuevas tecnologías de la Información y la Comunicación.

Un administrador y/o contador debe comprender los conceptos estadísticos y sus aplicaciones, lograr apreciar el poder de las herramientas estadísticas y, especialmente, reconocer situaciones en las cuales puede hacer uso efectivo de las mismas.

Las organizaciones recopilan datos sobre sus operaciones internas a través de sistemas contables y otros sistemas de reportes de datos. Estas masas de datos, por sí solas, tienen poco significado, por lo cual resulta necesario procesarlos y resumirlos con aplicación de métodos estadísticos.

Hoy en día los cursos de estadística para administradores y contadores han efectuado un cambio de paradigma, desplazando el cálculo hacia la atenta selección de métodos y la interpretación crítica de los resultados. Debido a este cambio, en la asignatura Métodos Estadísticos y Estadística para Administradores de la Facultad de Ciencias Económicas y Estadística de la Universidad de Rosario, los docentes nos estamos replanteando nuestro quehacer en el aula y evaluando la factibilidad de implementar el uso de nuevas tecnologías mediante el programa R Commander en el dictado de clases. Se elige este programa por ser un software de libre acceso, es decir gratuito, al que todo estudiante podrá tener alcance, además de ser de fácil manejo para el trabajo con bases de datos.

En el presente trabajo se muestra la implementación del software R-Commander como un estudio preliminar en sólo dos comisiones de la cátedra Métodos Estadísticos y Estadística para Administradores con el fin de mejorar la enseñanza y el aprendizaje en nuestras aulas.

Para llevar adelante la propuesta se hizo entrega de un material escrito. En el mismo, además de presentar y describir las bondades del software, se detallaron todos los pasos a seguir para su instalación y cómo utilizarlo en la resolución de un ejercicio de la práctica correspondiente a análisis descriptivo, el cual debió ser entregado en tiempo y forma. Uno de los objetivos principales fue la aplicación del software y la evaluación de la propuesta, así como la deducción, por parte de los estudiantes, del ahorro de tiempo en la resolución de ejercicios, propiciando mejoras para la interpretación de resultados.

Otra de las finalidades de la experiencia fue la valoración por parte de los alumnos del material brindado para incorporar las mejoras que se crean necesarias. Dicha devolución la realizaron a partir de una encuesta donde, entre otros puntos, mencionaron las dificultades con las que se encontraron en la lectura del material y aplicación del programa. Por otro lado, para evaluar el aprendizaje, se incorporó, en la primera actividad evaluativa, de las dos comisiones mencionadas, un ejercicio con una salida de este software para que a partir de la misma realicen una interpretación de los resultados con sus correspondientes fundamentos teóricos.

Toda esta información será tenida en cuenta para evaluar la factibilidad de una futura implementación del software no sólo en las demás unidades de la materia, sino también en todas las comisiones de la cátedra.

Referencias bibliográficas

- DRISCOLL, Marcy P, VERGARA, Adriana (1997) "Nuevas tecnologías y su impacto en la educación del futuro" Pensamiento Educativo. Vol. 21
- GUTIÉRREZ MARTÍN, Alfonso (1997): Educación Multimedia y Nuevas Tecnologías. Ediciones de la Torre.
- SALINAS, J., de BENITO, B. y PÉREZ, A. (1.999): "Tecnologías de la Información y la Comunicación en la Enseñanza universitaria: el caso de la UIB". Comunicación. I Simposium Iberoamericano de Didáctica universitaria: La Calidad de la docencia universitaria. Universidad de Santiago de Compostela.

<https://cran.r-project.org/doc/contrib/curso-R.Diaz-Uriarte.pdf>. <https://www.r-project.org/http://www.tutorialr.es/es/descripcion.html>

Palabras clave: Propuesta didáctica, R-Commander, Contador Público, Licenciatura en Administración, Análisis descriptivo

Área de aplicación: Enseñanza de la Estadística

PAQUETE SHINY: UNA HERRAMIENTA R PARA LA ENSEÑANZA DE DISTRIBUCIONES DE PROBABILIDAD

Mariana Silvia Santamaría

Departamento de Matemática, Universidad Nacional del Sur,

Bahía Blanca, Argentina

msanta@criba.edu.ar

Resumen

El estudio de las distribuciones de probabilidad forma parte de los programas de los cursos básicos relacionados con Probabilidad y Estadística a nivel de la enseñanza universitaria Argentina. Dado que en la actualidad el uso de la tecnología constituye un recurso didáctico de gran utilidad en el proceso de enseñanza/aprendizaje en diversas disciplinas, en este trabajo se presenta una propuesta didáctica de fácil manejo consistente en una aplicación web desarrollada con el lenguaje de programación R y su paquete Shiny. Sirve para representar gráficamente distintas distribuciones de probabilidad de manera interactiva, al seleccionar el tipo de distribución de una lista desplegable y los valores de sus parámetros mediante barras de desplazamiento. Resulta un complemento atractivo en la enseñanza de las distribuciones permitiendo al alumno comprender las diferencias entre distribuciones de variables aleatorias discretas y continuas, y revisar conceptos como simetría y curtosis y aproximación entre distribuciones de probabilidad. El programa R y el paquete Shiny son libres y se encuentran disponibles en internet a través del Comprehensive R Archive Network (CRAN, <https://cran.r-project.org/>). Shiny permite construir aplicaciones web interactivas sencillas mediante código R. Una característica importante es que utiliza programación reactiva y posee una comunicación rápida y eficiente entre el navegador web del usuario y R (Gómez, Molina, Mulero, Nueda, y Pascual, 2016). La aplicación propuesta fue desarrollada utilizando el paquete ggplot2 de R para realizar los gráficos de las distribuciones y las instrucciones o widgets provistos por Shiny para generar la

interactividad y reactividad. Las distribuciones de probabilidad incluidas fueron binomial, Poisson, normal, t de student, chi cuadrado y F (Canavos, 1988). La aplicación está compuesta por dos componentes: la interfaz del usuario, ui.R, y la función server, server.R. La interfaz ui.R contiene la información sobre la disposición y apariencia de la página web resultante, los comandos necesarios para generar la lista desplegable (select box) con los nombres de las distribuciones, las barras de desplazamiento (slider) con los valores de sus parámetros y un identificador para el gráfico de salida. La función server.R contiene una expresión reactiva que incluye una sentencia de control condicional para generar el gráfico de salida de la distribución seleccionada y que permite que un cambio en el tipo de distribución o en sus parámetros se refleje en una actualización inmediata del gráfico. La ejecución de aplicación se realiza desde el editor de RStudio abriendo los dos componentes y luego utilizando el botón Run App. Se abre una página web que contiene la lista desplegable de distribuciones de probabilidad, debajo las barras de desplazamiento con los parámetros y a continuación el gráfico de la distribución seleccionada. El uso de la aplicación en el aula resultó muy útil. La apariencia visual de la página web es atractiva y clara posibilitando una participación activa del alumno en el aprendizaje del tema propuesto. La aplicación se puede poner a disposición del alumno como página web a través de una ruta URL.

Referencias bibliográficas

Canavos, G. C. (1988). *Probabilidad y Estadística. Aplicaciones y Métodos*. México: McGraw Hill.

Gómez, D. S., Molina, M. D., Mulero, J., Nueda, M. J., y Pascual, A. (2016). Aplicaciones diseñadas con Shiny: un recurso R para la enseñanza de la estadística. *XIV Jornadas de Redes de Investigación en Docencia Universitaria*, 2029-2041. Recuperado de: https://rua.ua.es/dspace/bitstream/10045/59411/1/XIV-Jornadas-Redes-ICE_149.pdf

Palabras Clave: Propuesta didáctica – Aplicación Shiny – Distribuciones de probabilidad – Representación gráfica.

Área de Aplicación: Enseñanza de la Estadística.

Estadísticas Oficiales



EVALUACIÓN DEL DESARROLLO PROVINCIAL SEGÚN UN INDICADOR SINTÉTICO

M. Laura Vignera^{1,2}, Carla R Martínez¹

1. División Estadística, Departamento de Ciencias Básicas, UNLU

2. Área Estadística, Departamento Ciencias Básicas, UNNOBA

mlauravig@gmail.com

Resumen

Un indicador o índice sintético, es una combinación de indicadores individuales, que permite resumir en una única cifra el comportamiento de un conjunto de indicadores que dan cuenta de diversas variables relacionadas. Los indicadores sintéticos resumen información, de tal modo que facilitan la interpretación de la información y posibilitan la evaluación de fenómenos complejos y multidimensionales, de una manera sencilla. Son por lo tanto herramientas útiles para la toma de decisiones. (Joint Research Centre-European Commission, 2008).

Durante los últimos años los indicadores sintéticos se han configurado como uno de los instrumentos de medición analítica más utilizados en la práctica en multitud de campos de la realidad social. Cada vez más, los indicadores sintéticos son aceptados como herramientas útiles para la toma de decisiones y la comunicación de información (Domínguez Serrano, Blancas Peral, Contreras Rubio, & Ramírez Hurtado, 2011). Existen en la actualidad numerosos informes que analizan las variadas técnicas para el cálculo de indicadores compuestos. En este sentido, se han realizado revisiones generales que demuestran los progresos en las diferentes metodologías utilizadas (Saisana, Tarantola, & Saltelli, 2005), literatura en la que se debate los principales pros y contras en la utilización de indicadores compuestos.

Objetivo: El objetivo de este trabajo fue elaborar un Indicador Sintético de Desarrollo Provincial con el fin de medir el desarrollo de las provincias en la Argentina desde una perspectiva multidimensional, a partir de un sistema de indicadores simples. El propósito de este indicador compuesto, es facilitar una herramienta comunicacional y permitir elaborar un análisis comparativo.

Metodología: Sobre cada unidad de análisis, la provincia, se recopiló información estadística de acceso público, procedente de diferentes fuentes. Se identificaron en primera instancia las dimensiones que se deseaban medir y analizar. Para describir esas dimensiones se seleccionó un conjunto de variables que

constituyeron la base para explicar de manera eficaz el desarrollo provincial. Las dimensiones abarcaron aspectos de población y condiciones de vida, producción, educación y cultura, infraestructura y salud. Dichas variables dieron origen a los indicadores simples, a partir de los cuales se calculó el indicador sintético (IS). Con motivo de tornar a las variables comparables, previamente se normalizaron o estandarizaron los datos utilizando el método de re-escalamiento.

En el proceso de construcción del índice, se plantearon los siguientes interrogantes: ¿cada indicador o incluso cada dimensión que conforma el índice deberá tener el mismo peso o ponderación, o deberá tener cada uno un peso diferente? ¿Y si así se considera, cuál es la metodología adecuada para determinar la importancia de cada uno? La respuesta a estas preguntas, resulta muchas veces un punto crítico en la práctica. La ponderación consiste en atribuir un peso o importancia, y por ende un valor determinado a cada dimensión o a cada indicador simple. En este aspecto, el cálculo del IS se realizó a través del Método de Componentes Principales (ACP), técnica estadística de análisis multivariante (Cuadras, 2007),(Dallas, 2000), cuyo procedimiento matemático transforma un conjunto de p variables respuesta correlacionadas en un conjunto menor de variables no correlacionadas. ACP sólo es aplicable en el caso de que las variables se encuentren correlacionadas, condición que se verificó en este caso, observándose además un valor muy pequeño para el determinante de la matriz de correlaciones, hecho que demostró que el sistema se encuentra linealmente relacionado.

Las comunalidades obtenidas a través del ACP, fueron utilizadas en el cálculo del IS como los ponderadores de cada indicador simple. Éstas pueden interpretarse como la proporción de la variabilidad en una variable respuesta que puede ser explicada por las componentes principales. Finalmente, cada indicador ponderado por la comunalidad respectiva, fue agregado en un indicador sintético provincial. Asimismo, se obtuvieron indicadores parciales para cada una de las dimensiones establecidas.

Resultados: Siendo que la matriz original de datos consta de veinte indicadores simples, se logró reducir la dimensionalidad de las p variables a cinco componentes principales, las cuales explicaron el 78% de la variabilidad contenida en los datos.

La obtención del IS permitió establecer un ranking de desarrollo entre las provincias. Un retorno a los datos originales, manifestó las fortalezas y falencias de cada provincia. El procedimiento de agregación de información para la construcción de indicadores sintéticos implica, que el investigador deba tomar una serie de decisiones

subjetivas, que van a influir sobre los resultados. La elección de las variables e indicadores de partida, la forma de agruparlos, el uso de uno u otro método de normalización de los datos, la metodología de ponderación, constituyen todo un conjunto de acciones aportadas por el analista. Sin embargo, los indicadores sintéticos constituyen una herramienta de gran utilidad a la hora de evaluar conceptos multidimensionales, y en ocasiones, distintos métodos pueden resultar complementarios. Finalmente, la técnica metodológica aquí propuesta puede resultar aplicable en cualquier ámbito de investigación en el que se disponga de estadísticas válidas.

Referencias bibliográficas

- Cuadras, C. (2007). Analisis de Componentes principales. En C. Cuadras, *Nuevos Metodos de Analisis Multivariante* (págs. 63-76). Barceona, España: CMC Editions
- Dallas, J. (2000). *Métodos multivariados aplicados al análisis de datos*. Mexico: Thomson
- Domínguez Serrano, M., Blancas Peral, F. J., Contreras Rubio, I., & Ramírez Hurtado, J. M. (2011). Construcción de indicadores sintéticos: una aproximación para maximizar la discriminación. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 41-70. Obtenido de <http://www.upo.es/RevMetCuant/art.php?id=48>
- Joint Research Centre-European Commission. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD publishing

Palabras Clave: Herramienta – Multidimensional – Comparación

Área de Aplicación: Estadísticas oficiales.

Genética

An orange horizontal arrow pointing to the right, with the word "Genética" written in blue text inside it.

ANÁLISIS FACTORIAL MÚLTIPLE PARA LA CARACTERIZACIÓN DE PLANTAS DE BANANA SEGÚN EL AMBIENTE DE CRECIMIENTO INICIAL

Del Medico, A. P. ¹; Ermini, J. L. ¹; Tenaglia, G. ²; Lavalle, A. ³; Vitelleschi, M.S. ⁴;
Pratta, G.R. ¹

1 Instituto de Investigaciones en Ciencias Agrarias de Rosario, CONICET/UNR.

2 Instituto de Investigación y Desarrollo Tecnológico para la Agricultura Familiar,
INTA.

3 Departamento de Estadística. Universidad Nacional del Comahue.

4 Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística,
CIUNR/UNR.

mvitelleschi@fcecon.unr.edu.ar

Resumen

La regeneración a partir de cultivo *in vitro* de meristemas permite, entre otras aplicaciones, liberar a plantas de reproducción asexual de patógenos virales que eventualmente las han infectado. Una vez regeneradas, es necesario aclimatar estas plantas crecidas *in vitro* a las condiciones *ex vitro* siguiendo protocolos estándares en forma secuencial y condiciones controladas. Este proceso puede afectar su posterior desarrollo en condiciones de campo (Bairu et al., 2006).

En este trabajo se dispuso de 12 clones, algunos regenerados *in vitro* y otros donadores de meristemas, que crecieron inicialmente en dos ambientes distintos: Zavalla (Santa Fe) y Laguna Naineck (Formosa). En Zavalla se llevó a cabo la regeneración y aclimatación de las plantas, identificadas como Madres, que luego fueron llevadas a Laguna Naineck para su evaluación a campo. Una vez establecidas en este ambiente, se multiplicaron por rizomas, identificándose a las plantas así obtenidas como Hijos. Las variables cuantitativas que se analizaron fueron: diámetro de la planta, número de hojas, número de hojas en la primera etapa, número de hojas en la segunda etapa, número de hojas descartadas, largo de la hoja, ancho de la hoja y largo del peciolo. Por lo tanto, la matriz de datos resultó ser de tres modos (clones x variables x ambientes de crecimiento inicial).

El objetivo de este trabajo fue detectar diferencias o similitudes entre las plantas de banana según el ambiente de crecimiento inicial. Para tal fin, se llevó a cabo un Análisis Factorial Múltiple (AFM). Esta técnica no se limita a la obtención de una tipología de los individuos definida a través del conjunto de variables, sino que se

amplía en la búsqueda de posibles relaciones entre las estructuras obtenidas en el seno de cada uno de los “modos”. Es decir, intenta recoger la verdadera estructura presente en los datos y, así, generar conclusiones más completas que las obtenidas al realizar en forma aislada los análisis estadísticos multivariados tradicionales de dos modos (individuos x variables).

Los dos primeros ejes globales del AFM explicaron un 56,16% de la variabilidad total de los datos. El coeficiente RV fue igual a 0,36 indicando la presencia de diferencias entre las condiciones ambientales de crecimiento inicial. De la representación de las variables, se observó que la mayoría de ellas presentaban grandes discrepancias entre Madres e Hijos, distinguiéndose la variable número de hojas en la segunda etapa. A partir de la representación de las trayectorias de los individuos, también se percibió diferencias entre los ambientes Zavalla y Laguna Nainneck. Además, se distingue que los clones regenerados *in vitro* 12II y 522 se alejaban del resto con relación a los dos primeros ejes globales. En conclusión, el AFM permitió percibir las diferencias entre las plantas de banana según su ambiente de crecimiento inicial.

Referencias bibliográficas

Bairu, M.W.; Fennell, C. W.; Van Staden, J. (2006) The effect of plant growth regulators on somaclonal variation in Cavendish banana (Musa AAA cv. ‘Zelig’). *Scientia Horticulturae* 108(4): 347-351.

Pagès, J. (2004). Multiple Factor Analysis: Main Features and Application to Sensory Data. *Revista Colombiana de Estadística* 27 (1): 1-26.

Palabras Clave: Análisis Factorial Múltiple — Regeneración *in vitro* — plantas de reproducción asexual.

Área de Aplicación: Genética

Industria y Mejoramiento de la Calidad



ESTUDIO DEL EFECTO DE ERRORES DE MEDIDA SOBRE EL ERROR DE PREDICCIÓN DE UN ÁRBOL DE REGRESIÓN EN UN PROBLEMA REAL

Daniela F. Dianda, Marta B. Quaglino

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario.

ddianda@fcecon.unr.edu.ar

Resumen

Entre los objetivos principales del análisis de datos en contextos industriales, aparece la predicción, es decir, identificar una función que permita predecir el valor de una respuesta de interés a partir de los valores de otras variables consideradas potenciales predictores de esa respuesta. La posibilidad actual de disponer en poco tiempo de grandes cantidades de datos, ha hecho necesario desarrollar técnicas de análisis alternativas a las tradicionales, que permitan procesar y predecir la respuesta en tiempo real, de modo de poder accionar en forma dinámica sobre los procesos. Muchos de los nuevos métodos propuestos tuvieron su origen fuera del ámbito de la estadística y están basados en algoritmos automáticos desarrollados mayormente en el campo de la informática (Apté, 1997, Nisbet, Elder y Miner, 2009, Larose y Larose, 2015). No obstante, como sucede con cualquier técnica de análisis, la calidad de los datos es uno de los factores clave para asegurar la bondad de los resultados. Particularmente en el ámbito de procesos industriales, gran parte de los datos disponibles proviene de mediciones, por lo que cualquier pequeña distorsión en alguno de los sensores o dispositivos de medición, podría resultar en una gran cantidad de información “contaminada” en un período de tiempo muy corto. Diversos autores han demostrado que los errores de medición pueden distorsionar severamente los resultados de varias de las técnicas de análisis de uso frecuente en el campo de la mejora de procesos (Mittag, 1995, 1997, Bordignon y Scagliarini, 2002, Shishebori y Hamadami, 2009, Dianda, 2015). Frente a tales antecedentes, surge el interrogante respecto de si los algoritmos informáticos se ven afectados por los errores de medición, en cuanto a su habilidad para reconocer los patrones de relación entre las variables y proveer descripciones y/o predicciones confiables. Cuando se asume que en un proceso existen errores de medición, es necesario tener en cuenta que los valores observados no corresponden a la variable real, X , sino a una variable empírica, X^e (Mittag, 1997), afectada por tales errores. La presencia de errores sistemáticos genera un sesgo constante en las

mediciones, de modo que $X^e = X + c$, donde c es una constante real que representa la magnitud del sesgo del sistema de medición. En cambio, la presencia de una componente estocástica de error implica la existencia de una variable aleatoria V que la representa, resultando $X^e = X + V$. En este trabajo, se hace foco en los árboles de decisión CART (Classification and Regression Trees), uno de los algoritmos aplicables a problemas en los que el interés es predecir una respuesta de interés (Breinman, Friedman y Olshen, 1984). Se trata de un método no paramétrico de aprendizaje supervisado, cuyo objetivo es crear un modelo predictivo basado en reglas de decisión binarias inferidas de los datos. Mediante el algoritmo CART, se modelan datos reales de un proceso de producción de preparados para cemento de construcción, y su desempeño, en términos de capacidad predictiva, es analizado frente a la presencia de errores de medición tanto aleatorios como sistemáticos. Los resultados obtenidos son comparados con los obtenidos al emplear regresión lineal múltiple. La incorporación de errores de medición se realiza de manera artificial sobre los datos, simulando diferentes esquemas. Para el caso de errores aleatorios, se asume un modelo normal de media cero y variancia constante para cada variable $\sigma_{e_i}^2$. Los errores de tipo sistemáticos se incorporan adicionando a los valores reales de las variables una cantidad constante c_i . En ambos casos, el estudio considera diferentes valores de $\sigma_{e_i}^2$ y c_i , con el objetivo de evaluar también el efecto del cambio en la magnitud de los errores de medición. Entre los resultados obtenidos, se encontró que el error de predicción se ve siempre afectado por la presencia de errores de medida, aleatorios o sistemáticos, y aun cuando éstos sean de pequeña magnitud. No obstante, la presencia de errores sistemáticos se mostró mucho más nociva para la capacidad predictiva de los árboles que la de errores aleatorios, generando aumentos en el error de predicción de hasta un 60% aproximadamente. Estos resultados, aunque corresponden al estudio sobre un caso particular, resaltan la importancia de asegurar constantemente la calidad de la información recolectada, más aun en los contextos actuales, en los que la automatización permite generar y almacenar una gran cantidad de datos en períodos muy cortos de tiempo, y su análisis en tiempo real direcciona acciones correctivas de manera casi inmediata.

Referencias bibliográficas

Apté, C. (1997). Data Mining: An Industrial Research Perspective. *IEEE Computational Science & Engineering*, 4(2), 6-9.

- Bordignon, S. y Scagliarini, M. (2002). Statistical Analysis of Process Capability Indices with Measurement Errors. *Quality and Reliability Engineering International*, 18, 321-332.
- Breiman, L., Friedman, J., Olshen, R. y Stone, C. (1984). *Classification and Regression Trees*. Florida, USA, Chapman & Hall/CRC Press.
- Dianda, D. (2015). Estudio estadístico de sistemas de medida e indicadores de capacidad de procesos multivariados, en contextos de mejora de la calidad y la productividad. Tesis Doctoral. Universidad Nacional de Rosario.
- Larose, D. y Larose, C. (2015). *Data mining and predictive analytics*. New Jersey, USA, John Wiley & Sons, Inc.
- Mittag, H-J.(1995). Measurement Error Effect on Control Chart Performance. *Annual Quality Congress*, 49(0), 66-73.
- Mittag, H-J. (1997). Measurement Error Effects on the Performance of Process Capability Indices. *Frontiers in Statistical Quality Control*, 5, 195-206.
- Nisbet, R., Elder, J. y Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Amsterdam, Países Bajos. Elsevier Inc.
- Shishebori, D. y Hamadami, A.Z. (2009). The Effect of Gauge Measurement Capability and Dependency Measure of Process Variables on the MCp. *Journal of Industrial and Systems Engineering*, 4(1), 59-76.

Palabras Clave: CART - Regresión lineal - Error de medición - Error de predicción.

Área de Aplicación: Industria y mejoramiento de la calidad.

NUEVAS ESTRATEGIAS PARA LA CONSTRUCCION DE DISEÑOS 2^k-p PLEGADOS

Lucia Hernandez, Antonella Mazzeo , Jose Pagura

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario.

lucia.hernandez@fcecon.unr.edu.ar

Resumen

Los experimentos son un recurso esencial en muchas actividades relacionadas a la investigación y el desarrollo para caracterizar procesos o sistemas. Los diseños factoriales son la mejor estrategia experimental para estudiar los efectos de múltiples factores sobre una característica o variable respuesta de interés. La eficiencia de los

diseños factoriales aumenta, a medida que el número de factores considerados aumenta. Sin embargo, también aumenta el número de pruebas que deben ser realizadas. Si el número de factores a considerar es grande, una manera de reducir el número de pruebas es ensayar cada factor a dos niveles resultando en los llamados diseños 2^k . Aun cuando los factores sean ensayados a sólo dos niveles los diseños factoriales pueden requerir demasiadas pruebas. En estos casos, la mejor alternativa que se encuentra en la bibliografía es llevar a cabo una fracción de las pruebas requeridas por un diseño factorial completo. Un diseño 2^{k-p} , es un diseño para k factores cada uno medido a dos niveles que utiliza 2^{k-p} tratamientos. Al ensayar un número menor de pruebas que un diseño 2^k , se pierde, además de precisión en las estimaciones de los efectos, se pierde la capacidad de estimar algunos efectos y la de distinguir o separar las estimaciones de grupos de los demás efectos dándole lugar a la denominada estructura de confusión de alias. Los diseños 2^{k-p} pueden ser clasificados de acuerdo con su resolución que indica qué tipos de efectos están confundidos entre sí. A menor resolución, menor será la capacidad del diseño para distinguir las estimaciones de los efectos. Los diseños 2^{k-p} de resolución baja son en general utilizados como experimentos de screening, es decir experimentos cuyo objetivo es determinar cuáles factores de un gran grupo tiene efecto sobre la característica de interés. Esto ocurre usualmente en las primeras etapas de la investigación donde no se prioriza la información acerca de las interacciones, aunque tampoco se descarta si puede obtenerse (Mee (2009), Wu (2009), Oehlert (2010), Lawson (2015)).

En muchos casos, la confusión no afecta el análisis ya que los resultados obtenidos no generan ambigüedad alguna. Sin embargo, otras veces la confusión es crucial obteniendo resultados que dejan ciertas cuestiones sin resolver. En estos casos se necesitará datos adicionales o información externa de que ciertas interacciones son poco probables de ocurrir para poder concluir el análisis. Si se necesitan datos adicionales para poder resolver ambigüedades acerca de ciertos efectos, una estrategia comúnmente utilizada es llevar a cabo un nuevo experimento de acuerdo con otra fracción del diseño 2^{k-p} inicial. La técnica conocida como “foldover” o diseño plegado consiste en realizar la nueva fracción cambiando los niveles de todos los factores o de algunos de los factores por aquellos niveles no utilizados en la fracción inicial (o dicho más simplemente, invirtiendo los niveles de los factores). Cabe señalar que un foldover puede ser construido por dos motivos. Si el análisis del diseño inicial revela que un conjunto particular de efectos es significativo, entonces el foldover puede ser elegido

para resolver la confusión entre estos efectos en particular. Por ejemplo, si un factor dado es muy importante y no debe ser confundido con otros efectos, entonces el foldover consistente en invertir los niveles de ese factor es el apropiado. Ahora bien, si el objetivo es separar tantos efectos como sea posible, entonces la elección del foldover no es tan inmediata. La opción tradicionalmente utilizada consiste en ensayar la nueva fracción invirtiendo los niveles de todos los factores involucrados en el experimento. Esta estrategia es buena dado que el diseño combinado, es decir el diseño consistente de la fracción inicial y de la nueva fracción tendrá mayor resolución. Sin embargo, en los últimos años algunos autores han mostrado que se pueden obtener diseños con mayor resolución o con la misma resolución, pero con mayor capacidad para separar efectos que el resultante del foldover tradicional (Montgomery y Runger (1996), Li y Mee (2002), Li y Lin (2003)).

En este trabajo se realizó una reseña de las nuevas propuestas existentes en la bibliografía para construir foldovers comparándolas con los métodos tradicionalmente utilizados para mostrar las ventajas de cada una de ellas. Los resultados muestran que las nuevas estrategias conducen a diseños combinados (diseño inicial y foldover) con mayor capacidad para separar grupos de efectos confundidos entre sí que las estrategias tradicionales.

Referencias bibliográficas

- Lawson, John (2015). "Design and Analysis of Experiments with R". CRC Press
- Li, Huoy Mee, Robert W. (2002) "Better Foldover Fractions for Resolution III 2^{k-p} Designs". Technometrics.
- Li, William y Lin Dennis K. J. (2003). "Optimal Foldover Plans for Two-Level Fractional Factorial Designs". Technometrics.
- Mee, Robert W. (2009) "A Comprehensive Guide to Factorial Two-Level Experimentation". Springer.
- Montgomery, Douglas C. y Runger, George C. (1996). "Foldovers of 2^{k-p} Resolution IV Experimental Designs". Journal of Quality Technology.
- Oehlert, Gary W. (2010). "A First Course in Design and Analysis of Experiments". Gary W. Oehlert.
- Wu, C. F. Jeff y Hamada Michael S. (2009). "Experiments. Planning, Analysis and Optimization". John Wiley & Sons, Inc.

Palabras Clave: Experimentos de Screening – Experimentación Secuencial – Resolución – Mínima Aberración

Área de Aplicación: Industria y Mejoramiento de la Calidad.

**IMPACTO DE DISTINTOS MÉTODOS DE ESTIMACIÓN DE SCORES
FRENTE A DATOS FALTANTES SOBRE EL GRÁFICO T² DE HOTELLING
CON COMPONENTES PRINCIPALES**

Julia Inés Fernández, José Alberto Pagura, Marta Beatriz Quaglino

*Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad
de Cs. Económicas y Estadística, Universidad Nacional de Rosario - CONICET*

juliai.fernandezt@gmail.com

Resumen

Problemática/antecedentes:

El control estadístico de procesos es una importante estrategia para el mantenimiento y mejora de calidad en los ámbitos industrial y de servicio. Su objetivo es monitorear si el estado de un proceso a lo largo del tiempo se mantiene bajo control estadístico, en el cual las variaciones de las variables de calidad estudiadas se deben a causas comunes de variación, y detectar si se producen causas asignables de variación para corregirlas. Actualmente, la calidad de un producto depende de la combinación de valores de un conjunto de variables, lo cual hace esencial implementar estrategias de control multivariado de procesos (MSPC). Una propuesta es usar Análisis de Componentes Principales (PCA) y basar el control en un número menor de variables que retenga la información relevante, construyendo a partir de ellas el gráfico T² de Hotelling.

En el contexto de MSPC, es común encontrar datos faltantes debido a diferentes causas. Estos pueden afectar al MSPC en la etapa de construcción del modelo PCA, o de aplicación del control ante nuevas observaciones en un modelo ya construido. Para la segunda etapa, Arteaga y Ferrer (2002) proponen y analizan un conjunto de métodos para estimar scores de nuevas observaciones cuando hay datos faltantes, utilizando un modelo PCA conocido. En el presente trabajo se estudia por simulación el impacto que tiene la estimación de los scores sobre el gráfico T² de Hotelling diseñado a partir de componentes principales de un modelo PCA construido previamente sin datos faltantes.

Objetivo: Evaluar la eficiencia de estrategias de MSPC utilizando PCA frente a datos faltantes en caso de que el estado de control se mantenga, usando distintos métodos para estimar los scores de nuevas observaciones.

Metodología:

Se realiza un estudio por simulación usando el software libre R para evaluar el cambio que se produce en el comportamiento del gráfico multivariado bajo control estadístico, en presencia de valores faltantes en los vectores de información de calidad. A partir de dos conjuntos de datos reales, uno tratado en la literatura (Folch-Fortuny, Arteaga y Ferrer 2015) y otro del repositorio UCI Machine Learning Repository, se estiman modelos PCA con los cuales se diseña una estrategia T^2 de Hotelling considerando los datos como poblacionales. Se retienen la cantidad de componentes principales que explican al menos el 70% de la variación total de los datos. Para evaluar si la performance del gráfico se mantiene frente a la presencia de datos faltantes, se generan observaciones con pérdidas aleatorias de las poblaciones definidas y se aplica el control calculando los scores mediante cuatro técnicas (Arteaga y Ferrer ,2002), que no requieren de imputación de los valores faltantes: Trimmed Score (TRI), Known Data Regression (KDR), Trimmed Score Regression (TSR) y Projection to the Model Plane (PMP). Los escenarios construidos combinan tres porcentajes de datos faltantes: 5%, 10% y 15%, y cuatro métodos de estimación de los scores: TRI, KDR, TSR y PMP. En cada una de las situaciones estudiadas, se generan 10.000 observaciones normales con vector de medias $\mathbf{0}$ y con estructuras de covariancia definidas por las matrices de correlación de los conjuntos de datos considerados. Como límite de control se toman los valores de la estadística χ^2 con grados de libertad igual al número de componentes principales utilizadas y $\alpha=0.005$. Como medidas de eficiencia se evalúan los ARL (Average Run Length) bajo control en cada escenario, resumiendo el comportamiento a través del Error Cuadrático Medio (ECM). El gráfico fue diseñado de manera que, con datos completos, el valor del ARL sea de 200. También se comprueba si la presencia de datos faltantes altera la distribución de la estadística del gráfico de control T^2 .

Resultados:

En la mayoría de las situaciones estudiadas se encontró que la distribución del estadístico de prueba para el gráfico T^2 no sigue la distribución chi-cuadrado. En casi todos los escenarios, el valor del ARL estimado fue superior al ARL bajo control

esperado de 200 para $\alpha=0.005$. Se encontró que, en general, el método KDR presenta los menores ECM, lo cual coincide con lo presentado por Arteaga y Ferrer (2002).

Referencias bibliográficas

- Arteaga, F. y Ferrer, A. (2002). Dealing with missing data in MSPC: several methods, different interpretations, some examples. *Journal of Chemometrics*, 16, 408-418.
- Dua, D. y Karra Taniskidou, E. (2017). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- Folch-Fortuny, A., Arteaga, F. y Ferrer, A. (2015). PCA model building with missing data: new proposals and a comparative study. *Chemometrics and Intelligent Laboratory Systems*, 146, 77-88.
- MacGregor, J. y Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403-414.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>.
- Wold, S., Geladi, P., Esbensen, K. y Ohman, J. (1987). Multi-way principal components and PLS analysis. *Journal of Chemometrics*, 1, 71-84.

Palabras Clave: Control Multivariado de Procesos - datos faltantes - Análisis de Componentes Principales - gráfico T^2 de Hotelling

Área de Aplicación: *Industria y mejoramiento de calidad*

Salud Humana

A horizontal orange arrow pointing to the right, with the text "Salud Humana" written in blue inside it.

COMPARACIÓN DE MÚLTIPLES TRATAMIENTOS A PARTIR DE UNA REVISIÓN SISTEMÁTICA

Guillermina Isern, Cristina Cuesta

Instituto de Investigaciones Teóricas y Aplicadas en Estadística de la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario, Argentina

gisern@fcecon.unr.edu.ar

Resumen

Los estudios por metanálisis en medicina constituyen un método analítico que permite la compaginación de resultados individuales de artículos científicos. La selección de los artículos participantes del mismo debe llevarse a cabo de forma programada y exhaustiva a través de una revisión sistemática claramente diseñada. Las metodologías más desarrolladas para los metanálisis se dedican a la comparación de dos estrategias para el tratamiento de una enfermedad o evento clínico.

En el último tiempo se han comenzado a desarrollar metodologías para contrastar tratamientos que no fueron comparados en forma directa pero que, sin embargo, dicha comparación puede obtenerse en forma indirecta. Por ejemplo, si se quiere comparar el tratamiento B con el C, pero sólo se cuenta con la comparación de A con B y de A con C, la comparación de B con C se podría obtener en forma indirecta. Esta problemática se conoce como “Metanálisis en Red” y permite comparar más de dos tratamientos con el conocimiento de sólo algunas comparaciones entre ellos.

Las estrategias para llevar a cabo los metanálisis en red van desde herramientas graficas hasta estimaciones basadas en modelos. Los gráficos más comunes son: el gráfico de la red, el diagrama de calor (para representar la contribución de cada estudio al efecto tratamiento y su consistencia) y el gráfico de Bland-Altman (para visualizar la heterogeneidad entre los estudios). Los modelos para estimar las comparaciones de tratamientos se estiman por mínimos cuadrados ponderados (teniendo en cuenta como ponderación la variabilidad observada en cada comparación). Estos permiten realizar estimaciones de los efectos de los tratamientos y a partir de ellas estimar las comparaciones entre tratamientos tanto en forma directa como indirecta. Por último, se utilizan las estadísticas Q_{total} e I^2 para evaluar la heterogeneidad entre estudios dentro de la red.

La validez de estos análisis se sustenta en el cumplimiento de algunos supuestos. En principio debe verificarse la calidad de los estudios incluidos, el control de sesgos, la

similitud metodológica y la variabilidad entre estudios como ocurre con cualquier metanálisis. Por otro lado debe haber consistencia entre las evidencias procedentes de comparaciones directas y los resultados de las estimadas en forma indirecta. Finalmente, el supuesto más fuerte que debe sostenerse es el de transitividad, esto significa que si el tratamiento B es mejor que el tratamiento A y A es mejor que C, entonces se asume que A es mejor que C.

En este trabajo se realiza un metanálisis en red para comparar tratamientos disponibles de drogas uteroretroductoras para controlar la hemorragia durante el tercer período del parto (lapso comprendido entre la salida del bebé y la expulsión de la placenta). La atonía uterina es la causa más común de hemorragia posparto primaria. Los procedimientos de tratamiento incluyen, entre otros, el uso de medicamentos oxitócicos tales como: oxitocina, ergonovina/metilergonovina, carbetocina y misoprostol.

Referencias bibliográficas

- Bland, J. A. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327(8476), 307–310.
- Hu, Y. (2012). Algorithms for visualizing large networks. En O. S. U. Naumann, *Combinatorial Scientific Computing* (págs. 525–549). Chapman and Hall/CRC Computational Science.
- Krahn, U. B. (2013). A graphical tool for locating inconsistency in network meta-analyses. *BMC Medical Research Methodology*, 13(1), 35 .
- Krahn, U. B. (2014). Visualizing inconsistency in network meta-analysis by independent path decomposition. *BMC Medical Research Methodology*, 14(1), 131.
- Rücker, G. (2012). Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods* 3, 312–324.
- Rücker, G. S. (2014). netmeta: network meta-analysis with R. www.cran.R-project.org/package=netmeta. R package version 0.6-0.
- Rücker, G. S. (2015). Automated drawing of network plots in network meta-analysis. *Research Synthesis Methods*.
- Schwarzer, G. C. (2015). *Meta-Analysis with R (Use R!)*. Springer International Publishing.

Senn, S. G. (2013). Issues in performing a network meta-analysis. *Statistical Methods in Medical Research* 22, 169–89.

Veroniki, A. V. (2013). Evaluation of inconsistency in networks of interventions. *International Journal of Epidemiology* 42, 332–345.

Palabras Clave: Metanálisis en red – Hemorragia post-parto – Efecto tratamiento

Área de Aplicación: Salud humana.

DIAGNÓSTICO DEL PERFIL SOCIOECONÓMICO Y NUTRICIONAL DEL PACIENTE DIABÉTICO TIPO 2 A PARTIR DE MÉTODOS MULTIVARIANTES.

Santillán M.¹, S. M. Zabala¹, Martínez C.1, Doello, F.² y Olga S.Filippini¹

¹ Departamento de Ciencias Básicas, Área Estadística, Universidad Nacional de Lujan,
Lujan Buenos Aires, Argentina

² Servicio de Diabetología, Hospital de Moreno, Moreno, Buenos Aires, Argentina
marisolsantillan_unlu@yahoo.com.ar

Resumen

La diabetes es un síndrome caracterizado por una hiperglucemia que se debe a un deterioro absoluto o relativo de la secreción y / o la acción de la insulina. Según la ley 26.914 de Salud Pública es necesario un conocimiento tanto de la enfermedad como de sus complicaciones de acuerdo con los conocimientos científicos aceptados hasta este momento. Esto hace referencia a que si el paciente conoce sobre su enfermedad obtiene más herramientas para su control y disminuye con esto la posibilidad de contraer complicaciones. Asimismo, la cantidad de diabéticos se cuadruplicó desde 1980: hoy viven con el mal 422 millones de personas en el mundo. En nuestro país, uno de cada 10 adultos vive con diabetes. Es decir que casi tres millones de mayores de 18 años tienen diabetes tipo 2, asociada principalmente con el sedentarismo, la mala alimentación y el tabaquismo. (Ministerios de Salud, 2016). La diabetes tipo 2 es la forma más común: afecta al 90-95% de los casos. El resto tiene diabetes tipo 1, que suele diagnosticarse en los niños y los jóvenes (también se la conoce como insulino dependiente). A diferencia de la diabetes tipo 2, la OMS recuerda que aún se desconoce qué impide que el cuerpo produzca suficiente insulina. La forma gestacional, que aparece durante el embarazo, no sólo aumenta el riesgo de la mujer de desarrollar diabetes tipo 2, sino que también puede causarle complicaciones en la gestación, el parto o su bebe. La OMS insiste en la importancia de los controles prenatales.

Atendiendo estos antecedentes el objetivo de este trabajo es obtener información a partir de pacientes diabéticos tipo 2, que permita elaborar un manual que contemple programas para mejorar hábitos y consumos alimentarios saludables. En base a lo establecido por el ministerio de Salud y la última encuesta de Factores de Riesgos Cardiovasculares (2014), se trabaja en la elaboración de mensajes prácticos sencillos que contemplen factores genéticos, sociales, económicos y culturales que mejoren los hábitos de consumo de alimentos y de vida saludable. Para la recolección de la información se realiza un Grupo Focalizado en el Hospital Mariano y Luciano de la Vega, Moreno, Buenos Aires, donde se relevan las inquietudes de los pacientes con respecto a la dificultad de mantener una alimentación saludable. Por otra parte, para caracterizarlos, se utiliza un test (test de Findrisk) cuyo procesamiento de resultados posibilitara reconocer complicaciones y calcular el índice de masa corporal. Se utiliza un estudio Analítico Observacional retrospectivo, y se procesa la información con diferentes métodos multivariados. Los resultados obtenidos permiten caracterizar a los individuos según salud, economía y aspectos sociales asociados a esta enfermedad, mejorando el diagnóstico del estado nutricional de los pacientes y orientando eficientemente las recomendaciones para el programa propuesto.

Referencias bibliográficas

- Carles M. Cuadras. (*CMC Editions, Barcelona, España*). Nuevos métodos de análisis multivariantes.
- Benzadon M, Forti L, Sinay I. (B. Aires; 1 feb. 2014 [citado 2015 Aug 04]) Actualización en el diagnóstico de la diabetes. URL: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S002576802014000100016&lng=es&nrm=iso
- Norelis Paredes y otros. (Med Interna, Caracas. Vol. 30, N°1 año 2014). Aplicación del test Findrisk para cálculo del riesgo de padecer diabetes mellitus tipo 2. URL: <http://www.svmi.web.ve/ojs/index.php/medint/article/view/87/85>
- Ministerio de Salud. Año 2016. Guía de Práctica Clínica Nacional sobre Prevención, Diagnóstico y tratamiento de la diabetes Tipo 2. Para el primer nivel de atención. URL: http://www.msal.gob.ar/images/stories/bes/graficos/0000000076cnt-2012-08-02_guia-breve%20-prevencion-diagnostico-tratamiento-diabetes-mellitus-tipo-2.pdf

Palabras Clave: Diabetes - Focus Group - Medicina

Área de Aplicación: Salud Humana

Ciencias Agropecuarias



MONITOREO DE HUERTAS Y GRANJAS DEL PARTIDO DE LUJÁN

Edgardo Sampaolesi; Silvia Lorena Vargas; Facundo Parmiggiani; María Laura Vignera; Stella Zabala; Luciano Palacios; Viviana Escanes; Carla Martínez; O.;

Susana Filippini

*Departamento de Ciencias Básicas, Área Estadística, Universidad Nacional de Lujan,
Lujan Buenos Aires, Argentina*

Resumen

La agencia de extensión rural de INTA del partido de Luján, Provincia de Buenos Aires a través del programa Pro-huerta brinda semillas y animales de granjas a la comunidad. Los objetivos de las acciones de INTA son las de mejorar la situación alimentaria de la población por debajo de la línea de pobreza a través de la autoproducción de alimentos frescos en huertas y granjas a nivel familiar, escolar, comunitario e institucional. Se realizó una encuesta semi estructurada siguiendo la línea de cuestionarios de los últimos censos agropecuarios y las necesidades de información del INTA. El instrumento fue probado mediante la realización de encuesta piloto. Los encuestadores fueron capacitados mediante talleres y se les suministró un manual con el lenguaje técnico utilizado en la actividad hortícola, y en las labores de granjas. Se realizó la entrevista a 183 huerteros y/o granjeros evaluándose la eficacia lograda en el armado y producción de huertas y granjas, a partir de los kits de semillas y animales de granjas entregados en los períodos Otoño Invierno, Primavera Verano, 2016-2017.

A efectos de evaluar los resultados del programa, se decide trabajar sobre el partido de Luján, de acuerdo a la información suministrada por el INTA y a las planillas donde estaban enlistados los huerteros y/o granjeros. Para hacer la evaluación se decide realizar un diseño de muestra a efectos de poder, en un tiempo relativamente corto, evaluar el impacto del programa, siendo el muestreo Simple al Azar con una selección Sistemática de la Muestra; en una población conocida de 720 casos. Se analizaron variables de tipo cuantitativa y cualitativa desde la identificación del encuestado, aspectos productivos y socio- demográficos que fueron procesadas con el software Estadístico SPSS 14.

Los resultados de la encuesta muestran que la mayoría de los encuestados pertenecen al partido y realizan huerta, granja en menor porcentaje y ambas actividades en una condición intermedia. El registro de kits y cantidad de huertas presentan una correspondencia alta, ya que un alto porcentaje de los encuestados que recibieron kits hicieron al menos una huerta. La producción de huerta y de granja es mayormente

destinada al consumo. Las especies más sembradas fueron acelga, espinaca, lechuga, perejil, albahaca, tomate y morrón. Los animales de granja recibidos fueron gallinas ponedoras, pollos parrilleros y conejos. Mayoritariamente mencionan que el principal destino de la producción es para consumo, en menor porcentaje responde que además utiliza la producción para la venta con el objeto de comprar alimentos para abastecer a la granja.

Un elevado número de encuestados no recibió capacitación en la localidad de Luján y manifiestan interés en recibir capacitación para implementar la huerta y producir y disminuir la mortalidad de los animales.

Referencias bibliográficas

Cuadras C; (2014). Nuevos Métodos de Análisis Multivariante. Barcelona: CMC Editores:

García R. M. (2004). Inferencia Estadística y Diseño de Experimentos. Buenos Aires: Eudeba.

Lohr S. L. (2000). Muestreo: Diseño y Análisis. Mexico: Editorial S.A Ediciones Paraninfo.

Marradi A., Archenti N., Piovani J. I. (2007) Metodología de las ciencias sociales. Buenos Aires: Emece Editores.

IBM SPSS Statistical software, versión para Windows, 2014.

Palabras Clave: Monitoreo- Muestreo- Encuesta- Prohuerta

Área de Aplicación: Ciencias Agropecuarias

ANÁLISIS DEL CONTENIDO VOLUMÉTRICO DE AGUA EN EL SUELO A PARTIR DE MÉTODOS DESCRIPTIVOS Y EXPLICATIVOS MULTIVARIANTES

Stella M. Zabala^{1,2}, Olga S. Filippini¹, Diana A. Giorgini¹ y Héctor A. Salgado²

¹Departamento de Ciencias Básicas, Área Estadística, Universidad Nacional de Luján, Luján Buenos Aires, Argentina

²Riego y Drenaje, Facultad de Agronomía de Buenos Aires, CABA, Argentina

stezab@hotmail.com

Resumen

El contenido de agua en el suelo, su relación con el estado fenológico vegetal y las variables meteorológicas son fundamentales para el diagnóstico hídrico de los cultivos y la administración del riego. En la actualidad, monitorear el estado hídrico es

aplicable a cultivos bajo riego deficitario, manejando el Umbral de Riego, y logrando mayor eficiencia en el uso del agua y mejoras en las cualidades de cultivos hortícolas y viñas (Fereres & Soriano, 2007). Su medición ha sido un desafío tecnológico importante en las últimas décadas, dando origen a diversos tipos de instrumentos, métodos y técnicas. Se desarrollaron sensores, que permiten la medición del agua en el suelo sin alterar significativamente las condiciones hidrodinámicas del mismo, y el registro de variables climáticas de modo más preciso. El objetivo del presente trabajo es analizar la incidencia de diferentes variables climáticas y del suelo sobre el contenido volumétrico de agua (CVA) disponible para la vegetación para su predicción.

La Estación Automática de Medición de humedad del suelo y microclima está ubicada en una zona agrícola del Partido de Azul, Provincia de Buenos Aires, (Latitud 37°00'S; Longitud 59°30'W), Argentina, donde los suelos predominantes son Argiudoles, con textura franca arcillo-limosa, alto % de materia orgánica y buen drenaje. La determinación del CVA se efectúa mediante sondas dieléctricas que emplean la velocidad de propagación de un pulso electromagnético de microondas para calcular la constante dieléctrica (ϵ) del suelo. Además del CVA ($m^3.m^{-3}$), los equipos registran conductividad eléctrica CE ($S.m^{-1}$) temperatura del suelo y del aire T ($^{\circ}C$), velocidad (m/s) y dirección del viento, radiación global (W/m^2), humedad relativa (%) y precipitación (mm/d).

El agua disponible para la vegetación es calculada considerando el contenido hídrico por encima del Punto de Marchitez, para la correspondiente clase textural del suelo, de acuerdo a una ecuación adaptada de Saxton & Rawls (2006).

Los datos recopilados fueron analizados usando el programa estadístico SPSS (2014) e Info Stat (2008). El análisis se efectuó en dos etapas. En la primera, se aplicó el método de componentes principales (CP) y Análisis de Factores (FA), con rotación Varimax con normalización Kaiser y las puntuaciones según método de Barlett para reducir la dimensionalidad de la base de datos. Posteriormente, se aplicó un modelo de regresión lineal múltiple considerando a los factores obtenidos anteriormente como variables predictoras para analizar las relaciones y variabilidad del sistema estudiado.

Del análisis de componentes principales se evidencian tres factores claramente diferenciados: contenido volumétrico de agua en suelos, radiación y precipitación corroborados numérica y gráficamente.

Posteriormente se realizó un análisis de regresión lineal múltiple (pasos hacia adelante) considerando al CVA como variable dependiente debido a que, según la bibliografía, es un indicador sensible en el balance hídrico en suelos regados complementariamente.

En este modelo los factores Ce del agua en el suelo, radiación y precipitación resultaron significativos, siendo el coeficiente estandarizado de la Ce del suelo el más significativo para explicar la variabilidad del CVA.

Los resultados obtenidos demuestran que, la reducción de la dimensionalidad de la matriz de información obtenida ayudó a encontrar una función lineal general que permitirá predecir el contenido volumétrico del agua en suelos en la región pampeana húmeda de manera sencilla.

Referencias bibliográficas

Cuadras Carlos, (2014) Nuevos Métodos de Análisis Multivariante. CMC editores Manacor 30. Barcelona, España.

Garcia Roberto M. (2004). Inferencia Estadística y Diseño de Experimentos. Primera Edición. Eudeba.

Fereres E., y M.A. Soriano, 2007. Deficit irrigation for reducing agricultural water use. *Journal of Experimental Botany*, Vol.58, 2:147-159.

InfoStat, 2008. Statistical software. Group Info Stat. Ability of Agricultural Sciences, National University of Córdoba, Argentina. Versión 2008.

Salgado H., C. Romay, S.M. Zabala y L. Génova, 2012. Calibración de campo de un sensor dieléctrico para suelos en la Provincia de Buenos Aires. VI Jornadas de Riego y Fertirriego, Mendoza, Argentina, 7-9 nov. 2012, 8 pp. Disponible en Internet: http://www.riegoyfertirriego.com.ar/VI_Jornadas/VI_JARF_Trabajos_Completos/Salgado.pdf.

Saxton K. E. and W. J. Rawls, 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Sci. Soc. Am. J.*, 70:1569-1578.

IBM SPSS Statistical software, versión para Windows, 2014.

Palabras Clave: monitoreo de humedad del suelo- variables climáticas- sondas dieléctricas- estación meteorológica automática - métodos multivariados.

Área de Aplicación: Ciencias Agropecuarias

ESTRATEGIAS DE ANÁLISIS PARA DATOS CATEGÓRICOS DE NATURALEZA BINARIA

**O. Susana Filippini ^{1,2, 2}; Vilma Regules ²; Gabriela H. Gualdoni, ²; Jessica
Gamarra ²; Sofía Huarita², Iván Depiante ²; Vanina Delfino ²**

^{1,2}Dpto. de Ciencias Básicas, Estadística, Universidad Nacional de Luján. Bs.As. Arg.²

Dpto. de Matemática y Estadística, Universidad Nacional de Tres de Febrero, Bs. As,

Argentina

sfilippini@unlu.edu.ar

Resumen

Las pérdidas que causan las enfermedades de semillas en la Agricultura, se presentan de varias formas. Pueden ser inmediatas y de períodos cortos por una germinación reducida, bajo vigor de las plántulas, o plántulas anormales y otros daños al cultivo en cualquier etapa del crecimiento, desde que son plántulas hasta la cosecha y almacenamiento, lo que afecta los rendimientos económicos. En esta oportunidad se consideró el caso de un estudio experimental comparativo, aleatorizado, para evaluar el número de plantas anormales, en pruebas de vigor en semillas de Soja, dónde se consideraron como factores, además de las variedades (dos), los distintos tratamientos de estrés que se les realizó a las semillas en los tres niveles de exposición: EA 36°C, EA 41°C y el Cold test. Las respuestas analizadas fueron poder germinativo, porcentaje de plántulas anormales y otras variables para las pruebas de vigor. En este caso se analizan datos de porcentaje de plántulas anormales para dos variedades. Vemos entonces que lo que habitualmente llamamos proporción tiene una interpretación como promedio aritmético de variables binarias (las plántulas del ensayo para cada variedad y unidad experimental ó parcela, serán normales ó anormales, generando una variable respuesta que será una proporción ó porcentaje. Esto nos permite en principio usar todo el valioso arsenal de conocimientos sobre estadísticas de promedios a casos de proporciones y porcentajes. Por extensión, los análisis que indirectamente involucran proporciones, por ejemplo los que modelizan probabilidades, por ejemplo los logísticos (Hosmer, J. R., et al,1989) también admiten en principio una representación estadística como promedios y podrían ser también estudiados a partir de un Modelo Lineal Generalizado, Cayuela, L.(2009). El tema de las variables binarias es un clásico de la estadística contemporánea y ha sido estudiado por numerosos autores, en sus aspectos teóricos y prácticos, incluyendo contribuciones de los fundadores de la biometría

moderna, Leroy Folks, J, (1999). Sin embargo, el problema de las interconexiones y desprendimientos estadísticos de la binarización ofrece numerosas variantes que siguen siendo de interés estadístico. Por eso, resulta beneficioso tener como objetivo analizar los resultados obtenidos con datos de naturaleza binaria, a partir de diferentes alternativas con estrategias de análisis para datos categóricos dicotómicos. El llamado conteo de atributos es una suma de variables binarias. La tabla de contingencia asociada a la prueba de χ^2 para la significatividad de la diferencia entre dos proporciones binomiales, representa un conteo de variables binarias desglosado según dos atributos, por ejemplo, tratamientos, y plántulas (normales ó anormales). Se utilizó la prueba χ^2 con partición de tablas y comparación de grupos con tablas dos por dos, con corrección de Bonferroni para las comparaciones y regresión logística con una modelización explícita de la variable dependiente binaria en función de las variables independientes de interés. Se han propuesto muchas funciones de distribución para ser usadas en el análisis de una variable respuesta dicotómica, pero existen razones principales para elegir una función Logística, Hilbe, J. M (2009). Los métodos alternativos considerados, producen resultados similares. Según pudo observarse, los menores porcentajes de germinación de plántulas anormales se dieron tanto en la variedad V1 (A4201) como en la V2 (A5901) en el tratamiento de EA 36°C, no observándose diferencias entre las dos variedades. En el test de EA 41 °C se observaron diferencias entre la V1 y la V2.

Referencias bibliográficas

Cayuela, Luis. (2009) Modelos lineales generalizados (GLM).

Hilbe, Joseph, M.(2009). Logistic Regression models. Chapman and Hall Tests in Statistical Science. Taylor and Francis Group.

Hosmer JR., D. W., And Lemeshow, S. (1989) Applied logistic regression. Wiley, New York.

Leroy Folks, J. 1999. Ideas of statistics. Wiley, pp. 319-324.

IBM SPSS Statistical software, versión para Windows (2014).

Palabras Clave: Datos Categóricos- Comparaciones entre grupos. Bonferroni- Modelación logística

Área de Aplicación: Ciencias Agropecuarias.

ESTIMACIÓN ROBUSTA DE UN MODELO NO LINEAL MIXTO PARA EVALUAR EL CRECIMIENTO DEL MAIZ

María del Carmen García, Cecilia Rapelli

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario. CIC-UNR
mgarcia@fcecon.unr.edu.ar

Resumen

El desarrollo de modelos flexibles para el análisis de datos longitudinales se ha convertido en un problema importante en muchas ramas de la ciencia. Los modelos no lineales mixtos se usan, frecuentemente, para caracterizar longitudinalmente procesos biológicos. Para representar la variación poblacional, estos modelos expresan los parámetros específicos de las unidades en función de efectos fijos y aleatorios y para considerar la correlación entre las mediciones repetidas introducen errores intra unidad.

En la modelación longitudinal es probable que existan unidades aberrantes (algunas unidades pueden tener una magnitud inusual). La estimación de los parámetros de los modelos mixtos se realiza mediante el uso de métodos basados en la función de verosimilitud, que son sensibles a estas unidades atípicas. Los analistas deben ser cuidadosos ante la presencia de estos datos, pues pueden tener una importante influencia sobre los resultados del análisis. Varias metodologías para la estimación de los parámetros de los modelos no lineales mixtos han introducido procedimientos de estimación robustos. En este trabajo se utiliza un procedimiento robusto a los outliers, que utiliza el método linealización de Lindstrom y Bates (1990), para estimar los efectos fijos y componentes de variancia. El procedimiento reemplaza la suma de residuos al cuadrado, en la función de verosimilitud, por la función Bisquare de esos residuos, que es robusta a los outliers (Williams et al.,2015). Para la aplicación se utilizan datos de un ensayo llevado a cabo en la Facultad de Ciencias Agrarias de la UNR para evaluar el rendimiento de un híbrido de maíz, sembrado en parcelas tratadas con dos niveles de fertilización nitrogenada y dos niveles de azufre. Con el fin de medir la evolución del peso y humedad del grano se realizaron muestreos periódicos, hasta que el peso permaneciera constante, extrayendo en cada ocasión 50 granos que fueron secados para obtener el peso seco. Se contaron en total con 13 mediciones para cada unidad. Se realiza un análisis exploratorio detectándose la presencia de algunas observaciones

atípicas. Se ajustó un modelo no lineal mixto utilizando la curva de Gompertz, asumiendo que de sus tres parámetros (límite de crecimiento, parámetro de escala que indica el inicio de crecimiento de la curva y tasa de crecimiento) sólo el inicio de crecimiento de la curva varía estocásticamente entre las unidades. La estimación de ese modelo se realizó utilizando un método estándar y uno robusto. El análisis de los resultados obtenidos permite concluir que la estimación de los efectos fijos cambia levemente cuando se utiliza el método de estimación robusto. Sin embargo se nota un cambio en la estimación de la variancia del error intra unidad.

Referencias bibliográficas

Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics*, 46, 673-787.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and SPLUS*. New York: Springer-Verlag.

Yeap B.Y. and Davidian M. (2001). Robust two-stage estimation in hierarchical nonlinear models. *Biometrics*, 57, 266–272.

Williams J. D., Birch J. B. and Abdel-Salam A.- S.G. (2015). Outlier robust nonlinear mixed model estimation. *Statistics in Medicine*, 34, 1304–1316.

Palabras Clave: Datos correlacionados - Modelos no lineales mixtos – Estimación robusta – Crecimiento maíz

Área de Aplicación: Ciencias Agropecuarias

Ciencias Exactas y Naturales

An orange horizontal arrow pointing to the right, with a white outline of a double arrowhead at its tip.

ESTUDIO DE LAS PROPIEDADES DISTRIBUCIONALES DE LOS ESTIMADORES LASSO

Iván Millanes*, María Belén Allasia, Marta Beatriz Quaglino

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística (IITAE).

Facultad de Ciencias Económicas y Estadística, UNR.

*ivanmillanes26@gmail.com

Resumen

Para estimar los parámetros de un modelo de regresión lineal, es común utilizar el método de mínimos cuadrados ordinarios (MCO), método popular por su fácil aplicación y buenas propiedades. Sin embargo, cuando el número de variables explicativas (p) es mayor que el número de observaciones (n) el estimador MCO no es único y pierde sentido la interpretación de los resultados. Incluso cuando p es menor pero cercano a n -pese a la existencia de una estimación mínimo-cuadrática única- puede no ser conveniente la utilización de este método debido a la gran variabilidad de las estimaciones. En este contexto es común utilizar métodos de regularización, que minimizan una Suma de Cuadrados del Error Penalizada. Uno de ellos es la denominada regresión LASSO, método que contrae algunos coeficientes y fija en cero a otros, realizando un procedimiento similar al de selección de variables.

En el presente trabajo se realiza un estudio por simulación para estudiar propiedades distribucionales de los estimadores LASSO en escenarios donde p es mayor que n modificando el número de parámetros significativos en el modelo (k), esto es la cantidad de parámetros distintos de cero. Se simulan aleatoriamente conjuntos de observaciones de la variable respuesta a partir de un modelo pre-establecido y con cada conjunto se ajustan los caminos de solución LASSO para 50 valores del parámetro de penalización, reteniendo en cada caso sólo el estimador del vector de parámetros que produce el menor Error Cuadrático Medio (ECM).

Se construyen las distribuciones empíricas de los estimadores obtenidos considerando globalmente aquellos que son significativos y los que no lo son.

Se observa que para estimar parámetros no significativos los estimadores LASSO resultaron insesgados. Al estimar parámetros significativos y para k pequeño, los estimadores LASSO son sesgados, con poca variabilidad. Al aumentar k , el método LASSO fuerza a muchos estimadores a ser cero a pesar de que sean significativos.

Referencias bibliográficas

Bien, J. (2016). The simulator: An engine to streamline simulations. Submitted.

- Friedman, J., Hastie, T., and Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. R package version, 1(4).
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Amsterdam: Elsevier.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*, 2nd edition. New York, NY: Springer.
- Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Boca Raton, FL: CRC press.
- Larose, D. T. and Larose, C. D. (2015). *Data mining and predictive analytics*. Hoboken, NJ: John Wiley & Sons.
- Nisbet, R., Miner, G., and Elder IV, J. (2009). *Handbook of statistical analysis and data mining applications*. London: Academic Press.
- Osborne, M. R., Presnell, B., and Turlach, B. A. (2000). A new approach to variable selection in least squares problems. *IMA journal of numerical analysis*, 20(3): 389-403.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267-288.
- Tibshirani, R. (2017). *Machine Learning 10-702: Sparsity, the Lasso, and Friends*. Retrieved from: <http://www.stat.cmu.edu/ryantibs/statml/lectures/sparsity.pdf>.
- Tibshirani, R. J. et al. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456-1490.
- Yan, X. and Su, X. (2009). *Linear regression analysis: theory and computing*. Singapore: World Scientific.

Palabras Clave: LASSO - Big Data - Distribución empírica

Área de Aplicación: Ciencias Exactas y Naturales.

APLICACIÓN EN R PARA LA DETECCIÓN DE ASOCIACIÓN ENTRE IMÁGENES

Silvia P. Melo¹, Rodolfo C. J. Salomón²

1.-Departamento de Matemática, Universidad Nacional del Sur, Argentina

2.- Departamento de Geología, Universidad Nacional del Sur, Argentina

smelo@uns.edu.ar

Resumen

Al presente se encuentra muy extendida la utilización de imágenes en numerosas áreas de investigación y con diversos tipos de aplicaciones. En ocasiones, se cuenta con una sucesión de imágenes, obtenidas con el fin de lograr información referida al desarrollo de un determinado proceso en el tiempo. En este trabajo se presenta la implementación de una aplicación realizada en R-software (*R Core Team*, 2018) para la detección de asociación entre imágenes.

Primera parte: dada una sucesión de p imágenes que registran los cambios producidos por un determinado proceso y/o tratamiento en el tiempo, se realiza un estudio descriptivo, basándose en la aplicación de un Análisis de Componentes Principales a la matriz de información de orden $p \times n$, construida con los valores de luminosidad correspondientes a las p imágenes con n puntos de resolución. El análisis de las correlaciones entre Variables y Componentes que aporta el ACP permite identificar grupos de variables (Johnson & Wichern, 2007), esto es, grupos de imágenes. Aquéllas que presenten alta correlación con alguna de las nuevas variables construidas constituyen un grupo, es decir, la variable obtenida por ACP asocia puntos con luminosidad semejante. (Salomón & Melo, 2014).

Segunda parte: dado que la correlación variable / componente no nos permite argumentar sobre la correlación o asociación variable / variable (Johnson & Wichern, 2007) se aborda este tema aplicando el concepto de entropía. (Abramson & Miguel Menoyo, 1986). Para esto, se considera la información subyacente en cada imagen, como una matriz de $m \times n$ puntos de resolución, cuyos elementos corresponden a los valores de luminosidad tomados en escala de grises. En base a esta matriz, se estima la distribución de probabilidades de la escala de grises por medio de las frecuencias relativas observadas, lo que permite el cálculo de la entropía.

Finalmente para detectar la asociación – disociación entre imágenes se toma como punto de partida la entropía de la imagen inicial y se consideran las diferencias con los correspondientes valores de entropía de las restantes imágenes. Esto es:

pequeños cambios generados por el tratamiento aplicado, no modificarán sustancialmente la distribución de probabilidades inicial. Cuando estos cambios sean mayores, la disociación entre las imágenes quedará evidenciada, tornándose significativa la diferencia entre las entropías. (Melo & Salomón, 2017).

La información se complementa con los correspondientes gráficos.

En el trabajo en laboratorio, en el proceso e interpretación de imágenes, parte del mismo queda aún supeditado a la pericia del personal que lo realiza. La implementación de esta aplicación, tiene por objetivo brindar un método automático, repetible en el tiempo e independiente del usuario.

Referencias bibliográficas

Abramson, N., & Miguel Menoyo, J. (1986). *Teoría de la Información y Codificación* (5th ed.). Madrid: Paraninfo.

Johnson, R., & Wichern, D. (2007). *Applied multivariate statistical analysis*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Melo, S., & Salomón, R. (2017). *La entropía como método para comparar imágenes en el tiempo*. Presentación, Congreso Interamericano de Estadística XLV Coloquio de la Sociedad Argentina de Estadística XXII Reunión Científica del Grupo Argentino de Biometría Jornadas del Instituto Interamericano de Estadística, Rosario, Argentina.

R Foundation for Statistical Computing, (2018). R Core Team (Version 3.5.0) [Windows]. Vienna, Austria.

Salomón, R., & Melo, S. (2014). *Aplicación de Análisis de Componentes Principales como método para la detección de asociación de imágenes*. Presentación, XI Congreso Latinoamericano de Sociedades de Estadística CLATSE 2014 y XLI Jornadas Nacionales de Estadística, La Serena, Chile.

Palabras Clave: Componentes Principales – Imagen – Asociación – Entropía – Probabilidad – R software.

Área de Aplicación: Ciencias Exactas y Naturales.

Otras Ciencias de la Salud

An orange horizontal arrow pointing to the right, with the text "Otras Ciencias de la Salud" centered inside it.

MODELOS PARA DATOS DE CONTEO CON EXCESO DE CEROS: UNA APLICACIÓN SOBRE EL MAL DE CHAGAS

Guillermina Harvey, Gabriela Boggio

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística.

Universidad Nacional de Rosario. Argentina.

gharvey@fcecon.unr.edu.ar

Resumen:

Un problema frecuente al modelar datos de tipo conteo es que la variancia observada sea mucho mayor que la especificada por el modelo propuesto. Este fenómeno, denominado sobredispersión, es habitual al postular el modelo Poisson caracterizado por la igualdad entre su media y variancia, y trae como consecuencia la sobrestimación de la significación de los efectos de las covariables. En este trabajo se comparan diferentes alternativas para el tratamiento de la sobredispersión debida específicamente al exceso de ceros en datos provenientes de un estudio experimental sobre el tratamiento de la enfermedad de Chagas. El objetivo de la experiencia realizada por el Instituto de Inmunología de la Universidad Nacional de Rosario, fue comparar tres tratamientos versus un control y evaluar cuál producía una reducción mayor en el número de parásitos en sangre, en ratas inoculadas con *Trypanosoma cruzi*, agente causal de la enfermedad de Chagas. Una alternativa es utilizar la distribución binomial negativa (BN) o bien enfoques que reconocen dos mecanismos diferentes para la generación de los conteos. Uno de ellos es el modelo Hurdle bajo el cual un mecanismo determina que un conteo sea nulo o no mediante un modelo binario y si el conteo es positivo, un segundo mecanismo se explicita a través de un modelo truncado. Otro enfoque es el modelo “cero inflado”, mezcla de dos procesos estadísticos, uno que genera sólo conteos iguales a cero y otro, tanto conteos ceros como positivos (Lambert, 1992; Ridout et al., 1998). A estas alternativas de modelización se suma la propuesta desarrollada por Bonat et al. (2017) de considerar una nueva clase de modelos lineales generalizados basados en los modelos con dispersión Poisson-Tweedie (P-T) (Jørgensen y Kokonendji, 2016). La familia P-T está dada por una especificación jerárquica: $Y/Z \sim \text{Poisson}(Z)$ y $Z \sim \text{Tw}_p(\mu, \varphi)$, con parámetro de potencia $p \geq 1$ y φ , parámetro de dispersión. Para el caso especial de $p=2$ se obtiene el modelo BN, una mezcla Poisson-gamma. Las densidades Tweedie, $\text{Tw}_p(\mu, \varphi)$, se caracterizan por presentar funciones de variancia de la forma: $V(\mu) = \mu^p$, con $p \in \mathbb{R} \setminus (0,1)$. Este parámetro juega un rol

importante ya que permite abarcar un abanico de posibilidades para captar el exceso de ceros. Al ajustar el modelo Poisson, como era de esperar, los tres tratamientos resultan significativos mientras que bajo el modelo BN resultan significativos dos tratamientos en comparación con el control. Por su parte, al ajustar el modelo Hurdle y el “cero-inflado” se observan valores de AIC y BIC mayores que en el BN por lo que se prefiere este último, más parsimonioso. Al considerar la familia P-T, se encuentra un valor estimado para p igual a 1,84, muy cercano al correspondiente al caso gamma ($p=2$). Esto hace que los errores estándares estimados sean muy similares entre ambos modelos, manteniendo la significación de dos de los tres tratamientos. Las diferentes alternativas aplicadas permitieron responder eficientemente al interrogante biológico planteado. El uso de la familia P-T facilita la elección de la distribución eligiendo el mejor ajuste dentro de dicha familia, eliminando la incertidumbre que la elección de una distribución generalmente conlleva. Además, permite recurrir a modelos simples que no necesitan complejizarlo con un predictor lineal adicional para describir el exceso de ceros. Por último, la implementación en R de la familia P-T facilita su aplicación.


Referencias bibliográficas

- Bonat, W.H., Jørgensen, B., Kokonendji, C.C., Hinde, J., y Demetrio, C.G.B. (2017) Extended Poisson–Tweedie: Properties and regression models for count data. *Statistical Modelling*, 18, 1–26.
- Jørgensen, B. y Kokonendji, C.C. (2016) Discrete dispersion models and their Tweedie asymptotics. *Advances in Statistical Analysis*, 100, 43–78.
- Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34 (1), 1–14.
- R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ridout, M.S., Demetrio, C.G.B., y Hinde, J.P. (1998) Models for count data with many zeros. *Proceedings of the XIXth International Biometrics Conference*, Cape Town, South Africa.

Palabras Clave: datos de conteo – sobredispersión – modelos Poisson-Tweedie.

Área de Aplicación: Otras Ciencias de la Salud.

Otras Ciencias Económicas, Administración y Negocios



**ANÁLISIS DE LA SATISFACCIÓN MEDIANTE ECUACIONES
ESTRUCTURALES. EL CASO DE LOS MATRICULADOS EN EL CONSEJO
PROFESIONAL DE CIENCIAS ECONÓMICAS DE LA PROVINCIA DE
CÓRDOBA**

Mariana V. Gonzalez, Rocío M. Cerino, Fernando A. Pronello y Franco D. Vico

Departamento de Estadística y Matemática, Facultad de Ciencias Económicas,
Universidad Nacional de Córdoba mvgonzalez@eco.unc.edu.ar

Resumen

La consolidación de relaciones estables entre cualquier tipo de organización y sus principales clientes se ha convertido, en la actualidad, en una herramienta imprescindible para garantizar la supervivencia de dichas organizaciones (Schlesinger, Taullet, Bonillo Sánchez, 2014). Lógicamente esta situación también afecta a los Colegios Profesionales, como entidades de derecho público que nuclean a los egresados, reglamentando y ordenando el ejercicio de las profesiones. En este contexto, la presente investigación aporta un modelo de relaciones estructurales que permite identificar las variables observables y latentes que mayor influencia ejercen sobre la Imagen del Consejo Profesional de Ciencias Económicas de la Provincia de Córdoba (CPCE), entendida como la manera en que el profesional matriculado percibe a la organización de acuerdo a sus experiencias (Lewis y Soureli, 2006), la Valoración de la Matrícula de dicha institución, exigida para el ejercicio de la profesión y la Satisfacción, entendida como la percepción del cliente sobre la medida en que sus necesidades, metas, y deseos han sido cubiertas completamente (Oliver, 1999; Grigoroudis y Siskos, 2010).

Los datos utilizados en esta investigación provienen de una encuesta de tipo personal, realizada a matriculados en el CPCE, egresados de Universidades Públicas y Privadas de la Provincia de Córdoba. La muestra quedó conformada por 330 observaciones completas. Para la construcción de los indicadores, se seleccionó una escala de Likert con 5 categorías, permitiendo un nivel intermedio o indiferente (1: Totalmente en desacuerdo; 3: Ni desacuerdo ni en desacuerdo; 5: Totalmente de acuerdo).

Se definieron como variables latentes: Imagen del CPCE y Valoración de la Matrícula que actúan como variables exógenas del modelo (no reciben el efecto de otra variable) y Satisfacción, que se definió como endógena. Cada variable latente está medida mediante variables observables o indicadores.

Posteriormente se aplicó la Metodología de Ecuaciones Estructurales (Structural Equations Models SEM), que abarcan una familia de modelos estadísticos multivariantes que permiten estimar el efecto y relaciones entre múltiples variables. El método de estimación utilizado fue el de Máxima Verosimilitud (ML) que es coherente, no sesgado, eficiente, invariable al tipo de escala y normalmente distribuido si las variables observables responden a las condiciones de normalidad, que es lo que se verificó en este caso.

Luego de varias corridas de los modelos se estimó un ajuste con seis indicadores reflexivos significativos para los constructos exógenos Imagen y Valoración. La variable latente Imagen, además, fue utilizada para predecir el constructo endógeno Satisfacción, muy importante en el modelo, que quedó reflejada en tres indicadores. Los factores identificados en relación a la Satisfacción fueron los vinculados a la representación profesional, el acceso a los beneficios sociales y la mejora en las condiciones de la profesión frente a la sociedad. Se detectó, además, una correlación estadísticamente significativa entre las dos variables latentes Imagen y Valoración, como así también correlaciones entre algunas indicadores, que se incluyeron en el modelo a fin de mejorar el ajuste. Los indicadores de bondad de ajuste *RMSEA*, *SRMR* y *CFI* resultaron adecuados.

Referencias bibliográficas

- Acock, A. (2013), *Discovering Structural Equation Modeling Usig Stata*, Nueva York, Estados Unidos, Stata Press.
- Grigoroudis, E. y Siskos, Y. (2010), *Customer Satisfaction Evaluation: Methods for Measuring and Implementing Service Quality*, Texas, Estados Unidos, Springer.
- Lewis, B. & Soureli, M. (2006). The antecedents of customer loyalty in retailbanking. *Journal of Consumer Behaviour*.5(1), 15-31.
- Oliver, R. (1999). Whence consumer loyalty?. *Journal of Marketing*.63(1), 33-44.
- Ruiz, M. A., Pardo, A. y San Martín, R. (2010). Modelos de Ecuaciones Estructurales. *Papeles del Psicólogo*. 31(1). 34-45.
- Schlesinger, M. W., Cervera, A., Iniesta, M.A & Sánchez, R. (2014). Un enfoque de marketing de relaciones a la educación como un servicio: aplicación a la Universidad de Valencia. *Innovar*, 24(53), 113-125.

Palabras Clave: Valoración– Imagen

Área de Aplicación: Otras Ciencias Económicas, Administración y Negocios.

Otras Ciencias Sociales y Humanas

An orange arrow graphic pointing to the right, with a white outline of the arrowhead.

TRIANGULACIÓN METODOLOGICA PARA EVALUAR LA PRESTACIÓN DE UN SERVICIO

Norma Patricia Caro y María Inés Stimolo

Facultad de Ciencias Económicas – Universidad Nacional de Córdoba

pacaro@eco.unc.edu.ar; mstimolo@eco.unc.edu.ar

Resumen

La triangulación metodológica se refiere del uso de dos o más métodos de investigación sobre un mismo objeto de estudio y puede ocurrir a nivel de diseño o en la recolección de los datos. En la recolección de datos se combinan dos o más fuentes de datos con similares aproximaciones en el mismo estudio para medir una misma variable, utilizando distintos formatos y tipos de datos. También puede realizarse una triangulación entre métodos cuando un método fortalece la debilidad del otro y con esta combinación de métodos las observaciones alcanzan lo mejor y superan las debilidades. De esta manera se analiza un mismo grupo de datos con el propósito de validación.

La selección de información debe seguir los criterios de pertinencia (tomar en cuenta solo lo que se refiere a la investigación) y relevancia (en relación con el tema que se pregunta).

En este trabajo se abordan las características y ventajas del uso combinado y complementario de enfoques cuantitativos y cualitativos en recolección de datos, recopilando información a través de preguntas en un formulario estructurado y entrevistas, combinando un análisis cuantitativo y cualitativo, respectivamente. La finalidad es validar los resultados obtenidos durante el trabajo de campo

El estudio tiene como objetivo evaluar la prestación de un servicio asistencial de alimentación en una muestra de comedores escolares en la provincia de Córdoba en las dimensiones que corresponden a calidad del menú, condiciones del personal que presta el servicio, infraestructura y mobiliario y la importancia de la alimentación.

A los responsables de 482 comedores se les realizó una entrevista donde se le pidió una opinión general sobre aspectos positivos, negativos y a cambiar del servicio prestado en el comedor y por otra parte se les pidió que respondan un cuestionario con preguntas estructuradas donde se midieron los mismos aspectos. Por otra parte se utilizaron fichas de observación en 824 visitas (algunos comedores fueron visitados más de una vez)

La metodología utilizada consistió en un análisis cualitativo realizado con el software Atlas.ti, definiendo categorías y subcategorías de análisis a las que fueron

asociadas las opiniones que resultaron de las entrevistas realizadas, el que fue relacionado con análisis descriptivo de las respuestas cuantitativas obtenidas de los responsables, y las fichas de observación

De la mixtura metodológica realizada se obtienen resultados cuantitativos y cualitativos que contribuyen a la evaluación del servicio asistencial, en los aspectos mencionados, confirmando las conclusiones obtenidas con ambos métodos de análisis.

Los responsables del programa destacaron principalmente los beneficios relacionados con el aspecto nutricional y con el aporte que el programa tiene para la alimentación de la población más vulnerable.

Referencias bibliográficas

- Valencia, M. M. A. (2000). La triangulación metodológica: sus principios, alcances y limitaciones. *Investigación y educación en enfermería*, 18(1), 13-26.
- Cantor, G. (2002). La triangulación metodológica en ciencias sociales. Reflexiones a partir de un trabajo de investigación empírica. *Cinta de moebio*, (13).
- Cisterna Cabrera, F. (2005). Categorización y triangulación como procesos de validación del conocimiento en investigación cualitativa. *Theoria*, 14(1).
- Fielding, N., & Schreier, M. (2001, February). Introduction: On the compatibility between qualitative and quantitative research methods. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 2, No. 1).
- Hollstein, B. (2011). Qualitative approaches. *The SAGE handbook of social network analysis*, 404-416.
- Pacheco, E., & Blanco, M. (2002). En busca de la "metodología mixta" entre un estudio de corte cualitativo y el seguimiento de una cohorte en una encuesta retrospectiva. *Estudios demográficos y urbanos*, 485-521.

Palabras Clave: Triangulación Metodológica – Metodología Mixta – Servicio Asistencial

Área de Aplicación: Otras Ciencias Sociales y Humanas.

ESTUDIO COMPARATIVO DE LOS PATRONES DE TRANSICIÓN A LA ADULTEZ

Nancy Stanecka, Patricio Canalis y Macarena Cancinos

Facultad de Ciencias Económicas. Universidad Nacional de Córdoba. Argentina.

nstanec@gmail.com

Resumen

El pasaje de la niñez a la adultez, se manifiesta con características y duraciones diferentes de acuerdo a las distintas culturas, razas, clases sociales y género. Miradas en

distintos sentidos nos llevan a concluir que los límites de esta etapa de la vida son cada vez más difusos. Un criterio demográfico aceptado en occidente, considera a la juventud como un período etario de la vida humana que transcurre entre los 15 y los 29 años de edad.

Este estudio tiene por objetivo indagar sobre los cambios de patrones de transición a la vida adulta de los jóvenes de entre 15 y 29 años de la provincia de Córdoba y de la Ciudad Autónoma de Buenos Aires (Argentina) en el período que va de 1991 a 2010.

A tal efecto se tendrán en cuenta cinco eventos representativos como lo son la finalización de los estudios formales, el abandono del hogar materno, el ingreso al mercado de trabajo, la unión consensual y la tenencia de hijos.

Dado que no se cuenta con encuestas específicas sobre el tema se ha recurrido a la utilización los Censos Nacionales de Población y Vivienda correspondientes a los años 1991, 2001 y 2010.

Se sigue la metodología adoptada por Melo Vieira y Pau Miret (2010), consistente en el análisis de la entropía de combinación de estatus, la cual permite resumir en una medida concreta la intensidad de la transición hacia la adultez y el aporte a la misma de los distintos eventos que determinan la emancipación del joven.

Se pretende observar las tendencias manifestadas en general en estas dos poblaciones y las particularidades de cada una de ellas consideradas en una perspectiva comparativa. La elección de los casos se fundamenta en la disparidad que manifiestan respecto a dimensiones geográficas, económicas, sociales y demás.

Como anticipo del estudio, podemos afirmar que se evidencia un retraso en la edad ápice en el caso de Córdoba, en concordancia con la evidencia internacional, no así en el caso de la CABA, en la cual hay un quiebre en el censo de 2001, posiblemente motivado por la crisis económica.

Además se analiza los diferenciales por sexo, en lo que respecta a la cúspide de máxima entropía y al patrón de transición.

En un momento posterior, se eliminan estatus a los efectos de considerar la importancia relativa de cada uno de ellos en el índice de entropía. A partir de la eliminación de variables los resultados permiten detectar claramente la contribución al índice de cada evento, analizado por grupo etario.

Referencias bibliográficas

- Billari, F. C. (2001). The analysis of early life courses: Complex descriptions of the transition to adulthood. *Journal of Population Research*. 18(2), pp 119-142.
- Fussell, E. (2006). Structuring the transition to adulthood: An entropy analysis of the early life course in the United States, 1880 to 2000. En *Annual Meeting 2006, Population Association of America, Los Angeles*.
- Melo Vieira, J. y Miret Gamundi, P. (2010). Transición a la vida adulta en España: una comparación en el tiempo y en el territorio utilizando el análisis de entropía. *Revista Española de Investigaciones Sociológicas (Reis)*. (131), pp. 75-107.
- Montenegro, Á. (2011). Información y entropía en economía. *Revista de Economía Institucional*. 13(25), pp. 199-221.
- Theil, H., (1972), *Statistical Decomposition Analysis with Applications in the Social and Administrative Sciences*, Amsterdam/London: North-Holland Publishing Company.

Palabras Clave: Entropía – Transición – Juventud– Censo.

Área de Aplicación: Otras Ciencias Sociales y Humanas (Demografía)

ESTUDIO DE LA INCIDENCIA DE FACTORES DEMOGRÁFICOS SOBRE LA INSERCIÓN LABORAL Y LA VIDA EN PAREJA DE LAS JÓVENES CORDOBESAS

Nancy Susana Stanecka y Valentina Caro

Instituto de Estadística y Demografía Facultad de Ciencias Económicas. UNC

nstanec@gmail.com; valentina.car@unc.edu.ar

Resumen

La Capital cordobesa tiene una población con dinámica propia. Año a año presenta una alta influencia de jóvenes migrantes que llegan a esta ciudad desde distintas localidades, provincias y hasta desde otros países, con el anhelo de seguir una carrera universitaria, mejorar sus condiciones de vida y/o trabajar, lo que hace que se enfrenten al desafío de valerse por sí mismos.

Por otro lado, la independencia de la mujer y los avances en la autonomía femenina han determinado ciertos cambios, en los últimos tiempos, en el perfil de las mujeres jóvenes que habitan esta población. En particular, los porcentajes relacionados a la emancipación residencial, a la convivencia en unión y la participación en la fuerza laboral superan en el año 2010 a los alcanzados en el año 2001 (Stanecka y Caro, 2017)

En virtud de la inexistencia de encuestas específicas sobre juventud, en trabajos previos realizados por estas autoras, se usaron datos censales y la información proveniente de las EPH, para el análisis de la población de mujeres jóvenes, cuyas edades están comprendidas entre los 15 y los 29 años del departamento Capital de la provincia de Córdoba en los años 2001 y 2010. Los resultados muestran que el número de jóvenes ha ido en aumento pero la proporción de este grupo etario, con respecto al total de mujeres cordobesas, disminuyó en el último censo. En el año 2001, las mujeres jóvenes representaban el 26,87% de la población total de mujeres de este departamento, disminuyendo su participación al 25,96% en el año 2010. En el año 2001, por ejemplo, un 12,37% de las jóvenes que vivían en el departamento Capital provenían de otra localidad, municipio, provincia o país (a estas jóvenes la denominaremos no nativas), mientras que el porcentaje en el año 2010 fue del 11,52%³. A su vez, se detectó que hay un importante porcentaje de esta población que proviene del interior de la provincia, de otras provincias del país y también del extranjero, sobre todo en el rango etario de 20 y 24, que coincidiría con la edad en la cual las jóvenes se encuentran normalmente cursando su carrera terciaria o universitaria o buscando trabajo.

El objetivo de este trabajo fue indagar sobre la incidencia de la condición de procedencia de las jóvenes (nativas y no nativas) en el ámbito laboral y en que se encuentren en pareja o no, además de determinar el efecto de otros factores sociodemográficos.

Para ello se recurrió a modelos de regresión logística binaria (Hosmer and Lemeshow, 1989), utilizando como fuente principal la información proporcionada por la Encuesta Permanente de Hogares (EPH), para el estudio de los años 2016 y 2017, los permitieron detectar algunas relaciones significativas entre las covariables seleccionadas y las variables respuestas.

Referencias bibliográficas

Arnett, J.J. (1997). Young People's Conceptions of the Transition to Adulthood. *Youth & Society*, 29 (1), 3-23

Ciganda, D. (2008). Jóvenes en transición hacia la vida adulta: el orden de los factores ¿no altera el resultado?. En C. VARELA PETITO (coord.), *Demografía de una sociedad en transición: la población uruguaya a inicios del siglo XXI*, 69-82. TRILCE-UNFPA, Montevideo.

³ Fuente: Censos Nacionales de Población y Vivienda 2001 y 2010.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression*, Wiley, New York, 1989.

Laplante, B., & Street, M. (2009). Los tipos de unión consensual en Argentina entre 1995 y 2003: una aproximación biográfica. *Estudios Demográficos y Urbanos*, 24(2), 351-387.

Solís, P., Cerrutti, M., Giorguli, S., Benavides, M. y Binstock, G. (2008) “Patrones y diferencias en la transición escuela- trabajo en Buenos Aires, Lima y la Ciudad de México”. *Revista Latinoamericana de Población*, 1(2), 127-146.

Stanecka, N y Caro V. (2017) Mujeres cordobesas y su transición a la adultez. (Departamento Capital de la provincia de Córdoba. Años 2001 y 2010). Working paper presentado en las XIV Jornadas Argentinas de Estudios de Población y I Congreso Internacional de Población del Cono Sur Santa Fe. Argentina.

Palabras Clave: Regresión Logística – Mujeres – Jóvenes – Trabajo –Unión Conyugal

Área de Aplicación: Otras Ciencias Sociales y Humanas: Demografía

Otras Aplicaciones

An orange horizontal bar with a white arrow pointing to the right, containing the text "Otras Aplicaciones".

FACTORES DE RIESGOS DE ENFERMEDAD CARDIOVASCULAR EN PACIENTES CON DIABETES Y SU REPERCUSIÓN EN SUS ÓRGANOS BLANCO EN LA OBRA SOCIAL DE DOCENTES PARTICULARES (OSDOP)

Kucukbeyaz Diana Rosana, Amaya Luis , Hugo Abuin, Mary Prichoiko, Marcelo Frugoni, Guadalupe Mangialavori y Fernando Ventura.

Untref (Universidad Nacional de Tres de Febrero)

Osdop (Obra Social de Docentes Privados)

ranush2003@gmail.com; facundo.amaya@gmail.com

Resumen

La diabetes es una enfermedad prevalente en la República Argentina (8.5% de la población). Gran parte de los individuos desconoce padecerla y es uno de los factores de riesgo de enfermedad cardiovascular aterosclerótica.

A su vez el resto de los factores de riesgo de enfermedad vascular afectan el daño en los vasos de esta enfermedad metabólica.

El daño vascular de la diabetes se puede dividir en micro vascular y macro vascular.

Investigamos que factores de riesgo tienen preponderancia en las lesiones de pequeños vasos y grandes vasos de los pacientes de nuestra obra social durante el periodo Diciembre – Mayo 2018.

Los datos obtenidos, los comparamos con los relevados en la Encuesta Nacional de Factores de Riesgos realizada por la Obra Social 2016-2017, bloque Diabetes.

Objetivo: Relacionar los factores de riesgo de enfermedad aterosclerótica de la base de datos de los pacientes diabéticos de Osdop con su daño de órgano blanco.

Metodología: La Obra Social tiene una planilla que debe ser completada por todos los pacientes diabéticos. La misma contiene datos de filiación, tratamiento y otros datos acerca de los daños de órgano blanco y de los factores de riesgo de la enfermedad cardiovascular.

Analizamos dichos datos en forma descriptiva y comprobamos su carga sobre los órganos blanco.

Conclusiones: A partir del análisis de nuestra base de datos de 135.000 beneficiarios, se han autoreportado 1226 afiliados diabéticos que se encuentran en tratamiento.

Hemos comprobado una prevalencia inferior a la esperada de pacientes diabéticos y a su vez inferior a la relevada por la Encuesta de Factores de Riesgo

(Osdop). Nosotros esperábamos encontrar mayor prevalencia de DBT de acuerdo a los valores evaluados.

Reconocemos de esta manera nuestra falta de detección fundamentalmente temprana de esta enfermedad.

Como en la población general de pacientes diabéticos, en nuestra base de datos de beneficiarios diabéticos, durante el período diciembre- mayo 2018, se corrobora que el tabaquismo, obesidad y la dislipemia se comportan como factores de riesgo preponderantes en el accidente del cerebrovascular y el infarto agudo de miocardio respectivamente.

Referencias bibliográficas

Epidemiología de la diabetes en la Argentina, Programa Provincial de la Diabetes del InSSeP, Chaco, Argentina.

Segunda Encuesta Nacional de Factores de Riesgo 2009, Diabetes, Ministerio de Salud de la Nación Argentina.

Kelmansky, Diana.(2009). Estadística para todos, colección: Las ciencias naturales y la matemática; 1a ed. - Buenos Aires: Ministerio de Educación de la Nación. Instituto Nacional de Educación Tecnológica. ISBN 978-950-00-0713-9

Daniel, Wayne W.(2005). Bioestadística: base para el análisis de las ciencias de la salud. 4a. ed .Limusa-Wiley. ISBN: 9789681861643

Pagano, Marcello. (2001). Fundamentos de bioestadística. 2a. ed. Thomson Learning. ISBN: 9789706860743

Palabras Clave: Factores de riesgo – Órgano blanco – Diabetes -

Área de Aplicación: Estadísticas de Salud.

ESTUDIO COMPARATIVO DE LOS ALGORITMOS EM Y VB EN PROBLEMAS DE APRENDIZAJE AUTOMÁTICO DESARROLLADOS EN PYTHON

Federico Blasi*, María Belén Allasia, Marta Beatriz Quaglino

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística (IITAE).

Facultad de Ciencias Económicas y Estadística, UNR.

*federicoblasi@gmail.com

Resumen

El algoritmo de esperanza-maximización (EM), metodología que se utiliza para resolver problemas de estimación máximo-verosímil en modelos probabilísticos que dependen de variables no observables, se aplica frecuentemente en algoritmos de

agrupamiento en aprendizaje automático. Sin embargo presenta ciertas limitaciones que impiden su utilización en problemas complejos, por ejemplo, situaciones donde la distribución a posteriori no se puede explicitar. Una solución a este inconveniente, es la inferencia bayesiana variacional (VB, por sus siglas en inglés), la cual consiste en un proceso iterativo en donde no es necesario el conocimiento explícito de la distribución a posteriori ya que la misma puede aproximarse. De acuerdo a esta metodología, la distribución posterior de las variables ocultas puede aproximarse si son condicionadas a las observaciones conocidas, haciendo posible las inferencias a partir de la maximización del límite inferior de la función de verosimilitud.

El objetivo del presente trabajo es comparar el desempeño de ambas metodologías, EM y VB, para la estimación de los parámetros de una distribución normal mixta (NM) previamente definida, considerando dos escenarios que consideran muestras de distintos tamaños ($n=50$ y $n=1000$). A tal fin se desarrolla el algoritmo VB en Python, software actualmente de preferencia en aprendizaje automático para la implementación de algoritmos y metodologías.

En el algoritmo VB, se consideran distribuciones a priori Normal-Gamma (NG) para los parámetros media y precisión (definida en cada caso, como el recíproco del desvío estándar), tomando como valores iniciales las soluciones obtenidas mediante el algoritmo EM. El procedimiento programado en Python realiza el proceso iterativo de este algoritmo de actualización de las distribuciones a posteriori de las componentes de la NM, considerando para los correspondientes pesos de las componentes, una distribución a priori Dirichlet. Se observa que existe dependencia entre los parámetros de media y precisión en las distribuciones posteriores obtenidas bajo VB, ya que a menor precisión resulta mayor la incertidumbre para el valor medio. Con las muestras de tamaño 50, al promediar las distribuciones aproximadas, se observa que el algoritmo VB identifica adecuadamente los valores medios de la distribución NM original, pero no logra captar adecuadamente la dispersión. Sin embargo, al aumentar el tamaño de muestra a 1000 se consigue una correcta representación de ambas características de la distribución original. En el primer escenario, el algoritmo VB conserva 5 componentes, mientras que en el segundo únicamente 4, es decir, logra una mejor estimación de la distribución NM original sin necesidad de estimar una por una cada una de las componentes, lo que implica un ahorro computacional de consideración en el contexto de aprendizaje automático. Esta característica diferencia al algoritmo VB respecto a EM

que trabaja con el total de las componentes. Otra ventaja del método VB es que permite obtener las mismas estimaciones máximo-verosímiles que con EM, si se utiliza como a posteriori (denominadas estimaciones MAP).

Referencias bibliográficas

- Bishop, C. M., & Mitchell, T. M. (2014). Pattern Recognition and Machine Learning. New York: Springer.
- Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. International Conference on Learning Representations Proceedings. Banff, Alberta, Canada.
- McLachlan, G., & Krishnan, T. (2007). The EM algorithm and extensions (Vol. 382). John Wiley & Sons.,
- Moon, T. K. (1996). The expectation-maximization algorithm. IEEE Signal processing magazine, 13(6), 47-60.
- Orbanz, P. (2009). Construction of nonparametric Bayesian models from parametric Bayes equations. In Advances in neural information processing systems (pp. 1392-1400).
- Tzikas, D. G., Likas, A. C., & Galatsanos, N. P. (2008). The variational approximation for Bayesian inference. IEEE Signal Processing Magazine, 25(6), 131-146.

Palabras Clave: Aprendizaje Automático - EM- Inferencia Bayesiana Variacional

Área de Aplicación: Otras Aplicaciones (Aprendizaje Automático).

CONSTRUCCIÓN DE UN ÍNDICE DE DESIGUALDAD DE GÉNERO: APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES.

María Carolina Trogliero, Martín Saino

Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

martinsaino@gmail.com

Resumen

El análisis de las desigualdades, y particularmente, la medición de las inequidades de la sociedad, se ha centrado, en general, en las carencias y exclusiones que afectan a la mayoría en términos de ingresos, de educación y de participación. Con esta investigación se propone modificar el foco de atención y reconocer la existencia de otro tipo de inequidad imperante y transversal a todas las dimensiones de la sociedad: la desigualdad de género. El principal problema con este tipo de desigualdades, es que no solo implican intrínsecamente una injusticia, sino que también limitan el logro de una sociedad más igualitaria, cohesionada e incluyente (CEPAL, 2001).

De aquí surge la motivación de la construcción, mediante el uso de técnicas de análisis multivariante, de un índice de desigualdad de género para los aglomerados que conforman las provincias argentinas. La revisión de la literatura evidencia la novedad en la temática y la necesidad de explotar y aprovechar la información vigente en cada país. Hoy en día, existe una fuerte inclinación hacia la construcción de índices con propósitos de comparación internacional lo que obliga a que el mismo sea primitivo y a veces poco representativo de las sociedades más desarrolladas, dada la limitada información en algunos países subdesarrollados o en vías de desarrollo.

Se considera de gran importancia aportar herramientas que permitan medir en forma más objetiva la realidad Argentina, explorando los diferentes usos de los métodos multivariados tanto para realizar una fotografía descriptiva de la realidad como para construir índices sintéticos.

Objetivo: Construcción y medición de un índice de desigualdad de género de los aglomerados de las provincias argentinas y el total nacional para aportar herramientas de sondeo de la inequidad existente.

Metodología: Se hace uso de herramientas de análisis exploratorio de datos y de técnicas de análisis multivariante con el propósito de medir la desigualdad de género dentro de cada unidad de análisis y, luego, comparar la situación de cada unidad respecto al total del país. Dado que el problema de la desigualdad de género es transversal a todos los campos de una sociedad, se definen cuatro dimensiones teóricas de análisis: dimensión económica, dimensión educativa, dimensión política y dimensión de salud.

La fuente de datos es diversa. Se utilizará principalmente la base de datos del tercer trimestre del año 2013 de la Encuesta Permanente de Hogares (EPH) y la Encuesta sobre Trabajo No Remunerado y Uso del Tiempo realizada por el INDEC el tercer trimestre del mismo año. Además, se recolectan datos de las Estadísticas Vitales, los mapas de género de la Oficina de la Mujer de la Corte Suprema de Justicia y de las Cámaras de Diputados y Senadores de las provincias argentinas. Para aquellas variables/indicadores que no se encuentran contempladas en las EPH se utilizará información provincial como proxy.

La investigación fue dividida en dos etapas. La primera etapa está asociada a la construcción de mapas factoriales mediante el uso de Análisis de Componentes

Principales (ACP) con el propósito de situar y describir el estado de cada dimensión teórica previamente definida.

El ACP es un método multivariante de reducción de la dimensión, es decir, permite disminuir la dimensionalidad de un problema. Tiene como objetivo explicar la mayor parte de la variabilidad total con el menor número posible de componentes, de forma que se produzca la menor pérdida de información (Peña, 2002). El mismo transforma un conjunto de variables originales (p) correlacionadas entre sí, en un nuevo conjunto de variables (k), denominadas Componentes principales, los cuales se caracterizan por estar incorrelacionadas entre sí; donde $k < p$.

Se aplica el ACP a cada dimensión teórica considerando un total de 22 variables ($p=8$ en la dimensión económica, $p=6$ en la dimensión educativa, $p=4$ en la dimensión política y $p=4$ en la dimensión de salud). Al emplear el ACP se disminuye cada campo teórico a dos dimensiones ($R=2$), permitiendo graficar en los ejes cartesianos los mapas factoriales y el círculo unitario. Cada mapa factorial evidencia disparidades existentes entre hombres y mujeres en cada jurisdicción en los diversos campos teóricos definidos.

En una segunda etapa, se aplica el ACP con el propósito de obtener los ponderadores para la construcción del índice sintético. La fortaleza del método surge de brindar la posibilidad de solucionar dos típicos problemas asociados a la elaboración de índices. Por un lado, evita la doble contabilización de la información que puede existir en la agregación del sistema de indicadores. Por otro lado, asigna valores concretos a los ponderadores asociados a cada variable (Saino, 2016).

Existen diversos procedimientos que utilizan el ACP para obtener los valores de los índices sintéticos. Algunos autores proponen como índice sintético al primer componente principal, el cual es una combinación lineal de las variables originales donde las mismas están ponderadas por los vectores característicos. De esta forma, se consigue establecer un criterio objetivo de ponderación. Otros autores, sugieren utilizar como ponderadores de las variables originales estandarizadas el producto del porcentaje de la varianza explicada por cada componente y los coeficientes de cada variable original asociados a la definición del componente. Otra posibilidad es establecer como ponderador el producto entre los autovalores – varianza del componente principal- y los vectores propios. Se opta por el tercer procedimiento para definir los ponderadores.

Los vectores propios (a_{ij}) del componente j , son los valores por los que esta ponderada cada variable en dicho componente principal. Es decir; $CP_j = a_{1j}Z_1 +$

$a_{2j}Z_2 + a_{3j}Z_3 + \dots + a_{pj}Z_p$. Mientras que los autovalores (λ_{ij}) de cada componente principal se obtienen sumando las cargas factoriales al cuadrado correspondientes a dicho componente, donde las cargas factoriales no son otra cosa que los coeficientes de correlación entre la variable y dicho componente principal (ρ_{ij}). La varianza del componente principal j es entonces: $\lambda_j = (p_{1j})^2 + (p_{2j})^2 + (p_{3j})^2 + \dots + (p_{pj})^2$. De esta forma, el ponderador de la variable original i queda definido:

$$w_i = \sum_{j=1}^k a_{ij} \cdot \lambda_j$$

A su vez, siguiendo la metodología del Programa de las Naciones Unidas para el Desarrollo (PNUD), se construyó un índice sintético para cada jurisdicción según sexo, para luego tomarle la diferencia y así medir la desigualdad de género existente. El índice para cada unidad de análisis queda definido como:

$$I_x^g = \sum_{i=1}^p w_i \cdot Z_{ix}^g$$

Donde Z_{ix} es el valor que asume la variable i en la jurisdicción x .

Entonces, para la construcción del índice de desigualdad de género se realiza la diferencia entre el valor que asume el índice para los valores asociados a los hombres y el valor que asume el mismo con la información femenina, se considera el valor absoluto de la diferencia:

$$IDG(X) = I_X^H - I_X^M = \sum_{i=1}^p w_i \cdot Z_{ix}^H - \sum_{i=1}^p w_i \cdot Z_{ix}^M = \sum_{i=1}^p w_i \cdot (Z_{ix}^H - Z_{ix}^M)$$

Resultados

Para obtener los valores del $IDG(X)$ se utiliza el Software R, se corre el ACP considerando las 22 variables definidas con anterioridad. Se realiza el Test de Esfericidad de Barlett y se evalúa la medida de adecuación muestral de Kaiser-Meyer-Olkin. En ambos casos, los resultados son más que aceptables para la aplicación del ACP.

Como criterio de retención del número de componentes, se utiliza *el criterio de la media aritmética con variables normalizadas*. Es decir, se retienen aquellos componentes con varianza mayor a 1. Por lo tanto, se mantienen los 7 primeros componentes principales. De esta forma, se reduce la dimensionalidad del problema de 22 a 7 a través del ACP, y luego, se construye un índice unidimensional que sintetiza el

80% de la información. Los valores que asume el índice de desigualdad de género para cada jurisdicción se observa, ordenado de mayor a menor, en la siguiente tabla:

Tabla 1. Índice de Desigualdad de Género

Jurisdicción	I_x^H	I_x^M	IDG (X)	IDG (X) normalizado
Catamarca	-14,77748	28,97081	43,74829	1,0000
Santiago del Estero	-20,41587	17,62054	38,03641	0,8694
Tucumán	-15,87357	17,35295	33,22652	0,7595
Chaco	-49,68736	-18,69194	30,99542	0,7085
Salta	-3,37682	26,65009	30,02691	0,6864
Santa Fe	-15,55514	11,81669	27,37182	0,6257
Total Nacional Urbano	-6,67785	19,26121	25,93906	0,5929
Córdoba	3,85203	28,56546	24,71343	0,5649
San Juan	-19,64678	2,33332	21,98010	0,5024
Entre Ríos	1,94341	23,63416	21,69075	0,4958
San Luis	-17,02892	3,13233	20,16125	0,4608
Santa Cruz	16,39091	-3,44762	19,83854	0,4535
Neuquen	-5,57805	14,10774	19,68579	0,4500
Río Negro	-13,65116	3,15603	16,80719	0,3842
Chubut	-8,15886	7,65133	15,81019	0,3614
Buenos Aires	-3,87253	10,04650	13,91904	0,3182
La Pampa	-13,99522	-0,77344	13,22178	0,3022
Corrientes	-11,89551	1,17931	13,07482	0,2989
Mendoza	0,60601	13,31705	12,71104	0,2905
Jujuy	-2,14281	9,72445	11,86726	0,2713
Formosa	-19,69636	-10,65519	9,04117	0,2067
CABA	29,51661	37,63955	8,12294	0,1857
Tierra del Fuego	4,54502	7,18350	2,63848	0,0603
La Rioja	-20,33702	-20,77913	0,44212	0,0101
Misiones	-11,87892	-11,60345	0,27547	0,0063

Es preciso explicitar que se realiza una aplicación del ACP considerando observaciones activas (hombres) y suplementarias (mujeres). La obtención de los ponderadores se realiza considerando solo aquellas observaciones activas. Con el propósito de que los índices para cada sexo sean comparables y consistentes entre sí, se utilizan los mismos ponderadores para las observaciones suplementarias.

Se realizó una normalización por el máximo. Puede observarse que la provincia con mayor desigualdad de género es Catamarca, le siguen en el ranking Santiago del Estero, Tucumán, Chaco, Salta y Santa Fe. En los seis casos nombrados, la desigualdad de género es mayor que la media a nivel nacional. Era de esperar que las primeras cinco jurisdicciones tuvieran los valores más altos, esto se debe a que están entre las provincias más conservadoras del país. Llama la atención la posición de Santa Fe. El resto de las jurisdicciones se encuentran con un nivel de desigualdad menor que la media nacional.

Es menester señalar que al construir el índice sintético de la forma en la que se lo hizo, y al obtener la desigualdad como una diferencia entre índices, no se puede inferir si un alto valor del $IDG(X)$ implica inequidad favorable para hombres o mujeres. Sin embargo, la pormenorización de cómo afecta esta desigualdad a cada uno de los géneros fue analizada en apartados anteriores.

Referencias bibliográficas

Bardhan, K., y Klasen, S. (1999). *UNDP's gender-related indices: A critical review*.

World Development, 27(6), pp. 985-1010.

- Bericat A., Eduardo (2012). *The European Gender Equality Index: Conceptual and Analytical Issues*. Social Indicators Research, Vol. 108, No. 1, pp. 1-28.
- Branisa, Boris; Klasen, Stephan; Ziegler, Maria (2009). *New Measures of Gender Inequality: The Social Institutions and Gender Index (SIGI) and its Subindices*. Courant Research Centre: Poverty, Equity and Growth. Discussion Papers, No. 10
- Castaño, Cecilia (1999). *Economía y género*. Universidad Complutense de Madrid. Política y Sociedad, 32, pp. 23-42.
- Dijkstra, A. G. (2002). *Revisiting UNDP's GDI and GEM: Towards an alternative*. Social Indicators Research, 57, pp. 301-338.
- Frias, S. (2008). *Measuring structural gender equality in Mexico: A state level analysis*. Social Indicators Research, 88, 215-2.
- Klasen, S. (2006). *UNDP's gender-related measures: Some conceptual problems and possible solutions*. Journal of Human Development, 7(2), pp. 243-274.
- Klasen, S. y D. Schüler (2011). *Reforming the Gender-related Development Index (GDI) and the Gender Empowerment Measure (GEM): Implementing some specific proposals*. Mimeo. University of Goettingen
- López-Claros, A. y S. Zahidi (2005). *Women's Empowerment: Measuring the Global Gender Gap*. Davos: World Economic Forum.
- Peña, Daniel (2002). *Análisis de Datos Multivariantes*,
- Permanyer, Iñaki (2010). *The Measurement of Multidimensional Gender Inequality: Continuing the Debate*. Social Indicators Research, Vol. 95, No. 2, pp. 181-198.
- Saino, O. Martín; Trogliero, M. Carolina y Badino, J.P. (2016). *Análisis de Componentes Principales para la construcción de índices sintéticos*. Universidad Nacional de Córdoba.
- Social Watch (2005). *Roars and Whispers Gender and Poverty: Promises versus Action*. Montevideo: Social Watch
- United Nations Economic Commission for Africa (2016). *The African Gender and Development Index: Phase 4*. Concept Notes.
- Wieringa, Saskia E. (1997). *Report of the Workshop on GDI/GEM Indicators*. Mimeo. The Hague: Institute of Social Studies.
- Palabras Clave:** Desigualdad de género, Análisis de componentes principales, Índice sintético

Área de Aplicación: Otras aplicaciones: Género y Economía.

COMPARACIÓN DE METODOLOGÍAS PARA LA CONSTRUCCIÓN DE MAPAS DE PRECIPITACIONES EN LA PROVINCIA DE SANTA CRUZ

**Dora Maglione¹, Julio Soto¹, Oscar Bonfili², José Luis Sáenz¹, Carlos Talay¹,
Marisa Sandoval¹, Miguel Llancahuen¹**

¹Universidad Nacional de la Patagonia Austral - Instituto de Trabajo, Economía y Territorio - ²Servicio Meteorológico Nacional, Oficina Río Gallegos

dmaglione@unpa.uarg.edu.ar

Resumen

En muchos problemas de la ciencia de la tierra, en especial en la climatología, es fundamental generar superficies continuas de datos climáticos a partir de un número discreto de observaciones. Para realizar esta tarea, es necesario recurrir al método de interpolación para estimar el valor de una variable en un punto que no fue observado con la ayuda de valores observados en puntos vecinos. Los métodos de interpolación conocidos en la literatura se pueden clasificar como de uno o dos niveles. Dentro del primer grupo se encuentran modelos muy bien instrumentados como los diagramas de Voronoi, método lineal, bilineal, la representación polinomial, el método Spline y el de las distancias inversas, entre algunas de las muchas propuestas. En los de dos niveles la interpolación se realiza en dos etapas, en la primera se determina una función que resume la estructura estadística del campo la cual va a servir de base para la determinación de los pesos que se usarán en la interpolación y en la segunda etapa los pesos de la aproximación se calculan a través de un proceso de optimización. Como representante de esto último tenemos el método de Kriging descrito por Matheron (1971) donde la estructura estadística del campo se explica a través del semi-variograma, y también al de Gandín (Montoya et al, 2000) donde los pesos se estiman a partir de la función de autocorrelación.

El método de Gandín (1963) es una mejora del método propuesto por Cressman (1959). En el método propuesto por Cressman los pesos dependen de la distancia entre el punto a estimar y los vecinos que se encuentren en un radio de proximidad R , y si bien es muy simple y bastante eficiente desde el punto de vista numérico (Barth et al, 2008), presenta algunos problemas tales como no poder obtener una estimación en lugares donde no se encuentra ninguna observación dentro del radio R o de devolver

campos discontinuos en regiones con muy pocas observaciones, además de asumir que todas las observaciones tienen una varianza de error similar dado que la ponderación se basa en la distancia. Para intentar evitar las discontinuidades en imágenes generadas, una alternativa es el método de Barnes en donde los pesos se definen usando funciones gaussianas. Otro método similar al de Cressman que también resuelve el problema de las discontinuidades es el de las distancias inversas pesadas (IDW).

El objetivo de este trabajo fue comparar los métodos propuestos por Cressman, el de Barnes, el IDW y el de Kriging para datos de precipitaciones en un entorno de la provincia de Santa Cruz. Para ello se calcularon valores de precipitaciones medias acumuladas sobre una grilla regular a partir de los observados y se generaron las imágenes usando los cuatro métodos. A partir de los valores predichos por los modelos y el valor real observado en cada punto, se computaron distintas medidas de errores sobre los puntos con información a fin de seleccionar la mejor metodología para poder generar el mapa de precipitaciones acumuladas de la provincia.

La comparación de las medidas de error para distintos grillados regulares sobre la región bajo estudio, demostró que el mejor de los cuatro métodos utilizados resultó ser el método de Kriging. Luego le sigue el de Cressman, aunque no siempre es posible estimar los valores de la precipitación acumulada por lo que se generan discontinuidades en el mapa. En el ámbito de la climatología existe mucha controversia sobre qué tipo de método es mejor para generar superficies climáticas a partir de un conjunto de datos observados. Si bien los tres primeros modelos son muy sencillos de implementar, en el caso de la precipitación acumulada los resultados señalan que es conveniente, a fin de generar información más precisa, utilizar el método de Kriging para la construcción de mapas en esta región geográfica.

Referencias bibliográficas

- Barth A., Alvera Azcárate A., Joassin P., Beckers J.M., Troupin C. (2008). Introduction to Optimal Interpolation and Variational Analysis. *Geohydrodynamics and Environment Research*, Nov.26.
- Cressman G.P. (1959). An operational objective analysis system. *Month Wea. Rev.* 87, No 10
- Gandin L.S. (1963). *Objective analysis of meteorological fields*. Gidrometizdat, Leningrad, 287 p.

Matheron, G. (1971). The theory of regionalized variables and its applications:
Cahierscentre morph. Math. Fontainebleau, V. 5, 211 p.

Montoya, G., J. Eslava, A. Calderón & H. Vega. (2000): Implementación del método de
Gandin para interpolar datos de precipitación en Colombia. Meteorol. Colomb.
1:25-30. ISSN 0124-6984. Santa Fe de Bogotá, D.C. – Colombia.

Palabras Clave: Mapas- Kriging - Método de Cressie - Método de Barnes –IDW-
Precipitación

Área de Aplicación: Otras Aplicaciones (Climatología)



Saber y saber hacer con estadística

XLVI COLOQUIO ARGENTINO DE
Estadística
4º Jornada de Educación



Este e-book reúne las propuestas científicas compartidas en el transcurso de la Jornada de Educación Estadística y el Coloquio Argentino de Estadística. Organizado bajo el formato de cursos, talleres y comunicaciones orales y digitales, en las categorías metodológicas: Control de procesos; Datos categóricos; Datos de duración; Datos faltantes; Diseño de experimentos; Estadística espacial; Inferencia Estadística; Métodos Bayesianos; Métodos multivariados; Métodos no paramétricos; Métodos robustos; Minería de datos; Modelos de regresión; Muestreo; Probabilidad y Procesos estocásticos; Series de tiempo y Teoría sobre distribuciones de probabilidad; aplicadas a Biología; Ciencias Agropecuarias; Ciencias Exactas y Naturales; Ecología y Medio Ambiente; Economía; Educación, Ciencia y Cultura; Enseñanza de la Estadística; Estadísticas Oficiales; Genética; Industria y Mejoramiento de la Calidad; Otras Ciencias de la Salud; Otras Ciencias Económicas, Administración y Negocios; Otras Ciencias Sociales y Humanas y Salud Humana.

Los autores son responsables de las expresiones vertidas y los contenidos en cualquiera de las modalidades de participación



Facultad de
Ciencias Económicas

Programa Producción de Datos
y Econometría Aplicada



Universidad Nacional
de Río Cuarto

Secretaría
de Ciencia y Técnica



Ministerio de
**CIENCIA
Y TECNOLOGÍA**

**ENTRE
TODOS**

GOBIERNO DE LA PROVINCIA DE
CÓRDOBA

ISBN 978-987-688-265-1



UniRío
editora



Universidad Nacional
de Río Cuarto