



Siguiendo las Huellas de Moodle en la Medición del Rendimiento Académico Universitario con Datos Jerárquicos

Juan Luis Peñaloza Figueroa y María Pérez Martín
Departamento de Estadística e Investigación Operativa II (Métodos de Decisión)



Siguiendo las Huellas de Big-Data a Través de Moodle y los Datos Jerárquicos

El uso de grandes cantidades de datos (Big Data) para dirigir el aprendizaje de los alumnos ya es una realidad en el contexto de la educación y, en particular, en la enseñanza universitaria. Y da lugar a la formulación de una serie de preguntas, tales como:

- ¿Qué usos tendrán esos datos y quiénes controlaran esos datos?
- ¿Qué beneficios y qué riesgos traerán los grandes flujos de datos?
- ¿Será posible una formación individualizada?
- ¿Cuáles son los nuevos problemas relacionados con la privacidad de los datos de los alumnos?
- ¿Será posible descubrir nuevos patrones de comportamiento que ayuden a los profesores a mejorar la enseñanza?
- ¿Qué factores relacionados con los estudiantes, clases, grados y universidad influyen en la calificación media de los estudiantes?
- ¿Puede variar el impacto de una variable educativa si se considera el contexto (grupos o grados)?
- ¿La variación total observada entre grupos puede diferir por el volumen de alumnos por grupo?

- ❑ El uso de grandes cantidades de datos para analizar el funcionamiento de un sistema educativo, permite detectar la falta de oportunidades, actuar para reducirlas y crear un “mapa de oportunidades educativas”, además de analizar la naturaleza de los resultados del aprendizaje. De ahí que nuestro acercamiento a Big-Data lo hagamos siguiendo las huellas de Moodle con la idea de aventurar posibles respuestas a este conjunto de preguntas.
- ❑ La utilidad de estos datos reside en la capacidad de los usuarios para extraer información útil, detectar eventos y tendencias atractivas; apoyar decisiones y políticas basadas en análisis y razonamiento estadísticos; y explotar los datos para alcanzar objetivos académicos, prácticos o científicos.
- ❑ Cuando la escala de manipulación de datos y la exploración e inferencia crecen más allá de las habilidades humanas, lo lógico es echar una mirada a las nuevas tecnologías para facilitar las cosas. Moodle es una herramienta para extraer información relativa a la interacción, ya que de otro modo permanecería oculta en los petabytes (un 1 seguido de 15 ceros).

- ❑ Los procedimientos para extraer información de Moodle permiten conocer:
 - A nivel de alumnos:
 - El ritmo de estudio y avance de los estudiantes
 - La detección del cumplimiento o no de los objetivos establecidos en el programa de la asignatura por parte de los estudiantes
 - El grado y volumen de interacción entre los estudiantes; etc.
 - En el caso de los profesores:
 - Las didácticas educativas que mejoran el aprendizaje
 - El grado de aprovechamiento de las nuevas tecnologías disponibles
 - La forma de personalizar el aprendizaje
 - La manera de impulsar el trabajo colaborativo e interactivo; etc.

- En términos institucionales
 - La calidad del grado educativo
 - La manera de mejorar la gestión de los recursos escasos
 - Los logros de los estudiantes
 - El seguimiento de la actividad de sus profesores
 - etc.
- ☐ El manejo de grandes volúmenes de datos influye en áreas como:
 - Aprendizaje personalizado → configura nuevas experiencias formativas
 - Aprendizaje adaptativo → Adaptación de la didáctica a las necesidades de formación de los alumnos
 - Intervención educativa → búsqueda de soluciones al fracaso académico y fomento del trabajo por competencias

☐ **Big-Data y Datos Anidados**

La figura que sigue muestra la distribución global de todos los datos. En ella se observa que Big-data constituye la parte principal de los datos (estructurados y no estructurados), y que los “Open Data” (alimentado por el sector público) son una parte importante de Big Data. Esta distribución evidencia el carácter anidado de los “open data” (una de sus fuentes es Moodle):

Moodle está diseñado para conservar todos los atributos especificados, especialmente cuando no se conoce el alcance de los datos o su uso. De manera que tanto el diseño como los requerimientos de datos no se definen hasta que se consultan dichos datos.

How open data relates to other types of data

ILLUSTRATIVE



SOURCE: McKinsey Global Institute analysis



- Big Data no sólo ha cambiado el futuro de la investigación, sino también el de la educación. Por ejemplo, uno de los descubrimientos entre las cinco principales políticas relacionadas con los efectos académicos significativos fue el uso de datos para guiar la enseñanza-aprendizaje [Bhatia and G Vaswani (2014). "BIG Data—A Review."].
- El impacto de la presencia de grandes volúmenes de datos en la educación ha permitido el desarrollo de nuevos métodos para la enseñanza y el aprendizaje, la creación de nuevas carreras y opciones profesionales para los estudiantes, así como la explotación y aprovechamiento de acervos digitales generados por las instituciones educativas. Aunque persisten las limitaciones de acceso a la información y disponibilidad de datos generados internamente en el sistema universitario (p.e. a los datos de Moodle).
- Las tecnologías de seguimiento automático como Moodle y las técnicas de minería de datos y análisis estadístico de datos pueden proporcionar una retroalimentación rápida a estudiantes y profesores sobre su rendimiento y actividad académica. Esto es, los datos colectivos y a gran escala pueden predecir qué alumno necesita más ayuda, evitando así el peligro de fracaso o abandono [G. Siemens and D. Gasevic, "Guest Editorial-Learning and Knowledge Analytics.," Educ. Technol. Soc., vol. 15, no. 3, pp. 1–2, 2012.].



- En un futuro próximo, posiblemente veamos un mayor crecimiento de las habilidades relacionadas con el uso de datos en el sistema universitario, y es muy posible que el efecto de los grandes volúmenes de datos sobre el rendimiento académico sea aún más significativo y generalizado.
- Actualmente, se ha trascendiendo hacia un estado en el cual se necesitan nuevos métodos sustentados en la tecnología para poder hacer el seguimiento de los alumnos, mejorar las tutorías, obtener datos objetivos de sus evaluaciones, predecir los riesgos académicos o simplemente comprender el comportamiento de los grupos y la naturaleza de la interacción entre los estudiantes.
- El sector universitario, como cualquier otra área de actividad, también está inmerso en un proceso de digitalización y transformación, impulsados por la tecnología y por los nuevos modelos y formas de hacer las cosas en un mundo cada vez más digital.
- La herramienta Moodle como instrumento de digitalización de la actividad académica tiene el rol de estimular el proceso de aprendizaje, gracias a la mejora de la comunicación entre los participantes del sistema educativo universitario, y también al proceso de "personalización" de la enseñanza.

- Las barreras a la digitalización del sistema universitario provienen de:
 - La resistencia al cambio por parte de los profesores
 - La falta de conocimientos de los profesores en temas digitales
 - La limitada disponibilidad de herramientas especializadas
 - La Limitada inversión de recursos en infraestructura tecnológica y/o digital
 - La escasez de contenidos multimedia orientados al aprendizaje
 - Las dificultades del salto de las iniciativas individuales al uso masivo de tecnología

Todos ellos son temas que, de una forma u otra, con unos sesgos u otros, han estado siempre muy presentes a la hora de tratar los asunto relacionados con la introducción de la tecnología en las aulas universitarias y el evaluación del RA, en un contexto de creciente digitalización de sus actividades.

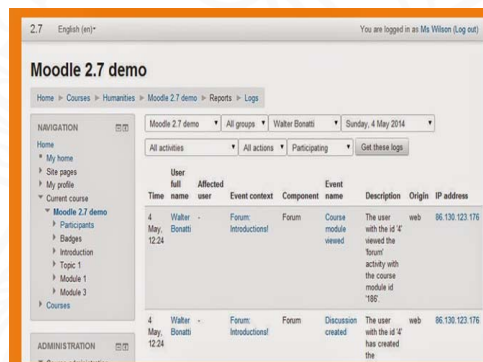
- Un ejemplo de los esfuerzos para superar estas limitaciones lo encontramos en el plan de la comunidad de Madrid para introducir la asignatura de “Programación” en los institutos. Plan a nivel autonómico que incluye a todos los Institutos de Educación Secundaria de la región: impresoras 3D, kits de robótica y el equipamiento informático necesario para impartir una asignatura de programación, además de extender en todos ellos el acceso ultrarrápido a Internet a 100 Mbps.

- Moodle como sistema de gestión del aprendizaje de fuente abierta es el más utilizado en la universidades e institutos en todo el mundo (Cole, 2005) (Las Universidades pueden acceder a sus códigos fuente y hacer los cambios que consideren oportunos). Y es generadora de información detallada sobre las actividades académicas online de los alumnos, profesores, grupos, etc.
- El procesamiento de estos datos implica:
 - Recolección de datos de Moodle.
 - Pre-procesamiento de los datos.
 - Aplicación de técnicas analíticas y estadísticas avanzadas (construir un modelo que permita descubrir y resumir el conocimiento de interés para el usuario: profesor, estudiantes o autoridad académica).
 - Interpretación, evaluación y resultados (realización de intervenciones o acciones adicionales para mejorar el aprendizaje de los estudiantes).

- ❑ Su diseño por módulos facilita la creación de nuevos cursos, añadiendo contenidos que motivan a los estudiantes, y está diseñado para soportar un estilo de aprendizaje denominado “pedagogía social constructivista” (Rice, 2006).
- ❑ Su estructura modular flexible en cuanto a actividades y recursos permite crear 5 tipos de materiales didácticos fijos:
 - Una página de textos
 - Una página web
 - Links a cualquier otra web
 - Una vista panorámica del directorio del curso
 - Una etiqueta que muestra cualquier texto o imagen
- ❑ También crea 6 tipos de materiales didácticos interactivos
 - Tareas,
 - Comentarios,
 - revistas,
 - lecciones,
 - cuestionarios y encuestas

- ❑ Además de 5 clases de actividades donde el estudiante interactúa con otros estudiantes
 - Chats
 - Foros
 - Glosario
 - Wikis
 - Talleres (workshop)
- (Romero et. al (2007))

- ❑ Moodle mantiene registros (logs) detallados de todas las actividades ejecutadas por los estudiantes y profesores. Esto es, registra cada clic que los estudiantes y profesores hacen al navegar y tiene un sistema de visualización de registros incorporado en él. Los archivos logs pueden ser filtrados por curso, participantes, día y actividad. También pueden ser utilizados por los profesores para determinar quién o quiénes han tenido una participación activa a lo largo del curso, qué es lo que hacen o cuándo lo hacen, o el tiempo utilizado en la ejecución de ciertas actividades, entre otras.



Time	User	Affected user	Event context	Component	Event name	Description	Origin	IP address
4 May, 12:24	Walter Bonatti		Forum: Introductions!	Forum	Course module viewed	The user with the id '4' viewed the forum activity with the course module id '195'.	web	96.130.123.176
4 May, 12:24	Walter Bonatti		Forum: Introductions!	Forum	Discussion created	The user with the id '4' has created the	web	96.130.123.176

Resumiendo, la generación de enormes cantidades de datos tiene un inmenso potencial tanto para mapear el rendimiento académico en base a una información detallada, como para personalizar la enseñanza, de modo que se facilite la acción del profesor, las autoridades o del propio Estado. Los riesgos asociados a la “violación” de la intimidad de los alumnos, no debe servir de justificación para no avanzar en el uso de los datos generados internamente en las universidades.



Una Mirada al Rendimiento Académico Universitario y su Medición

Juan Luis Peñaloza Figueroa y María Pérez Martín
 Departamento de Estadística e Investigación Operativa II (Métodos de Decisión)



Una Mirada al Rendimiento Académico Universitario y su Medición

Conceptualización del Rendimiento académico

- La literatura especializada ha puesto de manifiesto que el rendimiento académico no es el producto de una única capacidad, sino el resultado sintético de una serie de factores que actúan en, y desde, la persona que aprende.
- En el contexto educativo, el rendimiento académico es el resultado del aprendizaje suscitado por la actividad educativa del profesor y sintetizado en el alumno, aunque es claro que no todo aprendizaje es producto de la acción docente.
- El rendimiento académico es un concepto multidimensional, relativo y contextual, del que es difícil aportar una definición aceptada por todos (González-López, 2004). Sin embargo, desde una perspectiva operativa del término, muchos la definen como la nota o calificación media obtenida durante el periodo universitario que el alumno ha cursado (Tejedor, 1998).
- En la literatura especializada también se dice que el rendimiento académico de los alumnos universitarios se comporta como un indicador imprescindible para cualquier valoración de la calidad educativa en la enseñanza superior

- ❑ Desde el punto de vista conceptual, el modelo de RA se apoya en la noción de “función de producción educativa”, y representa el punto de partida para estudiar la efectividad existente entre el resultado del aprendizaje en las universidades y el conjunto de recursos disponibles con que cuentan para resolver el proceso de enseñanza y aprendizaje (Levin, 1983, Cooms y Hallak, 1987).
- ❑ Otro enfoque define RA como los mayores ingresos futuros de los estudiantes, debido a la incorporación de capital humano en la Facultad. La ventaja de esta medida es que se basa en el valor de mercado de los conocimientos adquiridos.
- ❑ Enfoques predominantes: el “Económico” basado en el modelo de “función de producción educativa”, en cuya estimación se utilizan distintos métodos econométricos. El “Sociológico” (y especialistas en educación) basado en el modelo de “integración” Vincent Tinto (1975) (El modelo de integración sugiere que los estudiantes se incorporan a la universidad con los atributos familiares y propios como individuo, pero tan pronto como el estudiante es admitido, una serie de factores del sistema social lo afectan y los profesores y el grupo de compañeros, lideran su desarrollo) y el Reproductivista de Pierre Bourdieu (El modelo Reproductivista sugiere que las formas de conocimiento y la distribución de la capacitación está enfocada a reproducir la división social del trabajo).

❑ **Medición del Rendimiento Académico**

- El acceso a la educación universitaria de un público estudiantil cada vez más heterogéneo en términos de su perfil socioeconómico, educativo y en aspiraciones académicas y laborales, demanda que las universidades exploren nuevos caminos pedagógicos e institucionales para lograr que estos jóvenes se gradúen, adquiriendo los conocimientos y habilidades necesarios para desenvolverse con éxito en el campo académico y profesional. Formar profesionales y científicos sin rebajar los niveles de calidad, y sobre todo buscando elevarlos, es actualmente un desafío de alta complejidad en el contexto de organizaciones de gran tamaño y modesto presupuesto (universidades).
- El estudio de los resultados académicos de los estudiantes universitarios constituye una herramienta sólida para construir indicadores que orienten la toma de decisiones en educación (universitaria) (Garbanzo, 2007).
- Se considera que los resultados de las distintas pruebas de evaluación, o algún tipo de promedio de notas obtenidas en las materias aprobadas, puede ser un primer indicador del rendimiento académico. Así, si dos alumnos iguales en otros aspectos obtienen distintas notas en los exámenes, esa diferencia indicaría que el alumno con la mayor nota ha incorporado más capital humano en su background de conocimientos.

□ Distintas formas de medición del rendimiento académico

- Como “producto” del proceso educativo sería el número de materias aprobadas por año. (Este es un indicador de productividad media del estudiante en el que el producto total es igual a la cantidad de materias aprobadas y el input la cantidad de años desde su ingreso a la Facultad).
- RA1 = promedio por materia aprobada = notas obtenida en todas las materias aprobadas incluyendo las distintas convocatorias / cantidad de materias aprobadas
- RA2 = productividad media o materias aprobadas por año = cantidad de materias aprobadas desde el año de ingreso a la facultad o grado
- RA3 = Rend1 * Rend2 = valor de la productividad media o nota promedio por año.
- RA4 = (Materias aprobadas/años desde el ingreso) / cantidad teórica (según el plan de estudios) de materias desde el ingreso.
- RA5 = RA4 x (promedio del estudiante/10). Esto es, pondera el indicador anterior por la distancia entre el promedio óptimo (10 puntos) y el promedio efectivo del estudiante.
- RA6 = La nota media de las notas de los exámenes parciales aprobados en las asignaturas

Todas estas medidas de rendimiento del estudiante consideran sólo el desempeño en la Facultad. Y permiten identificar claramente las múltiples dimensiones y determinantes que se esconden tras estos indicadores.

□ Factores que inciden en el RA

- Los resultados de diversos estudios confirman la existencia de factores en línea con trabajos anteriores. Sin embargo, otros factores como la organización y planificación personal, o la participación en actividades online, toman mayor influencia que en el pasado.
- Las universidades deberán incidir en la formación de competencias de organización y hábitos de estudio, así como en el fomento de planes de acción tutorial para contribuir a la mejora del rendimiento académico de los alumnos.

APORTACIONES RELEVANTES SOBRE LOS FACTORES QUE INFLUYEN EN EL RENDIMIENTO ACADÉMICO	
González y Cabanch, et. al (1991)	habilidades cognitivas: capacidades, conocimientos, estrategias y destrezas
Tejedor et. al (1991)	Habilidades motivacionales: disposición, intención y motivación
	Datos de identificación: sexo, estatus social, status económico
	Actitudes cívico-sociales:
	Variables contextuales
Artunduaga (2008)	Variables personales:
	Demográficas: edad, sexo, experiencia laboral, etc.
	Cognitivas: Aptitudes, rendimiento académico previo, capacidades y habilidades
	Actitudinales: interés, autoconcepto, habilidades sociales, etc.
	Variables contextuales
	Pedagógicas: actitud, formación, experiencia y personalidad del profesor, tamaño de grupo, Clima de clase, etc.
	Institucionales: tipo y tamaño de centro , políticas educativas
	Entorno social-familiar-cultural: nivel educativo de los padres, clima educativo familiar, integración social del estudiante, etc.
Tomas- Miquel et. al. (2014)	Ingreso familiar
	Ocupación laboral previa
	Miembros de la unidad familiar con formación universitaria
	test de inteligencia
	Deseo de aprender nuevo conocimientos
	Inserción laboral futura
	Percepción de obtener un título
	Personalidad activa o constante
	personalidad sociable o tolerante
	Horas de estudio
	Participación en clase
	Nivel de organización personal en el ámbito académico (organizar su tiempo)
Porto, alberto y Di Gresia, L	características del estudiante (sexo, edad);
	entorno familiar (educación de los padres, ingreso familiar
	la escuela en la que realizó los estudios del nivel secundario (pública o privada)
	la situación laboral (si trabaja o no; si trabaja, la cantidad de horas de trabajo
	motivación de los estudiantes y/o el grado de dificultad de la carrera
	características
	Características del proceso de aprendizaje son también importantes; en este grupo se incluyen los años desde el ingreso en la facultad y el año que cursa el estudiante
Gracia de Fanelli, Ana María	Asignaturas rendidas desde el momento del ingreso
	2. Asignaturas aprobadas desde el momento del ingreso
	3. Cociente entre 2 y 1: Indicador de eficiencia
	4. Cociente asignaturas aprobadas desde el momento del ingreso y años desde el ingreso: Productividad media
	5. Promedio de calificaciones (con o sin aplazos)
	6. Índice que multiplica 4 y 5: Valor de la productividad media
	7. Índice que multiplica 3 y 5
	8. Número de aplazos acumulados en la carrera
	9. Asignaturas aprobadas respecto asignaturas según duración teórica

☐ **Variables en el estudio del rendimiento académico**

▪ **Variable dependiente**

Rendimiento académico: para medir esta variable se ha considerado el promedio de las calificaciones obtenidas por el estudiante en las asignaturas del grado correspondiente. Intervalos de puntuación: [0-4,99], [5-5,99], [6-6,99], [7-7,99], [8-8,99] y [9-10].

▪ **Variables independientes**

A continuación presentamos el conjunto de variables por jerarquías, extraídas de Moodle y de otras fuentes internas de la Universidad, que serán utilizadas en este trabajo.

- En el nivel-1, el modelo de regresión es conocido como modelo de “dentro de la unidad básica de medida” (within-unit), describe los efectos en el contexto de un único grupo

Nivel-1. Entre las posibles variables (extraídas) de MOODLE asociados al primer nivel o dimensión de la jerarquía:

- Por ejemplo, las mediciones para el i-ésimo estudiante (unidad del nivel-1) en la j-ésima clase o grupo.

▪ Genero
▪ Número de veces que ha entrado en el CV
▪ Número de actividades entregadas en el CV
▪ Turno del alumno (mañana, medio día, tarde)
▪ Número de veces que se ha matriculado en una asignatura
▪ Número de veces que ha participado en un foro el alumno
▪ Si el alumno tiene beca o no
▪ Número de seminarios realizados
▪ Número de ejercicios colgados por el profesor en el CV
▪ Número de pruebas o controles realizados a través del CV
▪ Si tiene dado de alta uno o más foros
▪ Todo lo colgado el campus virtual por el profesor (diapositivas/artículos/lecturas/ etiquetas abiertas, noticias etc.)
▪ Número de etiquetas abiertas
▪ Número de Noticias publicadas
▪ Calificación de los alumnos (parciales y finales)
▪ Número de usuarios efectivos del CV
▪ Ubicación geográfica de la facultad o grado y las aulas
▪ Horarios de clase
▪ Número de veces que se han utilizado aulas de informática
▪ Número de mensajes creados en foros
▪ Dirección IP que identifica las actividades realizadas interna y externamente
▪ Número de ejercicios comentados en el foro
▪ Número de contenidos publicados en foros
▪ Número de mensajes actualizados
▪ Número de comentarios sobre comentarios de sus pares

Nivel-2. Los coeficientes de regresión del nivel-1 (β_0j , β_1j) son usados como variables respuesta y están relacionados con cada uno de las variables explicativas del nivel-2. Estos modelos (del nivel-2), también son conocidos como modelos “entre unidades”, describen la variabilidad presente a través de los múltiples grupos.

Determinar las características asociadas al grupo para determinar su efecto sobre el conjunto de alumnos en cada grupo.

- Por ejemplo, el método de enseñanza en la j-ésima clase o aula

▪ Número de alumnos matriculados en el grupo
▪ Proporción de alumnos que cursan por primera vez la asignatura
▪ ¿Qué cuelgas habitualmente en el campus? Varias opciones a marcar (Foros, tareas, chats, wiki, diapositivas, ...poner opciones)
▪ Nº de veces que ha impartido las asignaturas (listado asignaturas)
▪ Nº de veces que entra en el campus
▪ Número de actividades en grupo realizadas
▪ Proporción de alumnos que trabajan en el grupo
▪ Horarios de clase M/I/T
▪ Tiempo completo o parcial 3/4/6
▪ Grado de dificultad de la materia (desde el punto de vista profesor/ desde punto de vista del alumno)
▪ Nº campus en los que se imparte materias
▪ Doctor sí/no – Master sí/no
▪ Solo docencia/Solo tutorización TFG y TFM/ Compagina.
▪ El contenido del Campus Virtual (CV) (es totalmente nuevo cada año/ replica lo esencial pero actualiza/replica de un año a otro)
▪ Sistema de evaluación:
o ¿Pone parciales?
o No parciales solo final
o 1 parcial
o 2 parciales
o Respecto a lo anterior: Es obligatorio poner parciales?
o Trabajos calificados fuera de clase?
o Ejercicios calificados dentro y fuera de clase
o Número de exposiciones en clase
▪ Dpto al que pertenece el profesor
o ¿Siempre mismo dpto. o ha cambiado?

Nivel-3. Se corresponde con las variables relacionadas con el grado o facultad, p.e. la localización geográfica del grado y/o el presupuesto anual asignado a la facultad o universidad.

- La idea es identificar la porción de variabilidad en las calificaciones obtenidas asociadas con las aportaciones al desempeño de los alumnos por el grado o facultad.

▪ Condiciones institucionales (campus, edificio, despacho compartido/no, nº de profesores por despacho..)
▪ Tipo de grado: único o doble
▪ Área de conocimiento: Ciencias sociales/ ciencias puras/ Humanidades...
▪ Nombre del grado
▪ Materias (de los grados) básicas / OB/OP/CLES
▪ Nº de asignaturas por grado/año y cuatrimestre y nº de asignaturas por materias (en cada grado)
▪ Ubicación
▪ Turnos disponibles
▪ Participación alumnos actividades académicas e institucionales
▪ Infraestructura(sala de ordenadores / biblioteca / biblioteca digitalizada/ deportes /edificios (puntuar por edificio) / salas para alumnos (tuna, AIESEC...) / salas de estudio (individuales y grupo)
▪ Nº profesores (plantilla) por grado (teóricos con observados)
▪ Departamentos involucrados

- Resumiendo, las variables presentadas anteriormente forman parte de un fenómeno cuya información se cifra en muchas variables más o menos correlacionadas y anidadas. Por lo que es posible construir índices por niveles a partir de las variables originales, haciendo uso de técnicas multivariantes (p.e. Componentes Principales).



Modelo Multinivel Basado en Datos Moodle

Juan Luis Peñaloza Figueroa y María Pérez Martín
Departamento de Estadística e Investigación Operativa II (Métodos de Decisión)



Modelos Multinivel Basado en Datos Moodle

- El propósito de este trabajo es predecir el rendimiento académico de los estudiantes universitarios, en función de sus características educativas, en base a los datos generados por Moodle, aplicando modelos multinivel de varios niveles.
- Datos
 - La población bajo estudio está formada por los alumnos matriculados en los distintos grados de la facultad de Ciencias Económicas y Empresariales y de Comercio y Turismo, en los diferentes cursos académicos.
 - Las mediciones del perfil de los estudiantes corresponden a su desempeño académico, los cuales serán obtenidos de la plataforma de aprendizaje (Moodle) que gestiona la actividad e interacción de los alumnos entre los alumnos, grupos y grados.

Esta información será incorporada periódicamente a un único almacén de datos con un diseño orientado a ofrecer datos detallados sobre la interacción entre los alumnos, profesores y autoridades de los grados. Este proceso incluye la integración, depuración y formateo de los datos, siguiendo las técnicas usuales de pre-procesado (Dapozo y Porcel, 2005; Dapozo et al, 2007)

- ❑ El ejemplo más significativo de estudios de rendimiento a nivel internacional es el Proyecto para la Evaluación Internacional de los Alumnos (PISA). Este estudio permite realizar comparaciones internacionales entre los países participantes. Se realiza cada tres años, por iniciativa y bajo la coordinación de la Organización para la Cooperación y Desarrollo Económico (OCDE). Su objetivo es medir los conocimientos y destrezas de los alumnos de 15 años, edad próxima a la finalización de la escolaridad obligatoria, en las materias de Matemáticas, Lectura, Ciencias y Solución de Problemas.
- ❑ Para facilitar la comparación entre los resultados de los diferentes países participantes, independientemente de las diferentes formas de organización educativa y del currículo escolar, PISA no evalúa contenidos curriculares, no valora lo que se ha enseñado a los alumnos en las escuelas; es una evaluación de conocimientos y destrezas que se esperan de un estudiante que se encuentra a punto de acabar la escolaridad obligatoria.

❑ **MODELOS JERÁRQUICOS LINEALES**

- Respecto a la técnica de análisis utilizada es conveniente resaltar que los datos que proceden de las ciencias sociales y del comportamiento y, por ende de educación, tienen una estructura anidada. Por ejemplo, las calificaciones de los alumnos están agrupadas dentro de los grados o facultades, las cuales a su vez están anidadas dentro de universidades y éstas dentro de comunidades e, incluso, países.
- Los modelos multinivel representan cada uno de los niveles de agregación con un sub-modelo para expresar las relaciones que se producen entre variables dentro de un mismo nivel y especifican cómo las variables de un nivel influyen en otros niveles.
- La asociación jerárquica de los datos no es accidental, por lo que no debe ser ignorada. Ignorar los efectos de los grupos o grados puede invalidar las técnicas de análisis estadístico tradicionales que son utilizadas para el estudio de las relaciones entre datos de estas características.
- Estas técnicas estadísticas suelen incurrir en dos tipos de errores diferenciados (Hox, J.J., 1995): Falacia atomística (Alker, 1969) y Falacia ecológica (Robinson, 1950).

- ❑ Los modelos multinivel ponen solución a este problema trabajando con los diferentes niveles al mismo tiempo. Con estos modelos es posible diferenciar la varianza explicada por cada predictor en los diferentes niveles de agregación seleccionados. Además, es posible realizar inferencias con variables que actúan a diferentes niveles.
- ❑ Un problema multinivel concierne a una población con estructura jerárquica. Una muestra de tal población puede ser descrita como una muestra multicéntrica: primero tomamos una muestra de unidades del más alto nivel (por ejemplo grados), y luego muestreamos las subunidades del nivel inmediato inferior (grupos dentro de los grados), para finalmente muestrear las subunidades de nivel inferior (los estudiantes). En tales muestras, las observaciones individuales no son completamente independientes.
- ❑ El grado de homogeneidad de los contextos viene definido por la autocorrelación o correlación intraclase, que evalúa el grado de parecido entre unidades de nivel micro que pertenecen a la misma unidad macro. Entre las consecuencias de no tener en cuenta la correlación intraclase tenemos: la dependencia de orden del coeficiente de correlación y errores típicos muy pequeños debido a que los tests estadísticos se basan en el supuesto de independencia de las observaciones, cuando éstas realmente no lo son (Gaviria, J.L. y Castro, M., 2005)

- ❑ Los modelos multiniveles son diseñados para analizar variables de diferentes niveles simultáneamente, usando un modelo estadístico que incluye las diferentes dependencias.
- ❑ Si los efectos de una variable son fijos o aleatorios dependerá en gran parte del contexto o de los objetivos del estudio. En los modelos multinivel como unidades (agrupamientos) que definen los niveles, son vistos como efectos aleatorios (como muestras aleatorias de una población de estas unidades: grados, cursos, grupos, etc.). Estos efectos aleatorios se traducen en un modelo de coeficientes aleatorios que toman en cuenta la variabilidad entre agrupamientos o niveles, desde formas simples, a través de la variabilidad a nivel del intercepto, a formas más complejas, a través de variabilidades a niveles de inclinaciones de rectas.
- ❑ El modelo de regresión multinivel completo asume que hay un conjunto de datos jerárquicos, con una sola variable dependiente que es medida en el nivel más bajo y variables explicativas que existen para todos los niveles. Conceptualmente el modelo puede ser visto como un sistema jerárquico de ecuaciones de regresión.

- ¿Qué determina el número de niveles?
 - La heterogeneidad de la variabilidad presente en la variable respuesta
- ¿Cuántas dimensiones existen en una población bajo estudio?
 - Aportan a la variabilidad global las variables asociadas a los niveles de alumnos y/o los grupos (dos niveles)
 - Aportan información sobre la variabilidad global las variables asociadas a los niveles de alumnos, grupos y/o grados (3 niveles), y así, sucesivamente.
- El uso de estos predictores se verá limitado por la cantidad de varianza de Y existente en las distintas dimensiones de los datos estudiados.
- En este trabajo formulamos un modelo multinivel de tres niveles de agregación: los alumnos (nivel 1), los grupos (nivel 2) y los grados o facultades incluidas en estudio (nivel 3). Se pretende probar si las varianzas presentes en los tres niveles resultan ser significativas, es decir, mostrar que existen diferencias en el rendimiento académico entre los distintos grados y facultades incluidos en el estudio.

- El modelo general del rendimiento académico quedaría establecido como sigue:
 - **Nivel-1:** alumnos

$$Y_{ijk} = \beta_{0jk} + \sum_{p=1}^P \beta_{pjk} X_{pjk} + e_{ijk}$$

donde

Y_{ijk} = rendimiento académico del alumno i del grupo j y del grado k

β_{0jk} = rendimiento medio para todos los alumnos del grupo j y grado k

β_{pjk} = coeficientes de regresión del nivel-1

X_{pjk} = variables predictoras del nivel-1

e_{ijk} = efecto aleatorio en el nivel 1, bajo el supuesto de que $e_{ijk} \sim N(0, \sigma^2)$

- **Nivel 2: Grupos**

En el nivel 2 los coeficientes de regresión del nivel-1 son tratados como variables respuesta:

$$\beta_{pjk} = v_{p0k} + \sum_{q=1}^Q \beta_{pqk} Z_{qjk} + r_{pjk}$$

Donde

v_{p0k} = rendimiento medio para los grupos en el grado k

β_{pjk} = coeficientes de regresión del nivel-2

Z_{qjk} = predictores del nivel-2

r_{pjk} = efecto aleatorio en el nivel 2, bajo el supuesto de que tiene una distribución normal multivariante con media cero y matriz de covarianza (Σ) con una dimensión máxima de $(p + 1) \times (p + 1)$.

- **Nivel-3. Grados o Facultades**

En el nivel 3 los coeficientes beta del nivel-2 son tratados como variables respuesta. Esto es:

$$\beta_{pqk} = v_{pq0} + \sum_{s=1}^S \gamma_{pqs} W_{sk} + u_{pqk}$$

Donde

γ_{pqs} son los coeficientes del nivel-3

W_{sk} son los predictores del nivel-3

u_{pqk} = efecto aleatorio en el nivel-3, bajo el supuesto de que tiene una distribución normal multivariante con vector de medias igual a cero y matriz de covarianzas (Σ) con máxima dimensión igual a $\sum_{p=0}^p (Q_p + 1) \times \sum_{p=0}^p (Q_p + 1)$

□ De esta forma el modelo general de tres niveles estará integrado por:

$$Y_{ijk} = \beta_{0jk} + \sum_{p=1}^P \left(v_{p0k} + \sum_{q=1}^Q \left(v_{pq0} + \sum_{s=1}^S \gamma_{pqs} w_{sk} + u_{pqk} \right) z_{qjk} + r_{pjk} \right) X_{pjk} + e_{ijk}$$

- Una parte sistemática o fija que incluye todos los parámetros que definen la media del rendimiento de todos los alumnos, de todos los grupos y de todos los grados.
- Una parte aleatoria que muestra la estimación de la varianza en cada nivel de agregación, en este caso los alumnos (primer nivel), los grupos (segundo nivel) y los grados (tercer nivel), para cada parámetro incluido en la parte sistemática (estimación de la variación de las líneas individuales que representan a los niveles en torno a la línea promedio, y los rendimientos observados de los alumnos alrededor de la línea promedio de su nivel).

□ Ventajas y Desventajas de los Modelos Multinivel

- Las ventajas de usar modelos multinivel son muchas. Correctamente utilizados, estos modelos nos permiten obtener mejores estimaciones de los coeficientes de regresión y de su variación que con los modelos tradicionales. La gran flexibilidad ofrecida por los modelos multinivel se da en términos de la modelización de la estructura de varianza de los datos en función de variables explicativas y permite analizar datos cuya varianza no es homogénea; además explora en gran detalle el comportamiento de la variabilidad de los datos.
- Entre las limitaciones está la dificultad de interpretación de sus resultados, los cuales no siempre son evidentes, especialmente, cuando se utilizan estructuras complejas de variabilidad. Esta dificultad es ampliada por la falta de experiencia del investigador.

- Limitaciones adicionales surgen con el uso de los programas informáticos (MLwiN, HLM, SPSS, SAS, Stata, etc.). Los resultados de Martínez-Murillo (2014) revelan que, aunque en esencia los programas muestran los mismos resultados, existen diferencias en torno a la cantidad de niveles que soportan, el tratamiento de los residuos, o el método de estimación de los elementos de la varianza.
- MLwiN (Rabash et al. 2012) facilita el ajuste de todo tipo de modelo multinivel y minimiza los problemas de estimación. Aunque presenta limitaciones en la gestión de los datos y presentación gráfica, porque tiene su propio formato.
- HLM (Raudenbush, Bryk, Cheong y Congdon, 2000) también facilita el ajuste de todo tipo de modelo multinivel y minimiza los problemas de estimación, y presenta problemas relacionados con la gestión de datos y gráficas, debido al formato específico que utiliza.
- SPSS, SAS y STATA presentan limitaciones en torno al ajuste de los modelos multinivel e interpretación de los resultados; aunque sus ventajas aparecen en la gestión de los datos.

GRACIAS