

Análisis estadístico del rendimiento académico de una asignatura con relación a asignaturas anteriores

Rosalía Peña,
rpr@uah.es
Universidad de Alcalá

Ismael Sánchez
ismael@est-econ.uc3m.es
Universidad Carlos III de Madrid

Resumen

Este artículo analiza el rendimiento de una asignatura en función del rendimiento de otras asignaturas anteriores, mostrando que la probabilidad de cursar con éxito una asignatura disminuye considerablemente si no se ha superado alguna previa.

El conocimiento concreto de qué asignaturas están interrelacionadas resulta relevante para la planificación del alumno. Adicionalmente, garantizar que los discentes de un aula disponen de unos conocimientos mínimos en las materias relevantes puede facilitar la organización del proceso de enseñanza-aprendizaje y por tanto, mejorar el rendimiento académico.

1. Introducción

Son muchas las variables que afectan al aprendizaje y al rendimiento académico de los alumnos. Hay variables de tipo personal, familiar y de contexto. A toda organización docente y en particular a todo profesor, debe interesar conocer cuales son estos factores determinantes y en que medida se puede incidir para hacer más eficaz el proceso de enseñanza.

Si lo que pretende el proceso docente no es solo que el alumno conozca sino que comprenda e integre los nuevos contenidos, para ser capaces de aplicarlos e incluso evaluarlos (véanse las categorías del aprendizaje según la taxonomía de Bloom [3]), se hace inevitable que los nuevos conocimientos se relacionen con los previos del alumno.

Entendemos que entre los factores que afectan al rendimiento académico es fundamental contemplar como variable primaria el

conocimiento previo, es decir las aptitudes previamente adquiridas por el alumno. Específicamente relacionado con las aptitudes viene al caso la afirmación de Coll: "La acción educativa óptima no es nunca en términos absolutos sino en función de las características de los alumnos a las que se dirige" [4].

Dada la relación profesor/alumno existente en el aula universitaria, es necesario, con frecuencia, dirigir la acción docente al colectivo del alumnado, en lugar de al individuo. Es por ello que resulta conveniente que los conocimientos previos, de los alumnos dentro de un aula, sean suficientemente homogéneos.

En este sentido, algunas universidades se plantean trabajar de forma selectiva, haciendo posible la identificación de los alumnos de nuevo ingreso que tengan dificultades y la adopción de estrategias selectivas de apoyo para los mismos [1]. Este tipo de acciones puede ser extraordinariamente costoso, hasta el punto de resultar completamente inviable para la mayoría de las universidades.

Alternativamente, la Universidad ha pretendido conseguir una homogeneización de conocimientos previos, al menos en los cursos no introductorios, con el establecimiento de asignaturas llaves, es decir definiendo prerrequisitos para el acceso a una materia concreta, pero en la última década se aprecia una tendencia a eliminar estas incompatibilidades en los planes de estudio de las universidades públicas, tendencia que no parece reflejada en los de las privadas.

Esta eliminación de las llamadas asignaturas llave no implica que la comunidad universitaria niegue la estrecha relación entre el rendimiento académico entre determinadas asignaturas. Un estudio reciente de Zúñica et al. [6] en la

Universidad Politécnica de Valencia relaciona el rendimiento de la asignatura Tecnología de Computadores (TCO) con el obtenido por los alumnos en otras asignaturas. Se concluye en ese estudio que dicho rendimiento está estrechamente relacionado con el de asignaturas relacionadas con la informática, tales como la Programación (PR) y la Estructura de Computadoras (ESC). El estudio, sin embargo, no muestra una relación directa entre el rendimiento de dichas asignaturas, de forma que el rendimiento de una de ellas pueda verse determinado por las otras. Es precisamente esta determinación, en mayor o menor grado, la que nos interesa analizar en este artículo. Este análisis estaría así en la línea de las recomendaciones del Consejo de Universidades acerca de la realización de estudios estadísticos sobre el rendimiento académico como ayuda para la evolución de los planes de estudio vigentes [2].

Esta relación de determinación entre asignaturas consecutivas puede ser de gran utilidad para profesores y alumnos. Incluso en el caso de no volver a establecer asignaturas llaves como un impedimento, sería muy útil para el alumno conocer la relación real existente entre las asignaturas que evitaría matricularse de asignaturas en las que la probabilidad de aprobar sea muy reducida si antes no ha aprobado otras determinadas.

Asimismo, el profesor que sea consciente de dicha determinación, puede enfocar su método docente para facilitar al alumno la conexión de su asignatura con aquellas asignaturas anteriores que son determinantes del rendimiento académico. De esta manera, ayuda a aumentar la probabilidad de aprobar de aquellos alumnos que no tuvieron buen rendimiento académico en la asignatura previa. De esta forma, la toma de conciencia tanto del profesor como del alumno de dicha relación puede contribuir a la deseada homogeneización del aula.

En este artículo se realizará este análisis para algunas de las asignaturas de titulación de Ingeniería Técnica de Informática de Gestión (ITIG), de la Universidad Carlos III de Madrid (UC3M).

2. Los datos

Los alumnos sobre los que se realizará el análisis son de la titulación de ITIG, de la UC3M en su campus de Leganés. Se dispone de calificaciones de un total de 1032 alumnos que ingresaron en la ITIG de la UC3M entre los cursos académicos 1999/2000 y 2003-2004.

Para la realización del análisis se han considerado las asignaturas obligatorias de primer y segundo curso, cuyo acrónimo se presenta en la Tabla 1.

Para cada asignatura se ha registrado la última calificación de cada alumno. Las calificaciones están comprendidas entre 0 y 10.

| Código | Cuatrim. | Nombre |
|---------------|----------|------------------------|
| Primer curso | | |
| M1 | 1 | Matemáticas 1 |
| FIS | 1 | Física |
| ECO | 1 | Economía |
| PR1 | 1 | Programación 1 |
| LOG | 1 | Lógica de la Progr. |
| Segundo curso | | |
| M2 | 2 | Matemáticas 2 |
| M3 | 2 | Matemáticas 3 |
| PR2 | 2 | Programación 2 |
| ED | 2 | Estructura de Datos |
| TCO | 2 | Tecnol. de Comput. |
| Tercer curso | | |
| MD | 3 | Matemática Discreta |
| IT | 3 | Informática Teórica |
| FBD | 3 | Ficheros y B. de Datos |
| MET | 3 | Metodología |
| ESC | 3 | Estr. de Computadores |
| Cuarto curso | | |
| ES | 4 | Estadística |
| IS | 4 | Ing. de Software |
| SO | 4 | Sistemas Operativos |
| RR | 4 | Redes |
| GE | 4 | Gestión empresarial |

Tabla 1. Asignaturas de primer y segundo curso de ITIG

3. Relación entre Programación I y Estructura de Datos

3.1. Consideraciones generales

En esta sección analizaremos el rendimiento académico de la asignatura Programación I (PR1), que se imparte en el primer curso, primer cuatrimestre, y Estructura de datos (ED), que es de ese mismo curso pero del segundo cuatrimestre. De las calificaciones de las asignaturas puede verse que ED es la asignatura más difícil aprobar. Resulta entonces relevante saber qué asignaturas anteriores pueden determinar en cierta medida el éxito o fracaso en ED.

La normativa actual en la UC3M obliga a los alumnos a matricularse de todas las asignaturas de primer curso. Esta matriculación puede no ser beneficiosa para el alumno, pues como veremos más adelante, el rendimiento académico de las asignaturas del segundo cuatrimestre puede verse condicionado a las del primero. De esta forma, un alumno que no haya superado PR1, por haber suspendido o no haberse presentado, tiene menos probabilidades de superar ED. Esta relación no sólo tiene implicaciones en la gestión de la matriculación, sino en la propia estrategia del proceso de enseñanza-aprendizaje.

Está claro que cuanto mejor sea un estudiante mayor será su rendimiento académico en cualquier asignatura. Por tanto, será obvio que los alumnos que saquen buenas notas en PR1 tenderán a tener también buenas notas en ED sin poder concluir por ello que ambas asignaturas guardan algún tipo de relación.

Por tanto, para poder inferir que existe relación entre el rendimiento académico de dos asignaturas debe detectarse una influencia entre sus rendimientos académicos descontando el componente que sea debido a la capacidad natural del alumno. Para realizar el análisis se construirá, en primer lugar, una medida general del rendimiento académico del alumno, y que denominaremos *Capacidad del alumno*. Esta medida es necesaria para poder interpretar relaciones entre asignaturas.

3.2. Medida de la Capacidad del alumno en el primer cuatrimestre

Existen muchas formas de definir un indicador de la Capacidad del alumno. En este artículo, y con el fin específico de analizar la relación entre PR1 y ED, emplearemos una medida del promedio de las calificaciones obtenidas en las asignaturas del primer cuatrimestre. Esta medida se realizará mediante la técnica estadística de componentes principales (véase, por ejemplo Hair et al. [5]).

La técnica de componentes principales permite transformar un conjunto de K variables relacionadas en otro conjunto de K variables incorreladas. Las nuevas K variables, denominadas componentes principales, son combinaciones lineales de las variables originales. Estas componentes tienen una interpretación en el sentido de dimensiones subyacentes de las K variables originales. Cuando las K variables originales son homogéneas, en el sentido de representar mediciones de magnitudes comparables, la primera componente principal es un promedio de las variables originales, y tiene una interpretación de tamaño. Este promedio es tal que permite una ordenación de los individuos, de menor a mayor valoración, con varianza máxima; es decir, pondera las variables de manera que maximiza la distancia que hay entre los individuos.

Aplicado a nuestro caso, la primera componente principal será una especie de nota media donde aquella asignatura en la que los alumnos hayan tenido calificaciones muy similares (altas o bajas) tendrá una contribución menor en dicha nota media; mientras que aquella asignatura que haya supuesto una mayor dispersión de notas, y por tanto sea un reflejo del trabajo personal de cada alumno, tendrá un peso mayor en dicha nota media.

Por tanto, para medir la Capacidad del alumno utilizaremos la primera componente principal de las calificaciones de las cinco asignaturas del primer cuatrimestre. A la hora de realizar el cálculo del primer componente principal, se plantea la duda de cómo valorar a los alumnos que no se han presentado a alguna asignatura. Para evitar que las conclusiones puedan verse cuestionadas por esta decisión, se tratará a los no

presentados de dos formas alternativas. En una primera opción, eliminaremos a dichos alumnos, de manera que el análisis se realizará sólo entre los 642 alumnos que se han presentado a las 5 asignaturas del primer cuatrimestre. En una segunda opción, asignaremos un 0 como calificación en la asignatura en la que no se ha presentado. El cálculo de los componentes principales se ha realizado con las variables estandarizadas. La Tabla 2 muestra los coeficientes de la primera componente principal en las dos opciones consideradas.

La última fila de esta tabla muestra la proporción de la variabilidad total que está explicada por esta componente. En ambos casos puede verse que en la primera componente todas las asignaturas tienen un peso similar. La asignatura de Economía (ECO) tiene un peso inferior, lo que puede explicarse por la menor tasa de suspensos de esta asignatura. A esta primera componente principal, que utilizaremos como estimador de la Capacidad del alumno, la denominaremos CAPA.

| Pesos de la primera Componente Principal | | |
|--|-------------------|----------------------|
| Asignatura | Todos los alumnos | Sólo los presentados |
| M1 | 0.45 | 0.46 |
| FIS | 0.46 | 0.47 |
| ECO | 0.42 | 0.41 |
| PR1 | 0.46 | 0.45 |
| LOG | 0.46 | 0.44 |
| % varianza | 69% | 62% |

Tabla 2. Primera componente principal de las calificaciones del primer cuatrimestre

3.3. Dependencia entre Programación 1 y Estructura de Datos

Para analizar la dependencia entre PR1 y ED emplearemos una regresión logística (véase, por ejemplo Hair et al. [4]). En la regresión logística, la variable dependiente que queremos explicar es una probabilidad. En nuestro caso es la probabilidad de aprobar ED. Como variables explicativas de dicha probabilidad utilizaremos CAPA así como la calificación del resto de asignaturas de primer cuatrimestre. Los datos relativos a la probabilidad de aprobar ED será una

variable dicotómica que será 1 si el alumno ha aprobado ED y 0 en caso contrario. La denotaremos por AP_ED. La tabla siguiente muestra los resultados de la regresión logística donde la variable dependiente es AP_ED y las variables explicativas son CAPA y la calificación de PR1. En este modelo, la variable CAPA está basada en todos los alumnos, tanto presentados como no presentados (ver Tabla 2).

| Variable dependiente: AP_ED | | |
|-----------------------------|------------|---------|
| Variabes explicativas | Coficiente | p-valor |
| CAPA | 1.55 | 0.000 |
| PR1 | 0.20 | 0.000 |

Tabla 3. Regresión logística con CAPA basado en todos los alumnos

Vemos que tanto las variables CAPA como PR1 son, utilizando un nivel de significación del 5%, bastante significativas. Como era de esperar los coeficientes son positivos. Por tanto, a mayor capacidad del alumno, más probabilidades tiene de aprobar ED. Además, para un mismo valor de CAPA, a mayor calificación de PR1 mayor es la probabilidad de aprobar ED.

La Tabla 4 muestra estos mismos resultados pero utilizando sólo a los alumnos que se han presentado a todas las asignaturas del primer cuatrimestre.

| Variable dependiente: AP_ED | | |
|-----------------------------|------------|---------|
| Variabes explicativas | Coficiente | p-valor |
| CAPA | 1.30 | 0.000 |
| PR1 | 0.24 | 0.002 |

Tabla 4. Regresión logística con CAPA basado en los alumnos presentados

Los resultados utilizando sólo los alumnos presentados son cualitativamente similares. Por tanto, se puede concluir que, independientemente de la capacidad que tenga un alumno, un menor rendimiento en PR1 disminuye la probabilidad de aprobar ED.

Esta relación de dependencia no existe, sin embargo, entre las restantes asignaturas del primer cuatrimestre. Si repetimos el proceso anterior sustituyendo PR1 por el resto de las asignaturas

obtenemos las Tablas 5 a 8. En el caso de utilizar Matemáticas 1 como variable explicativa se obtienen los resultados de la Tabla 5.

| Variable dependiente: AP_ED | | |
|-----------------------------|--------------|---------|
| Variabes explicativas | Coefficiente | p-valor |
| CAPA | 1.54 | 0.000 |
| M1 | -0.005 | 0.956 |

Tabla 5. Regresión logística con CAPA basado en los alumnos presentados

El elevado p-valor refleja que, aparte de la influencia de su rendimiento académico en la asignatura de Matemáticas 1 sobre la medida global de la capacidad del alumno, la dependencia entre el rendimiento de ambas asignaturas no es significativo.

En el caso de la asignatura de Física (Tabla 6) puede verse que el p-valor está por encima del 5%, y además el coeficiente es negativo, por lo que no puede inferirse ninguna relación positiva entre FIS y ED.

Un coeficiente negativo no debe interpretarse que un buen rendimiento académico en Física es contraproducente para aprobar ED. Hay que notar que el modelo contiene la variable CAPA, por lo que la interpretación del efecto de una asignatura es siempre para valores constantes de CAPA. El signo negativo puede interpretarse como que el tipo de alumno que suele tener buenas notas en Física es de un perfil diferente al alumno que obtiene buenas notas en ED.

| Variable dependiente: AP_ED | | |
|-----------------------------|--------------|---------|
| Variabes explicativas | Coefficiente | p-valor |
| CAPA | 1.67 | 0.000 |
| FIS | -0.12 | 0.119 |

Tabla 6. Regresión logística con CAPA basado en los alumnos presentados

Para la asignatura de Economía se tiene:

| Variable dependiente: AP_ED | | |
|-----------------------------|--------------|---------|
| Variabes explicativas | Coefficiente | p-valor |
| CAPA | 1.79 | 0.000 |
| ECO | -0.30 | 0.002 |

Tabla 7. Regresión logística con CAPA basado en los alumnos presentados

que muestra una conclusión similar a la de FIS. Finalmente, para la asignatura de Lógica de la programación se tiene:

| Variable dependiente: AP_ED | | |
|-----------------------------|--------------|---------|
| Variabes explicativas | Coefficiente | p-valor |
| CAPA | 1.43 | 0.000 |
| LOG | 0.12 | 0.179 |

Tabla 8. Regresión logística con CAPA basado en los alumnos presentados

De nuevo, no resulta ser significativa, pese a que inicialmente, el nombre de la asignatura (y su contenido) podría sugerir lo contrario.

4. Relación entre las asignaturas del segundo cuatrimestre e IT

4.1. Consideraciones generales

Nuestro objetivo ahora es aplicar la metodología desarrollada en la sección anterior para encontrar relaciones entre las asignaturas del segundo cuatrimestre y la asignatura de Informática Teórica (IT), que es del tercer cuatrimestre (primer cuatrimestre del segundo curso). IT, junto con FBD, es una de las asignaturas más complicadas del cuatrimestre, por lo que es interesante ver qué asignaturas del cuatrimestre anterior pueden determinar en cierta medida el rendimiento académico en las mismas.

4.2. Relación entre los rendimientos académicos

Se usará la misma metodología expuesta anteriormente. En primer lugar construiremos una medida de la Capacidad del alumno mediante la primera componente principal de las

calificaciones de todas las asignaturas del primer curso (primer y segundo cuatrimestre). A continuación usaremos regresiones logísticas para analizar el rendimiento en IT en función de dicha componente y el rendimiento de cada una de las asignaturas del segundo cuatrimestre.

Para evitar que la acumulación de alumnos no presentados pueda sesgar la interpretación de las calificaciones, la construcción de la variable CAPA, que estima la capacidad del alumno, se hará únicamente entre los alumnos que se han presentado a los exámenes de las asignaturas de primer curso. La variable dependiente es una variable dicotómica que vale 1 si el alumno ha aprobado IT, y 0 en caso contrario. Esta variable se denotará por AP_IT. Al igual que se hizo anteriormente, se estima un modelo logístico para explicar AP_IT con cada una de las asignatura del segundo cuatrimestre, además de con la Capacidad del alumno. La Tabla 9 muestra las estimaciones de los coeficientes de las cinco asignaturas en las cinco regresiones estimadas, así como los respectivos p-valores. Por concisión, no incluimos las estimaciones del coeficiente de Capacidad del alumno.

| Variable dependiente: AP_IT | | |
|-----------------------------|--------------|---------|
| Variables explicativas | Coefficiente | p-valor |
| M2 | -0.07 | 0.455 |
| M3 | 0.10 | 0.338 |
| PR2 | 0.09 | 0.349 |
| ED | 0.28 | 0.000 |
| TCO | -0.39 | 0.000 |

Tabla 9. Resumen de las cinco regresiones logísticas, una por asignatura del segundo cuatrimestre, que explican la probabilidad de aprobar IT, incluyendo además a CAPA.

La Tabla 9 revela que sólo ED tiene una contribución positiva significativa para explicar el aprobado en Informática Teórica, que va más allá de la mera Capacidad general del alumno medida con CAPA. Uniendo este análisis con el de la sección anterior puede construirse lo que podríamos denominar Senda-Crítica, formado por PR1-ED-IT, de manera que el alumno puede ver claramente las repercusiones futuras de su actual rendimiento académico.

5. Conclusiones

Este artículo propone una metodología para analizar la influencia del rendimiento académico de una asignatura en asignaturas posteriores, descontando la influencia de la capacidad general del alumno. Esta metodología se ha aplicado para explicar el rendimiento académico en dos asignaturas de la titulación de Ingeniería Técnica en Informática de Gestión.

El análisis realizado en este artículo corrobora las conclusiones de Zúnica et al. [6] en el sentido de que la capacidad del alumno va a determinar su rendimiento académico futuro. El estudio actual, además, nos aporta una información que nos permite actuar positivamente sobre el proceso de enseñanza-aprendizaje. Hemos visto cómo profundizando en el análisis de los datos es posible obtener información más específica y útil, determinando un pequeño subconjunto de asignaturas que debieran ser superadas por el alumno antes de matricularse de una materia concreta.

La metodología expuesta en este artículo podría extenderse a todas las asignaturas de la titulación, permitiendo trazar un conjunto de sendas críticas que tanto el profesor como el alumno deben tener en consideración para aumentar el rendimiento académico.

Agradecimientos

Queremos agradecer a las autoridades académicas de la Universidad Carlos III de Madrid el acceso a los datos, en especial a Ricardo Camarillo, que pacientemente ha elaborado la base de datos que le solicitamos.

Referencias

- [1] Aglietto, M. T., Foradori, A. D., Zanazzi, J. L. *Un estudio de las condiciones iniciales de los alumnos para la predicción del rendimiento académico*. Universidad Blas Pascal. <http://www.ubp.edu.ar/investigacion/revistas/revista9/art20.html>
- [2] B.O.E. de 17 de enero de 1997.

- [3] B. Bloom, Englehart, E. M. Furst, W. Hill, and D. Krathwohl. Taxonomy of educational objectives: The classification of educational goals. New York: Longmans. 1956.
- [4] Coll, C. *Psicología y Curriculum*. Editorial Piados. 1994. ISBN:950-12-7304-0
- [5] Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W. *Análisis Multivariante*. Editorial Prentice Hall. 1999.
- [6] Zúnica L. Blesa, P., Alcover, R., Más, J., Valiente, J.M.; Estudio del rendimiento académico de una asignatura con relación a asignaturas de cursos anteriores. IX Jornadas de el Enseñanza Universitaria 2003, pag 137.