



ABANREDES: Development of predictive models of academic performance and early detection of dropout: following the footprint of students in educational platforms and social networks



EFECTO DIFERENCIAL DEL CONTEXTO INSTITUCIONAL SOBRE EL RENDIMIENTO ACADÉMICO: UN ENFOQUE MULTINIVEL

Juan Luis Peñaloza Figueroa

Dpto: Economía Financiera y Actuarial y Estadística

Universidad Complutense de Madrid

jluispf@ucm.es

PRESENTACIÓN



- Planteamiento del problema
- Ideas entorno al análisis de RA
- Modelización Multinivel y Metodología
- Discusión de los Resultados
- Conclusiones básicas

PLANTEAMIENTO DEL PROBLEMA

Nuestro planteamiento de investigación:

- 1) ¿Tiene algún efecto moderador el ambiente universitario en el resultado académico de los estudiantes?
- 2) ¿Cuál es su efecto diferencias sobre el RA?
- 3) Explica una proporción significativa de la varianza de los logros académicos promedio de los estudiantes las características de las titulaciones

□ CONTEXTUALIZACIÓN DEL PROBLEMA

- Nuestro planteamiento surge en el marco del proceso de digitalización de las instituciones universitarias.
- Las universidades se enfrentan a la necesidad de una adaptación dinámica y ágil a las nuevas realidades académicas e institucionales, a través de la generación de nuevos contextos de enseñanza y aprendizaje al que denominamos “efecto grado o titulación”.
- Realmente, es posible que NO exista un único contexto de aprendizaje y que los resultados académicos de los estudiantes posiblemente estén influenciados por el “ambiente” de la titulación en la que se matricularon.
- Es un hecho que el sistema universitario europeo atraviesa por cambios significativos en su organización y estructura académica e institucional, como el paso de las Licenciaturas de 5 o 6 años a los Grados académicos de 3 o 4 años, con el objetivo de lograr una homogenización de la educación superior a nivel europeo o la “división” de los masters en una vía profesionalizante y otra de investigación, y la concentración de los doctorados en las denominadas “Escuelas de doctorado”.

RELEVANCIA DEL PROBLEMA

- La percepción que tiene el estudiante de sus éxitos o fracasos está asociada al grado de motivación intrínseca y al clima institucional que vive en su actividad académica.
- La importancia que tienen los descriptores causales como el autoconcepto, la autoestima, la autoeficacia y el autocontrol, que constituyen un cuerpo de conocimientos teóricos y prácticos en los que es posible apoyarse para profundizar y buscar soluciones a los retos y problemas que el sistema universitario enfrenta.

OBJETIVOS e HIPÓTESIS

- El **objetivo general** de esta investigación es destacar la importancia de los aspectos ligados al grado académico, es decir, aquellas actividades que se realizan fuera del aula como las características académicas y sociales de la titulación y/o de las facultades.
- Nuestros **objetivos específicos** están orientados al análisis de las interacciones entre las variables contextuales y las variables que caracterizan a los estudiantes. Entre los objetivos propuestos, destacan la interacción observada entre las diferentes dimensiones del aprendizaje, lo que, a su vez, permite crear un esquema predictivo que tenga en cuenta la interrelación entre los diferentes dimensiones de medición de la actividad académica y de sus resultados.

- El RA posee un carácter complejo y multidimensional en su relación con los diferentes contextos académicos, lo que ha dado lugar a un espacio de discusión sobre los contenidos y las metodologías empleadas para su estudio.
- Se trata de identificar los aspectos que inciden de forma diferenciada sobre las distintas dimensiones del proceso de enseñanza-aprendizaje, así como la naturaleza de la interdependencia entre sus distintos niveles de medición.
- Las investigaciones que estudian los efectos diferenciales de las titulaciones universitarias varían en función de qué variables se consideran como el eje de la investigación. Por ejemplo:
 - Quienes estudian la eficacia diferencial según la situación socioeconómica del estudiante (Thomas, Sammons y Mortimore, 1997).
 - Quienes se decantan por estudiar la eficacia diferencial en función del grupo étnico (Murillo, 2007).
 - Quienes estudian el efecto de la diferencias en las calificaciones de acceso al sistema universitario (Rutter, et al., 1979).

Es más, estos efectos tienen continuidad en el tiempo, lo que refuerza el rol del contexto educativo en el rendimiento académico de los estudiantes universitarios (Sammons et al., 1995).

□ REVISIÓN DE LA LITERATURA ESPECIALIZADA

- En la investigación del sistema universitario se podría requerir medir el rendimiento en las titulaciones que utilizan diferentes métodos o enfrentan diferentes ambientes de aprendizaje. Sería un error analizar esos datos pensando que los estudiantes son muestras aleatorias simples de la población de estudiantes que aprenden bajo un método o ambiente en particular. Los alumnos son agrupados en clases (o cursos), los cuales a su vez son agrupados en Titulaciones o Grados. El desempeño de los estudiantes dentro de una clase están correlacionados, como el desempeño de los estudiantes dentro de la misma titulación. Estas correlaciones deben ser representadas en el análisis para una correcta inferencia e interpretación.
- En las muestras extraídas de la población estudiantil universitaria las observaciones no son independientes (p.e. *los estudiantes de un mismo grado académico tienden a ser similares entre si, porque comparten una historia común*).
- El problema de la dependencia entre observaciones individuales supone la violación del supuesto de independencia de las observaciones en los que se fundamentan los tests estadísticos convencionales (*estimaciones de los errores estándar demasiado pequeños y produce resultados significativamente espurios*).
- Las muestras también presentan variables medidas en todos los niveles posibles y estamos interesados en la relación entre todas esas variables, que presentan correlaciones intra-clase y tamaños muestrales diferentes para las distintas variables medidas.

□ REVISIÓN DE LA LITERATURA ESPECIALIZADA

- Existen muy pocos estudios sobre el efecto diferencial del contexto a nivel universitario. Esto es, la proporción de la varianza de los logros académicos que puede ser explicada, específicamente, por las características de la Titulación o Facultad o Universidad.
- Por ejemplo, a nivel de postgrado, Buela-Casal et al. (2011) en su estudio aplicado a una selección de 1948 doctorandos, identificados en la base TESEO y clasificados según las áreas de conocimiento del Consejo de Coordinación Universitaria de España, mostraron que las áreas con un mejor rendimiento fueron (en ese orden):
 1. Enseñanzas técnicas (49% de éxito) seguida por
 2. Ciencias de la salud (46%)
 3. ciencias experimentales (45,6%)
 4. Mientras que humanidades presenta la menor tasa de éxito (23.3%).

[Las diferencias encontradas fueron estadísticamente significativas (p -valor $< 0,05$ o $0,01$)]

- En síntesis, la búsqueda de variables relevantes que conecten el sistema institucional universitario y el sistema de enseñanza-aprendizaje desarrollado, se ha convertido en una tarea de gran importancia relacionada con los procesos socio-institucionales de carácter universitario, cuyo sentido es mostrar el efecto diferencial de las variables contextuales en la variabilidad del rendimiento académico universitario.

- El marco teórico conceptual que interpreta con mayor coherencia y exhaustividad la interacción entre variables de diferentes niveles y dependencia entre observaciones individuales, y sus efectos sobre el proceso de enseñanza-aprendizaje y sus resultados, es la modelización multinivel.
- Este tipo de modelización da amplitud a lo que Tikunoff (1979) denomina variables contextuales, a partir del entendimiento de que la enseñanza-aprendizaje en el aula está inmersa en una compleja estructura de variables interdependientes, situacionalmente específicas y que condicionan los procesos de aprendizaje y enseñanza.
- La primera aportación de carácter sustantivo de los modelos multinivel es, la de considerar las diferencias en el contexto como un factor relevante, a tener en cuenta, en la evaluación de los resultados académicos de los estudiantes.
- Una segunda aportación sustantiva de los modelos multinivel consiste en analizar, simultáneamente, los contextos y la heterogeneidad entre estudiantes, ya que no sólo se deben considerar las diferencias entre los contextos, sino también las diferencias entre los estudiantes.
- Es más, los modelos multinivel permiten formular y probar hipótesis sobre los efectos cruzados entre niveles (interacciones); por ejemplo, estudiar la relación entre las características del grado y el rendimiento de los alumnos en función de la nota media del curso académico o grupo.

- En el desarrollo de este estudio se utilizó una base de datos que registra la huella académica de los estudiantes y los factores inherentes a la organización académica de la Universidad Complutense de Madrid y de sus Centros adscritos: Villanueva del Campo, CUNEF y Cardenal Cisneros. Los registros fueron examinados y tratados con el software R de análisis estadístico y gráfico, en virtud de su doble naturaleza de programa y lenguaje de programación.
- La base de datos fue extraída del Sistema Integrado de Datos Instituciones (SIDI) de la Universidad Complutense de Madrid, la cual pretende ser la herramienta fundamental de soporte a la planificación estratégica y de información de la UCM y sus centros Adscritos.
- El conjunto de datos originales y anonimizados extraídos corresponde a los diferentes grados de las distintas especialidades de las CCSS. El tamaño inicial de la muestra era de 6763 registros relativos a los grados de Comercio, Sociología, Economía y ADE. La información recopilada abarca el periodo 2015 – 2021 con más de 60 campos o variables por estudiante. No obstante, la base de datos contenía campos vacíos en aquellas variables relacionadas con la calificación de las asignaturas, por lo que se procedió a eliminar los registros que no contaran con todas sus celdas completas.
- El conjunto de datos final, que no incluye valores perdidos o atípicos, fue de 2411 registros, de los cuáles el 47,74% corresponden al sexo femenino y el 52,26% al sexo masculino. Y se incluso en la BD sólo los datos de los estudiantes que han finalizado la carrera.

- Considerando los objetivos de nuestra investigación, se toma la Nota Media como variable dependiente, que representa el promedio de las calificaciones obtenidas por el estudiante en todas las asignaturas desde su matriculación hasta la finalización de la carrera.
- Son consideradas variables predictoras: la Nota de Acceso, que indica la puntuación de acceso a la universidad del estudiante conforme a la Prueba de Acceso a la Universidad (PAU) en el sistema español, la Nota del TFG o la calificación otorgada por el tribunal evaluador del trabajo fin de grado (TFG) como requisito para tener derecho al diploma de graduado, la variable Sexo que toma el valor 0 para las mujeres y 1 para los varones.
- En este trabajo la variable de agrupamiento es la titulación que forma parte del nivel 2 en la estructura jerárquica de los datos, dado que los estudiantes de una misma titulación comparte procesos de enseñanza-aprendizaje y condiciones académicas e institucionales similares.
- Esta variable de agrupación se manifiesta a través de la nota media por titulación y representa el factor de nivel superior que modera el resultado académico de los estudiantes.

La construcción del modelo jerárquico de dos niveles es un proceso en el cual se incluye un modelo nulo (intercepto), un modelo predictivo de variables métricas de primer nivel como la nota de acceso a la universidad, la nota del Trabajo de Fin de Grado (TFG) y una variable dicotómica de género (Hombre, mujer). Y un modelo predictivo de variables de segundo nivel como la calificación media por titulación que mide el “efecto Grado académico” o “efecto institucional” o “efecto contexto”.

La expresión matemática del modelo es:

$$\text{Nota_Media}_i = \beta_{0j} + \beta_{1j}\text{Nota_Acceso}_{ij} + \beta_{2j}\text{Nota_TFG}_{ij} + \beta_{3j}\text{Sexo}_{ij} + \varepsilon_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (3)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + u_{2j} \quad (4)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}Z_j + u_{3j} \quad (5)$$

Las ecuaciones (1), (2), (3) y (4) proporcionan la partición de la varianza en dos niveles: dentro de las titulaciones (ε_{ij}) y entre las titulaciones (u_{0j}). El intercepto β_{0j} está determinado por γ_{00} que representa la nota media de las titulaciones más un efecto aleatorio [u] $_{0j}$. Los coeficientes β_{1j} a β_{3j} son considerados efectos fijos de las variables predictoras asociadas a la nota media final de los estudiantes, y ε_{ij} el error residual para el estudiante i en la titulación j .

- Las ecuaciones 3, 4 y 5 establecen que la relación, descrita por los coeficientes de las pendientes β_{1j} , β_{2j} , β_{3j} , entre el RA y las variables predictores de primer nivel depende del contexto (Z). Por ejemplo, en el caso de β_{1j} , la relación que describe entre el RA y la “Nota de Acceso” depende, por ejemplo, si z es la “Nota media de la titulación”, entonces si γ_{11} es positivo, los efectos de la “Nota de Acceso” sobre el RA será mayor si la Nota media de la titulación es alta (Z), si γ_{11} es negativo, entonces los efectos de la Nota de Acceso sobre el RA será pequeña si la Nota media de la titulación es baja. El razonamiento es similar para los demás coeficientes explicitados en el modelo.
- Por lo tanto, la “Nota media de la titulación” actúa como variable moderadora en la relación entre el RA y las variables predictoras: “Nota de Acceso”, “Nota TFG” y “Sexo”. Esta relación variará en concordancia con el valor de la variable moderadora. Además, observamos que los coeficientes ‘ γ ’ no varían entre grupos (no tienen el subíndice ‘j’) por lo que serán válidos para todos los grupos o clases y, nos referiremos a ellos como coeficientes fijos. Por otro lado, toda la variabilidad no registrada por los β s, incluyendo el efecto de la variable del nivel superior, se asume que corresponde a la variación de los residuos, la cual es capturada por el término de error ‘uj’, el cual incluye el subíndice ‘j’ para indicar a que grupo pertenece dicho residuo..
- Los términos de perturbación u_{0j} , u_{1j} , u_{2j} , u_{3j} , corresponden a los términos de error aleatorios en el nivel de grupo (nivel superior) y, asume que: Los residuos u_{0j} , u_{1j} , u_{2j} u_{3j} tienen media cero y son independientes del error residual ‘eij” que corresponde al nivel de los estudiantes. Las varianzas de los residuos u_{0j} , u_{1j} , u_{2j} , u_{3j} son especificadas como $\sigma_{u_0}^2$, $\sigma_{u_1}^2$, $\sigma_{u_2}^2$ y $\sigma_{u_3}^2$. De manera similar las covarianzas entre los términos de error residual son denotadas como $\sigma_{u_{01}}$, $\sigma_{u_{02}}$, $\sigma_{u_{12}}$ y $\sigma_{u_{13}}$, los cuales, generalmente, no se asumen que sean cero.

- Como primer paso, generamos un modelo nulo o de media no condicionada (no incluye ninguna variable predictora), para examinar la variabilidad de los interceptos entre los grupos (titulaciones). Modelo que se toma como base para la estimación de la varianza explicada, que luego se evalúa a medida que incorporamos las variables predictoras.
- Para ello, debemos determinar la importancia de considerar que los datos puedan estar agrupados o jerarquizados, mediante el Coeficiente de Correlación Intraclase (ICC), Indicador que nos ofrece información sobre la proporción de la varianza total que se debe a las diferencias entre los grupos, que nuestro caso son las titulaciones o grados académicos.

Tabla 2. Coeficiente de Correlación Intraclase (ICC)

```
# Calculamos la ICC a partir del modelo de la media no condicionada o modelo nulo como sigue:  
> Reffects  
      grp          var1 var2          vcov          sdcor  
1 ID_Grado (Intercept) <NA> 0.09405409 0.3066824  
2 Residual             <NA> <NA> 0.47318115 0.6878816  
  
> ICC<-Reffects[1,4]/(Reffects[1,4]+Reffects[2,4])  
[1] 0.1658114
```

- El resultado nos indica que el efecto sobre la varianza total de 16,58% es atribuible a la variación entre las titulaciones o Grados académicos, mientras que el 83,42% es atribuible a la variación dentro de los Grados académicos. Significa que hay una buena porción de la varianza intra-grupo para modelizar la variable respuesta en función de las variables predictoras de interés del primer y segundo nivel.

DISCUSIÓN DE LOS RESULTADOS - II

Tabla 3. Solución del modelo de dos niveles

```
Linear mixed model fit: Nota_Media ~ 1 + Nota_Acceso + Nota_TFG + Sexo + (1 + Nota_mediana_Grupo | ID_Grado), Data: dlevel2

REML criterion at convergence: 4717.8

Scaled residuals:
   Min       1Q   Median       3Q      Max
-2.2976 -0.7462 -0.1480  0.5795  4.3960

Random effects:
   Groups Name              Variance Std.Dev. Corr
   ID_Grado (Intercept)      0.56341  0.7506
   Nota_mediana_Grupo      0.01518  0.1232  -0.95
   Residual                  0.40740  0.6383
Number of obs: 2411, groups: ID_Grado, 4

Fixed effects:
              Estimate Std. Error t value
(Intercept)  5.8497067  0.1437763  40.686
Nota_Acceso  0.0005638  0.0014139   0.399
Nota_TFG     0.1369426  0.0072291  18.943
SexoMUJER    0.1097306  0.0265121   4.139

Correlation of Fixed Effects:
      (Intr) Nt_Acc Nt_TFG
Nota_Acceso -0.077
Nota_TFG    -0.366 -0.023
SexoMUJER   -0.054 -0.070 -0.081

optimizer (nloptwrap) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.00274966 (tol = 0.002, component 1)
Model is nearly unidentifiable: large eigenvalue ratio
- Rescale variables?
```

Nuestros resultados indican que hay una asociación clara entre la nota_mediana de la titulación y el efecto diferencial asociado al rendimiento académico, tanto entre titulaciones y como entre estudiantes

Observamos, en el caso de los efectos fijos, que:

- El valor esperado del efecto del intercepto sobre la Nota_media es de 5,8
- La nota de acceso no es significativa y su efecto sobre el valor esperado de la Nota_media es relativamente pequeña y cercana a cero.
- La nota del TFG si tiene un efecto significativo ($t > \text{valor crítico}$) sobre el valor esperado de la Nota_media con un 0,13
- La variable Sexo también presenta un efecto significativo ($t > \text{valor crítico}$) sobre el valor esperado de la Nota_media, un 0,11 si se trata de una mujer y cero si se trata de un varón.

Con relación a los efectos aleatorios los resultados reportan lo siguiente:

- Hay una diferencia sustancial entre los estudiantes en el valor esperado de la Nota_media global ($sd = 0,75$)
- Si observa una diferencia sustancial entre titulaciones en la relación entre el rendimiento académico y la variable del segundo nivel (Nota_mediana_grupo) ($sd = 0,12$).
- La covarianza entre el intercepto aleatorio y la pendiente aleatoria fue de -0,95 (correlación inversa), lo cual indica que tienen valores de intercepto más altos para la Nota_media, tenían menos probabilidades de mostrar una asociación con la “nota_mediana_grupo”. Esto es, una asociación fuertemente negativa entre el RA y la nota mediana por titulación.

- Los resultados obtenidos respaldan parcialmente nuestras hipótesis de trabajo, y nos dan pistas para ir perfilando y ajustando un modelo de predicción multinivel del rendimiento académico considerando variables accesibles, disponibles y de uso cotidiano, para los profesores.
- La variable contextual considerada en este estudio establece, desde una perspectiva global, relaciones significativas con la variable rendimiento académico, tal y como fue planteado, inicialmente, en nuestras hipótesis. En este sentido, pudo constatarse que estas relaciones no son fijas y permanentes, sino que están determinadas por variables moderadoras específicas (Titulación, Facultad, Área de conocimiento, etc.).
- La medición global del rendimiento académico de los estudiantes universitarios limita la medición de la influencia del profesorado, ya que son diversas las tipologías de profesores que en cada semestre imparten docencia. Las características de los docentes se difuminan en el resultado global.
- Si bien se constata la existencia de inter-relaciones significativas entre las variables que forman parte de los niveles 1 y 2, y que inciden en las predicciones del RA. Al parecer las interacciones son más significativa en las evaluaciones de materias específicas, pero no así en las notas promedio globales por titulación.
- Las diversas estrategias de cambio institucional han hecho posible que la titulación y/o facultad juegue un papel más relevante en el desempeño académico de los estudiantes.



ABANREDES: Development of predictive models of academic performance and early detection of dropout: following the footprint of students in educational platforms and social networks

Financiado por el Ministerio de Ciencia y Tecnología



**CENTRO DE
INTELIGENCIA INSTITUCIONAL**

Por último, queremos dar las gracias al Centro de datos SIDI que es el “Sistema Integrado de Datos Institucionales” de la Universidad Complutense de Madrid y a sus responsables por poner a nuestra disposición los datos relativos a la huella digital de los estudiantes complutense y centros adscritos.



ABANREDES: Development of predictive models of academic performance and early detection of dropout: following the footprint of students in educational platforms and social networks

Financiado por el Ministerio de Ciencia y Tecnología

**¡GRACIAS POR
VUESTRA ATENCIÓN!**