

## Línea de Trabajo fin de Máster

(Fecha última actualización: 16/10/2020)

<b>Máster en Estadística. CURSO ACADÉMICO 2019-20</b>	
<b>Título</b>	Aplicación de modelos de respuesta discreta
<b>Profesor(es)</b>	Manuel Escabias Machuca y Ana María Aguilera del Pino
<b>Descripción</b>	<p>En la formulación de los modelos de regresión de respuesta categórica existen multitud de variantes dependiendo de si la respuesta es binaria o de categorías múltiples (ordinal o nominal) o de la función link asociada al modelo lineal generalizado. Así, por ejemplo, distinguimos entre modelos logit y probit para respuesta binaria, o entre los modelos logit de categoría de referencia para respuesta nominal y los modelos logit acumulativos y los modelos de categorías adyacentes para el caso de respuesta ordinal, así como muchos otros menos conocidos como los “continuation-ratio logit models”, “mean response models”, “discrete-choice multinomial logit models”, etc (ver Agresti, 2002). El problema de multicolinealidad que afecta a la estimación de estos modelos se puede resolver mediante el uso de técnicas de reducción de dimensión, dando lugar a otros modelos de respuesta discreta como los modelos basados en regresión en componentes principales o en regresión PLS. Estos modelos pueden también ser generalizados al caso de predictores funcionales.</p> <p>Por otro lado, la mayoría del software estadístico estándar existente (SPSS, STATGRAPHICS,...) sólo posee módulos para el ajuste de los modelos más comunes: modelo logístico binario, de respuesta nominal de categoría de referencia o de respuesta ordinal acumulativo de ventajas proporcionales. Hay otros entornos de programación como el software libre R que permiten adaptar módulos diseñados con otros fines para ajustar modelos menos habituales. Sin embargo no poseen herramientas específicas para estos modelos más particulares, con la consiguiente ausencia de métodos de análisis de bondad de ajuste, análisis de los residuos, problemas de multicolinealidad, separación, etc, asociados a un conjunto de datos.</p> <p>El objetivo principal del trabajo fin de máster será el desarrollo de aplicaciones de modelos de respuesta discreta con datos reales de distintas áreas como las ciencias de la salud, la investigación de mercados o cualquier otro ámbito de aplicación. Los resultados se obtendrán mediante el desarrollo de algoritmos computacionales con el software libre R.</p>
<b>Objetivos particulares</b>	<ul style="list-style-type: none"> <li>• A partir de unos datos reales plantearse un problema de estimación de una variable de respuesta discreta a partir de covariables relacionadas.</li> <li>• Elegir el modelo de respuesta discreta adecuado para resolver el problema planteado: formulación del modelo, interpretación de parámetros y posibles restricciones sobre los mismos parámetros, así como estimación del modelo, inferencia, bondad de ajuste y validación.</li> <li>• Manejo del software estadístico R para la aplicación a los datos reales objeto de estudio.</li> <li>• Presentación de resultados como un informe estadístico.</li> </ul>
<b>Prerrequisitos y recomendaciones</b>	Se requieren por parte del alumno conocimientos a nivel medio de Matemáticas, Estadística y Cálculo de Probabilidades así como el manejo del software estadístico R. En concreto se precisan conocimientos previos básicos sobre Análisis de Regresión y Análisis de Datos Multivariantes.



	<p>En relación al máster, es obligatorio cursar la asignatura Modelos de Respuesta Discreta. Aplicaciones Biosanitarias y se aconseja al alumno cursa la asignatura Técnicas Estadísticas Multivariantes y Aplicaciones.</p>
<p><b>Plan de trabajo</b></p>	<ul style="list-style-type: none"> <li>• El trabajo comenzaría con una revisión bibliográfica de las publicaciones relacionadas con los aspectos metodológicos y aplicados del modelo de respuesta discreta objeto de estudio.</li> <li>• En segundo lugar se haría un resumen de la metodología estadística a utilizar.</li> <li>• Finalmente, se aplicaría la metodología a un conjunto de datos reales con el objetivo de explicar una variable de respuesta discreta a partir de variables cuantitativas y cualitativas relacionadas. Para ello se procederá a la elaboración de los programas necesarios en R, interpretación de resultados y extracción de conclusiones.</li> </ul>
<p><b>Competencias generales y específicas</b></p>	<p>COMPETENCIAS GENERALES</p> <p>CG1 - Los titulados han de saber aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.</p> <p>CG3 - Los titulados han de saber comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades</p> <p>CG4 - Los titulados deben poseer las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.</p> <p>CG5 - Los titulados han de demostrar una comprensión sistemática del campo de estudio y el dominio de las habilidades y métodos de investigación relacionados con dicho campo.</p> <p>CG6 - Los titulados deben demostrar la capacidad de concebir, diseñar, poner en práctica y adoptar un proceso sustancial de investigación con seriedad académica.</p> <p>CG8 - Los titulados deben ser críticos en el análisis, evaluación y síntesis de ideas nuevas y complejas.</p> <p>CG9 - Los titulados deben saber comunicarse con sus colegas, con la comunidad académica en su conjunto y con la sociedad en general acerca de sus áreas de conocimiento.</p> <p>CG10 - Los titulados han de ser capaces de fomentar, en contextos académicos y profesionales, el avance tecnológico, social o cultural dentro de una sociedad basada en el conocimiento.</p> <p>COMPETENCIAS ESPECÍFICAS</p> <p>CE1 - Conocer métodos para el Análisis de Datos.</p> <p>CE3 - Adquirir conocimientos avanzados en Probabilidad y Procesos Estocásticos.</p> <p>CE4 - Profundizar en las técnicas de Modelización Estocástica.</p> <p>CE6 - Aprender y entender técnicas de Estadística Multivariante.</p> <p>CE15 - Ser capaz de identificar la información relevante para resolver un problema.</p> <p>CE16 - Utilizar correcta y racionalmente programas de ordenador de tipo estadístico.</p> <p>CE17 - Adquirir capacidades de elaboración y construcción de modelos y su validación.</p> <p>CE18 - Ser capaz de realizar un análisis de datos.</p> <p>CE21 - Conocer, identificar y seleccionar fuentes estadísticas.</p> <p>CE22 - Ser capaz de interpretar resultados a partir de modelos estadísticos.</p> <p>CE24 - Ser capaz de extraer conclusiones y redactar informes.</p> <p>CE25 - Ser capaz de identificar relaciones o asociaciones.</p> <p>CE26 - Saber utilizar con destreza entornos de programación y análisis estadístico.</p> <p>CE28 - Ser capaz de desarrollar un pensamiento y razonamiento cuantitativo.</p> <p>CE29 - Potenciar la habilidad para poder sustraer o deducir lo esencial de un concepto o situación determinada con objeto de extraer la información importante y generalizar el aprendizaje a situaciones nuevas.</p>
<p><b>Bibliografía</b></p>	<p><b>Agresti, A. (2002).</b> Categorical Data Analysis (2ª edition). Wiley.</p> <p><b>Aguilera del Pino, A.M. (2001).</b> Tablas de Contingencia Bidimensionales. Colección</p>



Cuadernos de Estadística (15). La muralla.

**Aguilera del Pino, A.M. (2006).** Modelización de tablas de contingencia multidimensionales (Colección Cuadernos de Estadística, 33). La Muralla.

**Aguilera, A.M., Escabias, M. y Valderrama, M.J. (2006).** Using principal components for estimating logistic regression with high dimensional multicollinear data. *Comp. Statistics and Data Analysis*, 50(8) 1905-1924.

**Crawley, M.J. (2007).** The R book. Wiley.

**Christensen, R. (1997).** Log-Linear Models and Logistic Regression. Springer.

**Hosmer, D.W. y Lemeshow, S. (1989).** Applied Logistic regression. Wiley.

**Kateri, M. (2010).** Contingency table analysis. Birkhäuser.

**Kleinbaum, D.G. (1994).** Logistic Regression. A Self-Learning Text. Springer.

**McCullagh, P. y Nelder, J. (1983).** Generalized linear models. Chapman and Hall.

**Power, D.A. y Xie, Y. (2000).** Statistical Methods for Categorical Data Analysis. Academia Press.

**Ruiz-Maya, L., Martín Pliego, F.J., Montero, J.M. y Uriz Tomé, P. (1995).** Análisis Estadístico de Encuestas: Datos Cualitativos. Ed. AC.