



UNIVERSIDAD DE SEVILLA

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

TESIS DOCTORAL:

**MODELIZACIÓN MULTINIVEL DEL
RENDIMIENTO ACADÉMICO
UNIVERSITARIO**

AUTORA:

Mary Elizabeth Morocho Quezada

DIRECTORES:

Prof. Dr. D. José Luis Pino Mejías

Prof. Dr. D. Juan M. Muñoz Pichardo

Sevilla - España

2015

CERTIFICACIÓN

Prof. Dr. D. José Luis Pino Mejías, Profesor Titular del Departamento de Estadística e Investigación Operativa de la Universidad de Sevilla

Prof. Dr. D. Juan M. Muñoz Pichardo, Catedrático del Departamento de Estadística e Investigación Operativa de la Universidad de Sevilla

Certificamos que la presente memoria titulada:

MODELIZACIÓN MULTINIVEL DEL RENDIMIENTO ACADÉMICO UNIVERSITARIO

Ha sido realizada bajo nuestra dirección por

D. Mary Elizabeth Morocho Quezada

Y constituye su tesis para optar al grado de Doctora

Y para que conste, en cumplimiento de la legalidad vigente y a los efectos que haya lugar, firmamos el presente documento

En Sevilla, 03 de junio del 2015



Prof. Dr. D. José Luis Pino Mejías



Prof. Dr. D. Juan M. Muñoz Pichardo

AGRADECIMIENTO

El camino para alcanzar esta meta ha sido extenso, pero el apoyo incondicional de mis directores de tesis me ha sostenido para llegar a cumplirla. Un especial agradecimiento para los Doctores José Luis Pino Mejías y Juan Manuel Muñoz Pichardo, por su constante aliento, inspiración y paciencia a lo largo de estos años. Gracias por su sabiduría, generosidad y grandes características de seres humanos, me siento honrada por su ayuda.

Al Doctor Pedro Luis Luque Calvo, quien con bondad también compartió su tiempo y conocimientos.

Al Doctor José Barbosa Corbacho, Rector – Canciller de la Universidad Técnica Particular de Loja, quien autorizó el uso oficial de la información de la institución que preside.

Al Doctor Juan Luis Moreno Rebollo, Director del Departamento de Estadística e Investigación Operativa de la Universidad de Sevilla y a su personal administrativo Francisco Méndez y Francisco Valverde por su apoyo en las tareas diarias.

A María Teresa Cáceres, María Dolores Sánchez, María Dolores Cubiles, Rafael Pino, Francisco Solís, Elena Martín, María Teresa Gómez y Manuel Díaz, quienes con su amistad, cariño y atenciones permanentes me hicieron sentir en Sevilla como si estuviese en casa.

A mis padres José y Rosa, a mis hermanos Fernando y Vanessa, por su amor y apoyo infinito, por su compañía a pesar de la distancia, sin ellos no hubiese caminado hasta el final.

Gracias de corazón a todas las personas citadas en esta página y a todas aquellas que en la travesía también contribuyeron para que esta memoria pueda finalizarse.

ÍNDICE

ÍNDICE DE TABLAS	11
ÍNDICE DE GRÁFICOS	13
INTRODUCCIÓN	15
CAPÍTULO 1. RENDIMIENTO ACADÉMICO UNIVERSITARIO	19
1.1. CONCEPTUALIZACIÓN	19
1.2. INDICADORES DEL RENDIMIENTO ACADÉMICO UNIVERSITARIO	22
1.3. FACTORES QUE INCIDEN EN EL RENDIMIENTO ACADÉMICO	34
1.3.1. FACTORES INHERENTES AL ESTUDIANTE	34
1.3.2. FACTORES INHERENTES A LA ORGANIZACIÓN ACADÉMICA UNIVERSITARIA	38
1.3.3. FACTORES INHERENTES AL DOCENTE	40
1.4. MODELOS Y TÉCNICAS ESTADÍSTICAS UTILIZADOS EN LA MEDICIÓN DEL RENDIMIENTO ACADÉMICO	46
CAPÍTULO 2. MODELOS MULTINIVEL	49
2.1. INTRODUCCIÓN A LOS MODELOS MULTINIVEL	49
2.2. MODELOS CON DOS NIVELES: INTERCEPTS ALEATORIOS	52
2.3. MODELOS CON DOS NIVELES: COEFICIENTES ALEATORIOS	57
2.4. INFERENCIA SOBRE EL MODELO	61
2.4.1. ESTIMADORES DE MÁXIMA VEROSIMILITUD	61
2.4.2. ESTIMADORES DE MÁXIMA VEROSIMILITUD RESTRINGIDA	65
2.4.3. DISTRIBUCIONES ASINTÓTICAS DE LOS ESTIMADORES	67
2.4.4. CONTRASTES DE HIPÓTESIS SOBRE LOS PARÁMETROS	67
2.4.5. AJUSTE DEL MODELO	69
CAPÍTULO 3. MODELIZACIÓN DEL RENDIMIENTO ACADÉMICO UNIVERSITARIO	71
3.1. POBLACIÓN	71
3.2. BASES DE DATOS Y TRATAMIENTO DE LA INFORMACIÓN	73
3.2.1. BASE DE DATOS ALUMNOS	73

3.2.2. BASE DE DATOS BECAS	80
3.2.3. BASE DE DATOS PROFESORES	81
3.3. DATA CLEANING: ADAPTACIÓN DE DATOS CORRESPONDIENTES AL PRIMER PERIODO	87
REORDENACIÓN DE LOS DATOS	87
SELECCIÓN DE REGISTROS	88
PROCESO DE UNIÓN DE REGISTROS	88
SELECCIÓN DE REGISTROS	89
PREPARACIÓN DE LAS VARIABLES DE NIVEL 2	89
DETERMINACIÓN VARIABLE BECA	90
ALMACENAMIENTO DEL DATA.FRAME	90
3.4. ANÁLISIS MULTINIVEL DEL PERIODO OCTUBRE/2010 - FEBRERO/2011, INDICADOR NOTA MEDIA PONDERADA	92
3.4.1. PASO 1A: INFLUENCIA DE LA CARRERA COMO EFECTO ALEATORIO	92
3.4.2. PASO 1B: MODELO LINEAL MÚLTIPLE RESPECTO A LAS VARIABLES DE NIVEL 1	94
3.4.3. PASO 2: MODELIZACIÓN DE LA ALEATORIEDAD DEL INTERCEPT	96
3.4.4. PASO 3: MODELIZACIÓN DE LA VARIABILIDAD DEL COEFICIENTE DE EDAD	101
3.4.5. PASO 4: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE SEXO	107
3.4.6. PASO 5: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE MODALIDAD	109
3.4.7. PASO 6: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE SEX*MOD	120
3.4.8. PASO 7: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE HÁBITAT	123
3.4.9. PASO 8: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE LA VARIABLE BECA	125
3.5. RESULTADOS	130
3.5.1. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011	130
3.5.2. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO ABRIL 2011 – AGOSTO 2011	132
3.5.3. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO OCTUBRE 2011 – FEBRERO 2012	134
3.5.4. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO ABRIL 2012 – AGOSTO 2012	135

3.5.5. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011	136
3.5.6. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO ABRIL 2011 – AGOSTO 2011	138
3.5.7. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO OCTUBRE 2011 – FEBRERO 2012	139
3.5.8. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO ABRIL 2012 – AGOSTO 2012	140

CAPÍTULO 4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS **141**

4.1. ANÁLISIS MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011	141
4.2. ANÁLISIS MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011	147
4.3. ANÁLISIS DEL MODELO AJUSTADO POR MEDIA PONDERADA POR PERIODOS	154
4.4. ANÁLISIS DEL MODELO AJUSTADO POR PORCENTAJE DE CRÉDITOS SUPERADOS POR PERIODO	157
4.5. ANÁLISIS COMPARATIVO DE LOS MODELOS AJUSTADOS POR NOTA MEDIA PONDERADA Y PORCENTAJE DE CRÉDITOS SUPERADOS	160

CAPÍTULO 5. DISCUSIÓN, LIMITACIONES, FUTURAS INVESTIGACIONES **163**

5.1. DISCUSIÓN	163
5.1.1. DISCUSIÓN MODELO NOTA MEDIA PONDERADA	163
5.1.2. DISCUSIÓN MODELO PORCENTAJE DE CRÉDITOS APROBADOS	166
5.2. LIMITACIONES	169
5.3. FUTURAS INVESTIGACIONES	170

BIBLIOGRAFÍA **171**

ANEXOS. TRATAMIENTO DE INFORMACIÓN **181**

ANEXO 1. UNIÓN DE ARCHIVOS ALUMNOS	181
ANEXO 2. ÁREA DE FORMACIÓN – TITULACIONES - CÓDIGOS	185
ANEXO 3. CENTROS UNIVERSITARIOS - MODALIDAD	190

ANEXO 4. CIUDAD – PROVINCIA - CANTÓN	191
ANEXO 5. HÁBITAT	193
ANEXO 6. FECHA DE NACIMIENTO	194
ANEXO 7. EDAD	195
ANEXO 8. CRÉDITOS	196
ANEXO 9. NOTA MEDIA PONDERADA – PORCENTAJE DE CRÉDITOS APROBADOS – PORCENTAJE DE ASIGNATURAS	201
ANEXO 10. BECAS	203
ANEXO 11. UNIÓN DE ARCHIVOS ALUMNOS - BECAS	204
ANEXO 12. SELECCIÓN VARIABLES PROFESORES	205
ANEXO 13. CREACIÓN VARIABLES DOCENTES	207
ANEXO 14. UNIÓN BASE PROFESORES – ESTUDIANTES - BECAS	210

ÍNDICE DE TABLAS

Tabla 1. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE AUSTRIA.....	23
Tabla 2. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE ESTADOS UNIDOS DE NORTEAMÉRICA.....	24
Tabla 3. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE CANADÁ	25
Tabla 4. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE ESPAÑA	26
Tabla 5. TABLA VALORATIVA DE BECAS DE FORMACIÓN DE PERSONAL DOCENTE E INVESTIGADOR ESPAÑOL	26
Tabla 6. ESCALA VALORATIVA DE LAS UNIVERSIDADES DEL ECUADOR.....	27
Tabla 7. TABLA DE EQUIVALENCIAS DE LA UNIVERSIDAD DE GRANADA.....	28
Tabla 8. EQUIVALENCIAS POR PERCENTILES	33
Tabla 9. MODELOS Y TÉCNICAS ESTADÍSTICAS.....	47
Tabla 10. NÚMERO DE ESTUDIANTES POR PERIODO ACADÉMICO Y SEXO.....	71
Tabla 11. PORCENTAJE DE ESTUDIANTES POR ÁREA DE FORMACIÓN, MODALIDAD Y SEXO.....	72
Tabla 12. PORCENTAJE DE ESTUDIANTES POR ÁREA DE FORMACIÓN, HÁBITAT Y SEXO	72
Tabla 13. BASE DE DATOS DE ALUMNOS.....	73
Tabla 14. ÁREA ADMINISTRATIVA.....	74
Tabla 15. ÁREA BIOLÓGICA Y BIOMÉDICA.....	75
Tabla 16. ÁREA TÉCNICA	75
Tabla 17. ÁREA SOCIO - HUMANÍSTICA.....	76
Tabla 18. CENTROS UNIVERSITARIOS - HÁBITAT URBANO	77
Tabla 19. CENTROS UNIVERSITARIOS - HÁBITAT RURAL	78
Tabla 20. CENTROS UNIVERSITARIOS - HÁBITAT EXTRANJERO.....	79
Tabla 21. BASE DE DATOS DE BECAS	80
Tabla 22. BASE DE DATOS DE DOCENTES.....	82
Tabla 23. BASE ALUMNOS - BECAS - PROFESORES.....	86
Tabla 24. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA NOTA MEDIA PONDERADA PERIODO OCTUBRE 2010 - FEBRERO 2011	131
Tabla 25. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA NOTA MEDIA PONDERADA PERIODO ABRIL 2011 - AGOSTO 2011.....	133
Tabla 26. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA NOTA MEDIA PONDERADA PERIODO OCTUBRE 2011 - FEBRERO 2012	134
Tabla 27. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA NOTA MEDIA PONDERADA PERIODO ABRIL 2012 - AGOSTO 2012.....	135

Tabla 28. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS PERIODO OCTUBRE 2010 - FEBRERO 2011	137
Tabla 29. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS PERIODO ABRIL 2011 - AGOSTO 2011.....	138
Tabla 30. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS PERIODO OCTUBRE 2011 - FEBRERO 2012	139
Tabla 31. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ : MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS PERIODO ABRIL 2012 - AGOSTO 2012.....	140
Tabla 32. EDAD * ÁREA DE FORMACIÓN	143
Tabla 33. MODALIDAD*SEXO*EDAD MEDIA DOCENTES*PORCENTAJE DOCENTES MUJERES	145
Tabla 34. EDAD*ÁREA DE FORMACIÓN*EDAD MEDIA DOCENTES.....	148
Tabla 35. MODALIDAD * SEXO	150
Tabla 36. ÁREA DE FORMACIÓN * EDAD	151
Tabla 37. EDAD MEDIA DOCENTE * EDAD	152
Tabla 38. MODELO MEDIA PONDERADA POR PERIODOS	155
Tabla 39. MODELO PORCENTAJE DE CRÉDITOS SUPERADOS.....	158

ÍNDICE DE GRÁFICOS

Gráfico 1. EDAD * ÁREA DE FORMACIÓN.....	143
Gráfico 2. EDAD *ÁREA DE FORMACIÓN * EDAD MEDIA	149
Gráfico 3. VARIACIÓN DE CRÉDITOS – 3D.....	153
Gráfico 4. VARIACIÓN DE CRÉDITOS – CONTORNO.....	153

INTRODUCCIÓN

El planteamiento de esta tesis surge en el marco de la realidad de las instituciones de educación superior, quienes enfrentan la necesidad de adaptación en forma dinámica y ágil a nuevas realidades debiendo formar profesionales con un perfil que responda a los procesos de transformación social y a las demandas de la emergente sociedad del conocimiento.

En este diverso entorno el rendimiento académico estudiantil constituye un elemento fundamental para determinar el resultado de una las funciones sustantivas de las universidades, su accionar académico. Sin embargo el éxito académico de los estudiantes es difícil determinarlo, pues en éste concepto se conjugan distintos factores multicausales provocados por variables inherentes al estudiante, a la organización académica universitaria y al profesorado que se ubican en diversos niveles de influencia e interaccionan entre sí.

Ante esta compleja relación se establece la aplicación de una metodología para el análisis de datos con complejos patrones de agrupamiento conocidos como modelos jerárquicos lineales o multinivel, éstos modelos proporcionan un entorno natural en el que se expresa y compara las teorías acerca de las relaciones estructurales entre variables de cada uno de los niveles en una jerarquía organizativa o de muestreo. (Gaviria & Castro, 2005)

Así también permiten la corrección de la subestimación de los errores estándares, el examen de las diferencias entre los diversos grupos, la eliminación del problema de agregación, la estimación de la variabilidad de los coeficientes entre los grupos, mientras más disímiles sean los grupos, mayor sesgo existirá en las estimaciones mínimo cuadráticas y más importante será la aplicación de estos modelos. (Gómez, 2009)

De esta forma y con las ventajas que presentan la utilización de los modelos jerárquicos se plantea en esta investigación la modelización multinivel del rendimiento académico universitario.

El objetivo de estudio es establecer la relación multinivel entre factores inherentes al estudiante, organización académica universitaria y docente con el rendimiento académico universitario medido por la nota media ponderada y porcentaje de créditos aprobados. Las variables relacionadas con la organización académica universitaria y docente se consideran referidas a la titulación (carrera) cursada por el estudiante.

Este objetivo general se concreta en los siguientes objetivos específicos:

Objetivo 1. Investigar la relación multinivel entre los factores inherentes al estudiante (edad, género, becas y hábitat de los estudiantes) y el rendimiento académico estudiantil.

Objetivo 2. Establecer la incidencia multinivel de los factores inherentes a la organización académica universitaria (área académica y tipo de estudios) en el rendimiento académico de los estudiantes.

Objetivo 3. Determinar la influencia multinivel de los factores inherentes al docente (experiencia docente, retribución, edad, sexo) en el rendimiento académico de los estudiantes.

Para alcanzar estos objetivos, la presente investigación se articula en cinco capítulos. En el primer capítulo se expone la conceptualización e indicadores del rendimiento académico universitario, los factores que inciden en el rendimiento universitario, así como los modelos y técnicas estadísticas utilizadas en su medición.

En el segundo se presentan los fundamentos teóricos de los modelos multinivel, introducción a los modelos multinivel, modelos de dos niveles con intercepts aleatorios y modelos de dos niveles con coeficientes aleatorios. Se finaliza con una aproximación general a la inferencia sobre el modelo.

En el tercero se desarrolla la modelización del rendimiento académico universitario. Para ello, se comienza con la determinación de la población de estudio y la descripción y análisis de las bases de datos disponibles. Posteriormente, se procede a desarrollar la acción *data cleaning* y

de adaptación de datos para la aplicación del análisis multinivel. Este proceso se ilustra sólo con los datos correspondientes al primer periodo académico, aunque se ha realizado con los cuatro periodos contenidos en el presente estudio. Finalmente se desarrolla el análisis multinivel del primer periodo con el indicador de rendimiento académico “nota media ponderada”.

Todo lo anterior realizado para el primer periodo académico y el indicador “nota media ponderada” se realiza con el resto de periodos y con otro indicador de rendimiento académico: “porcentaje de créditos superados” por el alumno. No obstante, dado que los procedimientos son totalmente paralelos y similares, sólo se incluye el correspondiente al primer periodo y primer indicador, aunque si se incluyen los resultados de los ocho modelos resultantes al combinar periodo e indicador.

En el cuarto se analizan los resultados obtenidos en el capítulo anterior. El análisis de tales resultados se realiza de forma pormenorizada como se indica a continuación. Primero se examinan los resultados del modelo con variable objetivo “nota media ponderada” en el primer periodo; segundo, se procede de forma similar con el modelo con variable objetivo “porcentaje de créditos superados” en el primer periodo; en tercer lugar se realiza un análisis comparativo de los modelos resultantes para la nota media ponderada en los cuatro periodos considerados; cuarto, se realiza similar análisis comparativo para el indicador “porcentaje de créditos superados”; y finalmente, se realiza un análisis comparativo entre los modelos obtenidos para ambos indicadores.

En el capítulo 5 se realiza una discusión general de los modelos multinivel para el análisis del rendimiento académico universitario, partiendo de los resultados obtenidos en los capítulos anteriores, finalizando con un análisis de las limitaciones de esta aproximación al problema del análisis del rendimiento, así como las futuras líneas de investigación.

Capítulo 1. RENDIMIENTO ACADÉMICO UNIVERSITARIO

1.1. CONCEPTUALIZACIÓN

A lo largo de los años, el estudio del rendimiento académico ha sido una preocupación constante en el campo de la investigación educativa (Ayán & Díaz, 2011), (Porto & Di Gresia, 2004), numerosos estudios han tratado de delimitar y explicar su concepto; sin embargo en todos los casos ha constituido una tarea compleja, dado el carácter multidimensional que da cuerpo a esta variable del área educativa (Montero, Villalobos, & Valverde, 2007).

La medición del rendimiento académico forma parte del análisis de resultados de las actividades universitarias. Entre las actividades que desarrollan las universidades están las de formación, investigación, administración; y extensión o vinculación con la sociedad, que requieren instrumentos específicos para analizar sus resultados.

En la presente memoria nos centramos en los resultados de la actividad académica. Los niveles institucionales de análisis más habituales son el sistema universitario, la institución, el centro, el departamento, la titulación o la materia; y, por otra parte, están las comparaciones centradas en los estudiantes y egresados. El resultado en el que suelen centrarse los análisis es el rendimiento académico del estudiante.

Tradicionalmente el rendimiento ha sido definido como el nivel de aprendizaje alcanzado por el alumno, basado en las calificaciones o notas que obtiene en una o más pruebas que constituyen su evaluación sumativa (García L. , 1988), es una calificación cuantitativa y cualitativa, que si es consistente y válida será reflejo del aprovechamiento por parte del alumnado o del logro de los objetivos preestablecidos (García & García, 2013), hacen referencia a los resultados en un proceso de acción docente y, en un sentido más amplio, a un

proceso educativo (Sarramona, 1993); y, es visto también como un predictor importante del rendimiento en otros niveles de la educación y de otros resultados del trabajo, como el desempeño laboral y el salario (Fenollar, Román, & Cuestas, 2007).

Según Coto (2007), el rendimiento académico es el nivel de conocimientos demostrado por el alumno mediante alguno de los diferentes métodos de evaluación del aprendizaje, en un área o materia en relación con los objetivos propuestos en el proceso de enseñanza-aprendizaje.

Otra definición la presenta Olmos (1997), manifestando que el rendimiento es una variable latente formada por un conjunto de características observables –las calificaciones, entre otras– y algunos otros rasgos que pueden englobarse en el error aleatorio.

Desde una perspectiva de la evaluación de la calidad, el rendimiento académico, es un factor imprescindible y fundamental para la valoración de la calidad educativa en la enseñanza superior pues es uno de los principales indicadores que ofrece un parámetro aceptable de la eficacia y eficiencia de las instituciones de educación superior y del propio sistema educativo. (García L. , 1987), (Rodríguez, Fita, & Torrado, 2004), (Edel, 2003), (McKenzie & Schweitzer, 2001), (de Miguel Díaz *et al.* 2002), (Win & Miller, 2005), (Vargas, 2013). Esta afirmación deriva de la coherencia que debería existir entre las normas académicas y prácticas de evaluación, entre la calidad de la enseñanza y la calidad de las disposiciones para el aprendizaje, entre el costo de la carrera y el potencial de empleabilidad de los graduados; y, entre la reputación de los cursos y programas académicos y los empleadores. (Wimshurst, Wortley, Bates, & Allard, 2006).

Desde un punto de vista práctico, el rendimiento se identifica con resultados inmediatos y diferidos. Los primeros determinados por las calificaciones que obtienen los alumnos durante su carrera hasta la obtención del título correspondiente y se definen en términos de éxito/fracaso en relación a un determinado período temporal; y el rendimiento diferido que tiene conexión con el mundo del trabajo, en términos de eficacia y productividad; y, se vincula con criterios de calidad de una institución. (Tejedor & Muñoz-Repiso, 2007), (Moncada & Rubio, 2011).

Otros autores, señalan que el rendimiento académico es el resultado del aprendizaje, suscitado por la intervención pedagógica del profesor o la profesora, y producido en el alumno. No es el producto analítico de una única aptitud, sino más bien el resultado sintético de una suma (nunca bien conocida) de elementos que actúan en, y desde la persona que aprende, tales como factores institucionales, pedagógicos, psicosociales y sociodemográficos. Tournon (1984) citado en Montero, Villalobos & Valverde (2007).

Vargas (2007), puntualiza que el rendimiento académico es la suma de diferentes y complejos factores que actúan en la persona que aprende, y ha sido definido con un valor atribuido al logro del estudiante en las tareas académicas que se mide mediante las calificaciones obtenidas, con una valoración cuantitativa, cuyos resultados muestran las materias ganadas o perdidas, la deserción y el grado de éxito académico.

Rubio (2009), señala que el rendimiento académico es un fenómeno multifactorial determinado por elementos cognoscitivos, motivacionales, socio-ambientales, y pedagógicos o metodológicos. 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. (Sánchez, Marín, & López, 2011).

Jiménez, Izquierdo & Blanco (2000), Cortés & Lever (2008), manifiestan que el rendimiento académico se concibe como un constructo en el que no sólo se contemplan las pruebas de aptitudes y calificaciones en grados precedentes a la universidad, o la motivación del alumno, sino también otras variables intervinientes como los aspectos docentes, la relación profesor-alumno, relación con los compañeros, aspectos socioeconómicos, entorno familiar, habilidades de estudio, influencia de los padres, contexto socio-histórico, educación de los padres, programa curricular, ambiente escolar, relación con los compañeros.

En este marco de conceptualización, podemos afirmar que el rendimiento académico en la educación superior es un concepto multidimensional que para ser valorado en su conjunto requiere del análisis y evaluación de los factores y elementos que intervienen en el proceso de enseñanza – aprendizaje del estudiante universitario.

1.2. INDICADORES DEL RENDIMIENTO ACADÉMICO UNIVERSITARIO

El rendimiento académico, como se mencionó en el apartado anterior, tradicionalmente se mide en una calificación cuantitativa y/o cualitativa, una nota que, si es consistente y válida, será el reflejo de un determinado aprendizaje, o si se quiere, del logro de los objetivos preestablecidos.

Las calificaciones constituyen en sí mismas el criterio social y legal del logro académico de un alumno en el ámbito universitario, son indicadores que funcionan para establecer los grados de avance académico, pues a través de la valoración cuantitativa, se muestran las materias ganadas o perdidas y la deserción en los diferentes componentes del aprendizaje.

Generalmente, el criterio más utilizado como indicador del rendimiento académico es la nota media del expediente o el porcentaje de créditos superados o matriculados. (de Miguel Díaz *et al.* 2002). Estos indicadores pueden calcularse para los egresados como medida del rendimiento del conjunto de los estudios o para otros niveles de análisis, por ejemplo:

- Asignatura
- Cuatrimestre o semestre
- Curso
- Ciclo

Los estudios de García, Melchor & García (2008); Farsides & Woodfield (2007); Byrne & Flood (2008); Porto & Di Gresia (2004); Chamorro-Premuzic & Furnham (2003); emplean en sus estudios, la nota media por curso o el porcentaje de créditos aprobados como indicadores de rendimiento académico.

En las universidades de habla inglesa el indicador más utilizado es el Grade Point Average (GPA) que es una medida ponderada de las notas, en la que los pesos son los créditos de las materias.

Estudios como los de Richardson, Abraham & Bond (2012); Junco (2012); Sheard (2009); Wimshurst, Wortley, Bates & Allard (2006); Duff *et al.* (2004); McKenzie & Gow (2004);

Lizzio, Wilson & Simons (2002); McKenzie & Schweitzer (2001); Cantwell, Archer & Bourke (2001), utilizan el GPA como medida de los logros académicos.

Existen diversas escalas para el cálculo del GPA, así por ejemplo en las universidades australianas, esta medida, se calcula asignando los valores numéricos a la calificación final de cada curso que se muestran en la Tabla 1.

Tabla 1. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE AUSTRIA

CALIFICACIÓN FINAL	PUNTOS DE CALIFICACIÓN
HD	7
D	6
C	5
P	4
MF	3
F	1.5
LW	-

Fuente: Federation University Australia (2014)

En las Universidades Norteamericanas, el GPA se calcula estableciendo los puntos para la calificación final de la Tabla 2.

Tabla 2. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE ESTADOS UNIDOS DE NORTEAMÉRICA

CALIFICACIÓN FINAL	PUNTOS DE CALIFICACIÓN
A	4.00
A-	3.70
B+	3.33
B	3.00
B-	2.70
C+	2.30
C	2.00
C-	1.70
D+	1.30
D	1.00
D-	0.70

I (Incompleto); y, W (Retiros): No reciben puntos de calificación y no tienen un efecto en el GPA

Fuente: ForeignCredits.Translation and Evaluation, (2013)

En algunas universidades canadienses, el GPA se calcula utilizando la escala de la Tabla 3.

Tabla 3. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE CANADÁ

CALIFICACIÓN FINAL	PUNTOS DE CALIFICACIÓN
A	85 - 100
A-	80 - 84
B+	77 - 79
B	73 - 76
B-	70 - 72
C+	67 - 69
C	63 - 66
C-	60 - 62
D+	57 - 59
D	53 - 56
D-	50 - 52
F	0 - 49

Fuente: University of Toronto Scarborough (2015)

Por otra parte, las universidades españolas en aplicación de lo dispuesto en el Real Decreto 1125/2003, del 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y validez en todo el territorio nacional, califican cada una de las materias del plan de estudios en función de la escala numérica de 0 a 10 puntos descrita en la Tabla 4.

Tabla 4. ESCALA VALORATIVA DE LAS UNIVERSIDADES DE ESPAÑA

CALIFICACIÓN FINAL	PUNTOS DE CALIFICACIÓN
Sobresaliente (SB)	[9 - 10]
Notable (NT)	[7 - 9]
Aprobado (AP)	[5 - 7]
Suspenso (SS)	[0 - 5]

Fuente: Universidad de Sevilla (2010)

Hay otras formas de valorar los resultados, por ejemplo en varias convocatorias de becas de formación de personal docente e investigador español, se utiliza las equivalencias de la Tabla 5.

Tabla 5. TABLA VALORATIVA DE BECAS DE FORMACIÓN DE PERSONAL DOCENTE E INVESTIGADOR ESPAÑOL

CALIFICACIÓN FINAL	PUNTOS DE CALIFICACIÓN
4	[10]
3	[9,10]
2	[7,9]
1	[5,7]
0	[0,5]

Fuente: Ministerio de Educación, Cultura y Deporte (2014)

En Ecuador, las Instituciones de Educación Superior, para medir los logros académicos, utilizan escalas diferentes y para objeto del presente estudio las calificaciones se consignan en la escala numérica de cero a cuarenta puntos por cada semestre, periodo y/o ciclo académico.

La aprobación de cada componente, materia y/o asignatura, se realiza con un puntaje de [28 a 40] puntos, reprobando un componente con un puntaje de [0 a 27] puntos, como se señala en la Tabla 6.

Tabla 6. ESCALA VALORATIVA DE LAS UNIVERSIDADES DEL ECUADOR

DEFINICIÓN	PUNTOS DE CALIFICACIÓN (SOBRE 40 PUNTOS)
Aprobado	28 - 40
Reprobado	0 - 27

Fuente: Universidad Técnica Particular de Loja (2015)

A la vista de los ejemplos anteriores, que evidencian la disparidad de escalas utilizadas, debe tenerse cuidado a la hora de comparar los resultados obtenidos en las investigaciones que tratan, como la presente, de analizar los factores que influyen en el rendimiento académico de los estudiantes universitarios.

Existen dos formas no excluyentes de abordar la equivalencia de calificaciones. La primera es aprobar una tabla que contempla las equivalencias con las calificaciones de una universidad, así por ejemplo la Universidad de Granada utiliza la Tabla 7.

**Tabla 7. TABLA DE EQUIVALENCIAS
UNIVERSIDAD DE GRANADA**

	SUSPENSO	APROBADO	NOTABLE	SOBRESALIENTE	Opción a M.H.
AMÉRICA					
CANADÁ 1	F-E	D-C+	B-B+	A - A	A+
CANADÁ 2	F	E-C	C+ - B+	A - A	A+
EEUU 1	F	D - C-	C - B	B+ - A-	A
EEUU 2	F	C - C+	B - B+	A - A	A+
MÉXICO	Reprobado / No suficiente	Regular / Suficiente	Bien	Muy bien	
	0 - 59 / 0 - 5,9	60 - 79 / 6,0 - 7,9	80 - 89 / 8,0 - 8,9	90 - 100 / 9,0 - 10	
PUERTO RICO	0 - 46 (F)	47 - 59 (D) 60 - 74 (C)	75 - 87 (B)	88 - 100 (A)	
PANAMÁ	0 - 60 (Fracaso)	61 - 70 (Mínima de promoción) 71 - 74 (Regular)	74 - 80 (Regular) 81 - 90 (Bueno)	91 - 99 (Sobresaliente)	100
ARGENTINA	Insuficiente/Reprobado (0 - 3,9) (0 - 3) (0 - 5)	Aprobado (4 - 5,9) Bueno (6 - 6,9) (4 - 5) (6)	Bueno (6 - 7) (7) (7 - 8) Distinguido (7 - 8,9)	Distinguido (9 - 9,9) (8 - 9) (9)	Excelente (10)
BOLIVIA	0 - 50 (Reprobado)	51 - 63 (Suficiente) 64 - 69 (Bueno)	70 - 76 (Bueno) 77 - 89 (Distinguido)	90 - 99 (Sobresaliente)	100
BRASIL	0 - 4,9 0 - 49 F (0 - 59)	5 - 6,9 50 - 69 D (60 - 69)	7 - 7,9 C (70 - 79)	8 - 8,9 80 - 89	9 - 10 90 - 100
CHILE	Insuficiente (0 - 3,9)	Suficiente (4 - 4,9)	Bueno (5 - 5,9)	Muy bueno/ Distinguido (6 - 6,9)	Muy bueno/ Excelente (7)

	SUSPENSO	APROBADO	NOTABLE	SOBRESALIENTE	Opción a M.H.
COLOMBIA	0 - 2,9 (Reprobado)	3,0 - 3,6	3,7 - 4,3	4,4 - 4,7	4,8 - 5
COSTA RICA	0 - 6 0 - 69	7 70 - 79	8 80 - 89	9 90 - 99	10 100
CUBA	Reprobado/Suspense (0-2)	Aprobado (2)	Notable/Aprovechado/ Bien	Sobresaliente/Excel ente	
ECUADOR	0 - 50 0 - 5,9 0 - 23	51 - 69 6 - 7,4 24 - 28	70 - 89 7,5 - 8,9 29 - 34	90 - 99 9 - 9,9 35 - 38	100 10 39 - 40
PARAGUAY	1 (Reprobado/Insuficiente)	2 (Regular/Aprobado)	3 (Bueno)	4 (Distinguido)	5 (Sobresaliente/ Excelente)
PERÚ	0 - 10,9 / 0 - 69	11 - 13,9 / 70 - 79	14 - 16,9 / 80 - 89	17 - 18,9 / 90 - 100	19 - 20
URUGUAY	0 - 2 (Deficiente)	3 - 4 (Regular/Aprobado)	5 - 7 (Bueno)	8 - 10 (Muy bueno)	11 - 12
VENEZUELA	0 - 9,9	10 - 13,9	14 - 16,9	17 - 18,9	19 - 20
ASIA					
CHINA	0 - 59 / 2	60 - 74 / 3	75 - 89 / 4	90 - 99 / 5	100
COREA	0 - 59 / F	60 - 73 / D - C-	74 - 86 / C - B	87 - 96 / B+ - A	97 - 100 / A+
INDIA	0-39	40-59	60-69	70-75	75-100
JAPÓN	0-59 (F)	60-79 (D-C)	80-89 (B)	90-100 (A)	
SINGAPUR	0 - 49 (0.0)	50 - 69 (1.0 - 2.6)	70 - 76 (2.7 - 3.2)	77 - 85 (3.3 - 4.2)	86 - 100 (4.3)
ÁFRICA					
EGIPTO	0 - 49	50 - 68	69 - 88	89 - 97	98 - 100
ISRAEL	0 - 49	50 - 68	69 - 88	89 - 97	98 - 100

	SUSPENSO	APROBADO	NOTABLE	SOBRESALIENTE	Opción a M.H.
JORDANIA	0 - 4,9	50 - 68	69 - 88	89 - 97	98 - 100
MARRUECOS	0 - 9,9	10 - 13,9	14 - 15,9	16 - 18,9	19 - 20
TÚNEZ	0 - 9,9	10 - 13,9	14 - 15,9	16 - 18,9	19 - 20
EUROPA					
ESPAÑA	0-4,9	5-6,9	7-8,9	9-9,9	10
ALEMANIA	5	4 - 3,1	3,0 - 1,6	1,5 - 1,1	1+ / 1,0
AUSTRIA	5	4 - 3,1	3,0 - 1,6	1,5 - 1,1	1+ / 1,0
BÉLGICA	0 - 9,9	10 - 11,9	12 - 14,9	15 - 17,9	18 - 20
BULGARIA	Slab (2)	Sreden (3)	Dobur (4)	MnogoDobur (5)	Otlichen (6)
CHIPRE	0 - 4,5	5 - 6,5	6,5 - 8,5	8,5 - 10	
DINAMARCA	00	02 - 4	7	10	12
ESLOVAQUIA	Nevyhovél (Fail)	Dobry (Good)	Vel'midobry (Very Good)	Vyborny (Excellent)	
ESLOVENIA	Nezadostno/Nezadovoljav a (0-4,9)	Zadostno/Zadovoljava (5-6,9)	Dobro (7-7,9)	Prav Dobro (9-9,9)	Odlicno (10)
ESTONIA	0 / F	Kasin / 1 / E: 51% - 60% Rahuldav / 2 / D: 61%-70%	Hea / 3 / C: 71% - 80%	Vägahea / 4 / B: 81% - 90%	Suurepärane / 5 / A: 91% - 100%
FINLANDIA	Hylatty (0-0,9)	Tyydyttävä (1-1,9) ErittäinTyydyttävä (2-2,5)	ErittäinTyydyttävä (2,6-2,9) Hyvä (3-3,9)	ErittäinHyvä (4-4,9)	Kiitettävä (5)
FRANCIA	0 - 9,9	10 - 11,9	12 - 14,9	15 - 17,9	18 - 20

	SUSPENSO	APROBADO	NOTABLE	SOBRESALIENTE	Opción a M.H.
GRECIA	0 - 4,5	5 - 6,5	6,5 - 8	8,1 - 10	
HUNGRÍA	Elégtelen(1-1,9)	Elégséges(2-2,9)	Közepes(3-3,9)	Jó(4-4,9)	Jeles(5)
IRLANDA	0-39	40-59	60-69	70-75	75 - 100
ISLANDIA	0-4,99	5-5,99	6-7,24	7,25-9,99	10
ITALIA	0-17	18-23	24-27	28-30	30 e lode
LETONIA	Neapmierinosi 0-3,9	GandrizViduveji (4-4,9) Viduveji(5-5,9) GandrizLabi(6-6,9)	Labi(7-7,9) Loti Labi(8-8,9)	Teicami(9-9,9)	Izcili(10)
LITUANIA	1-4,9	5-6,9	7-8,9	9-9,9	10
LUXEMBURGO	0 - 9,9	10 - 13,9	14 - 15,9	16 - 18,9	19 - 20
MALTA	F	D+ - C	C+ - B	B+ - A	A+
NORUEGA	6-4,1	4-2,9	2,8-1,7	1,6-1,1	1,0
PAÍSES BAJOS	0-5,5	5,6-6,9	7-7,9	8-9,9	10
POLONIA	Niedostateczny(0-2)	Dostateczny(2,1-3)	Dobry(3,1-4)	BardzoDobry (4,1-5)	19 - 20
PORTUGAL	0 - 9,9	10 - 13,9	14 - 16,9	17 - 18,9	19 - 20
REINO UNIDO	0-39	40-59	60-69	70-75	75-100
REPÚBLICA CHECA	Nevyhovél (4)	Dobre (3)	Vel'miDobre (2)	Výborne (1)	
RUMANÍA	0-4,9	5-6,9	7-8,9	9-9,9	10
SUECIA	F/U	G	G+	VG - VG+	VG++
SUIZA	0-4 / 0 - 3,9	5 / 4 - 4,4	6 / 4,5 - 5,4	7-9 / 5,5 - 5,9	10 / 6

	SUSPENSO	APROBADO	NOTABLE	SOBRESALIENTE	Opción a M.H.
TURQUÍA	Noksan/PekNoksan(1-4)	Orta (5-6)	Iyi(7-8)	Pekiyi(9-10)	
RUSIA	Neudovletvoritel'no(2)	Udovletvoritel'no(3)	Khorosho (4)	Otlichno(5)	
UCRANIA	nezadovil'no(2)	zadovil'no(3)	dobre(4)	vidminno(5)	
		OCEANÍA			
AUSTRALIA	Failure(0-49)	Conceded, Conditional or Compensatory Pass – Pass (50-64)	Credit(65-74)	Distinction(75 – 84)	High Distinction (85-100)
NUEVA ZELANDA	D y E	C	B	A	A

Fuente: (Universidad de Granada, 2010)

La otra forma de buscar equivalencias es la utilización de percentiles que permite la comparación de notas obtenidas con diferentes sistemas. En el Espacio Europeo de Educación Superior las categorías recomendadas son las establecidas en la Tabla 8.

Tabla 8. EQUIVALENCIAS POR PERCENTILES

CALIFICACIÓN FINAL	CATEGORÍA	PERCENTIL
A	Excelente (EXCELLENT)	[90,100]
B	Muy Bueno (VERY GOOD)	[65,90]
C	Bueno (GOOD)	[35,65]
D	Satisfactorio (SATISFACTORY)	[10,35]
E	Suficiente (SUFFICIENT)	[0,10]

Fuente: ECTS: Guía de uso. European Commission (2009)

La anterior revisión demuestra la importancia de establecer recomendaciones para permitir la comparabilidad de los resultados de los estudios sobre el rendimiento académico universitario.

De hecho dentro de un mismo sistema universitario e incluso en una misma institución la utilización de criterios distintos en la forma en que se codifican los datos académicos, producen resultados no comparables lo que debe tenerse en cuenta a la hora de confrontar los resultados de los diversos estudios sobre rendimiento académico universitario.

1.3. FACTORES QUE INCIDEN EN EL RENDIMIENTO ACADÉMICO

Desde los años 70 hasta la actualidad, varios estudios han examinado los factores que influyen en el rendimiento académico universitario, encontrando dificultad para identificarlos, “pues dichos factores o variables conforman muchas veces una tupida maraña, una red tan fuertemente entretrejida, que resulta ardua la tarea de acotarlas o delimitarlas para atribuir efectos claramente discernibles a cada uno de ellos” Álvaro Page y otros, 1990, citado en (Tejedor & Muñoz-Repiso, 2007).

En la literatura encontramos investigaciones de diversos países, que entre sus objetivos abordan diferentes factores que explican el rendimiento académico en la educación universitaria combinando factores inherentes al estudiante, factores inherentes a la organización académica universitaria y factores inherentes al profesorado.

1.3.1. FACTORES INHERENTES AL ESTUDIANTE

Los factores inherentes al estudiante, que se consideran en el presente estudio son la edad, género, becas y el hábitat de los estudiantes.

EDAD

Numerosas investigaciones han examinado la relación entre la edad y el rendimiento académico, han indagado si el rendimiento aumenta con la edad, disminuye con la edad, o si primero aumenta y a partir de una cierta edad disminuye. (Houston & Rimmer, 2005).

Así, Richardson (1995), en su estudio realizado a 38 estudiantes maduros (mayores a 23 años) y a 60 estudiantes no maduros (18 – 22.9 años) de una Universidad del Reino Unido, encontró que el rendimiento académico de los estudiantes adultos (23,1 – 49.9 años) era por lo menos tan bueno como la de los estudiantes no maduros. Van den Berg & Hofman (2005),

en su estudio realizado a 9000 estudiantes universitarios encuentran el efecto negativo de la edad, estudiantes mayores obtienen menos éxito de estudio que los estudiantes más jóvenes.

En otro estudio, Richardson & Woodley (2003) encuentran que el rendimiento académico de 228.790 estudiantes de pregrado del Reino Unido varía considerando los grupos de edad, y determina que los estudiantes adultos (mayores a 21 años) tienen niveles significativamente más altos de rendimiento que los estudiantes no maduros (menores de 21 años). Así mismo, Sheard (2009), en su estudio efectuado a 134 estudiantes de la Universidad Urbana del Noreste del Reino Unido, sugiere que los estudiantes en edad madura (21 años o más), logran un mayor éxito académico, en comparación con los jóvenes universitarios (menores a 21 años).

De igual forma, Jansen & Bruinsma (2005); van Rooyen, Dixon, Dixon, & Wells (2006); Wojciechowski & Palmer (2005); El Ansari (2002), (2003); Cantwell, Archer & Bourke (2001), Ofori (2000) afirman que los estudiantes maduros tienen mejor rendimiento que los estudiantes jóvenes.

De estos estudios se observa que cada uno de sus autores considera diversos rangos de edad, por lo que en el momento de comparar resulta difícil extraer conclusiones.

GÉNERO

La relación entre el género y el logro académico, ha atraído también la atención a otro grupo de investigadores, así Garkaz, Banimahd & Esmaeili (2011), en su estudio realizado a 450 estudiantes de las carreras de contabilidad de la Universidad Islámica Azad, encuentran que las mujeres muestran mejores resultados que los hombres. De igual forma Barrow, Reilly & Woodfield (2009), encontraron en el estudio de tres cohortes de los estudiantes de pregrado de la Universidad de Sussex que las estudiantes superan académicamente a sus homólogos masculinos.

En igual sentido Castagnetti & Rosti (2009) en el estudio realizado a 26.006 estudiantes italianos y Smith & Naylor (2001) con estudiantes de las universidades del Reino Unido, muestran que las mujeres obtuvieron mejores resultados que los varones; así mismo, Gibb, Fergusson & Horwood (2008), en su estudio realizado a un cohorte de 1.265 individuos estudiados desde el nacimiento hasta los 25 años, también encontraron que había una tendencia generalizada para que las mujeres superen a los hombres en el rendimiento académico; y, Dayioğlu & Türüt-Aşık (2007) han encontrado en una de las universidades públicas más grande de Turquía, que las calificaciones más altas son alcanzadas por las alumnas.

La OCDE (2004) por su parte también señala que en promedio, en la mayoría de sus países miembros, las alumnas superan a los varones en los logros académicos (Castagnetti & Rosti, 2009). Conclusiones similares se establecen en los estudios de Sheard (2009); Mills, Heyworth, Rosenwax, Carr & Rosenberg (2009); Farsides & Woodfield (2007); Woodfield, Jessop & McMillan (2006); Jansen & Bruinsma (2005); El Ansari (2003); Richardson & Woodley (2003); Gracia & Jenkins (2003); Cantwell, Archer & Bourke (2001)

Bermúdez *et al.* (2011), en su estudio a 3.463 doctorandos (1.724 mujeres y 1.739 hombres), comparando el número de tesis doctorales registradas en la base de datos de Tesis Españolas Ordenadas, TESEO del Ministerio de Educación del Gobierno de España entre 2002 y 2009, encontraron que cuando los doctorandos son beneficiarios de una beca predoctoral y en condiciones igualitarias, no existen diferencias de rendimiento en programas de doctorado en función del sexo; y, que no hay diferencias significativas en el rendimiento de hombres y mujeres en cada una de las áreas de conocimiento.

BECAS

La investigación también ha demostrado que otros factores como la situación socioeconómica del estudiante influyen en el rendimiento académico, éste componente conducirá al alumno a la solicitud de ayudas económicas a través de distintos tipos de becas.

En este sentido Considine & Zappalá (2002), en su trabajo realizado a más de 3.000 estudiantes de recursos económicos limitados, presentan que las familias de mayor estatus socioeconómico y educativo reflejan estudiantes con logros académicos superiores.

En los estudios de Sackett *et al.* (2009); Cantwell, Archer & Bourke (2001); White (1982); Lumb & Vail (2004), presentan que el estatus socioeconómico no tiene influencia sobre el rendimiento académico.

HÁBITAT

Otra variable que parece discriminar el logro académico de los estudiantes es el lugar en donde habitan los estudiantes, pues en países en vías de desarrollo, las condiciones geográficas, tecnológicas, sociales y culturales son diferentes en el sector rural y urbano. Estudios específicos que analicen la relación entre el hábitat y el rendimiento académico universitario son escasos; sin embargo se han encontrado investigaciones que examinan cómo puede afectar al rendimiento académico el hecho que los estudiantes no vivan en el lugar en donde se encuentran ubicadas las universidades.

En este contexto, Halpern (2007) manifiesta que mientras es mayor el tiempo que tarda un estudiante para ir y volver de la universidad, peor es su rendimiento académico, acentuándose estos factores en universidades ubicadas en los grandes centros urbanos como Londres, debido al tamaño de la ciudad y el costo de vida. Dutton & Dutton (2005), por su parte en un estudio aplicado a 159 estudiantes que asisten al módulo de Economía y Estadística Empresarial, de la Universidad Estatal de Carolina del Norte en el semestre de otoño de 2001, encontró que los estudiantes que viven a más de diez kilómetros del campus tenían niveles significativamente más bajos de rendimiento académico que los que viven dentro de los diez kilómetros del campus.

Por otra parte Al-Hattami (2012), señala que el logro de los estudiantes de escuelas urbanas es mayor al rendimiento de los estudiantes de escuelas rurales, mientras que Stanley *et al.*

(2008) encontraron que no hubo diferencias significativas en el rendimiento de los estudiantes en las zonas rurales y urbanas.

1.3.2. FACTORES INHERENTES A LA ORGANIZACIÓN ACADÉMICA UNIVERSITARIA

Los factores inherentes a la organización académica universitaria en las que el estudiante desarrolla sus actividades académicas estudiadas son el área de formación en la que la carrera o titulación se encuentra establecida y la modalidad de estudios que desarrolla un estudiante.

ÁREA DE FORMACIÓN Y TITULACIÓN

Otras variables que afectan al rendimiento académico es el área de formación y la titulación en la que se desarrollan las actividades universitarias de los estudiantes; sin embargo son pocos los estudios que analicen estas variables a nivel de pregrado que se han encontrado. Crawford & Wang (2014) en su investigación sobre las diferencias en el rendimiento académico entre estudiantes chinos y británicos en una universidad del Reino Unido a través de dos cohortes de licenciatura, determinan que entre los estudiantes chinos, su marca final de carrera no está influenciada por el programa de grado o titulación. 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 el área con un mejor rendimiento es la de enseñanzas técnicas (49% de éxito) seguida por ciencias de la salud (46%), mientras que humanidades presenta la menor tasa de éxito (23.3%). El área de ciencias experimentales se sitúa en tercer lugar, lo cual tiene una notable importancia ya que era la que presentaba un 45.6% del total de alumnos matriculados, las diferencias encontradas son estadísticamente significativas.

MODALIDAD DE ESTUDIOS

El tipo de estudios que realizan los estudiantes en universidades presenciales y a distancia es otra variable que se incluye en los factores externos que inciden en el rendimiento académico; sin embargo las investigaciones que han analizado la influencia de las dos modalidades en el rendimiento académico de la educación superior es bastante limitado. Duvall & Schwartz (2000), en su estudio analiza los datos de una evaluación continua del programa de educación a distancia para estudiantes de negocios a nivel de postgrado en una universidad privada, abordando el impacto del aprendizaje asistido por la tecnología en el rendimiento académico entre los estudiantes a distancia y sus homólogos en el campus. Los resultados indican, al ajustar por sexo, que no existió diferencias significativas en el rendimiento académico entre los estudiantes a distancia y presenciales mientras que muestra diferencias significativas entre los estudiantes de tecnología y aquellos sin conocimientos tecnológicos.

Se han encontrado también estudios que analizan características propias de los estudiantes en modalidad presencial o a distancia, como la asistencia a clases, compromisos laborales y tiempo de dedicación a los estudios. Autores como Buckalew, Daly & Coffield (1986), en su investigación aplicada a más de 200 estudiantes de psicología de dos universidades, encuentran una fuerte relación entre la asistencia a clases y el rendimiento académico. Woodfield, Jessop & McMillan (2006), en dos investigaciones realizadas en la Universidad de Sussex, presenta una relación positiva entre la asistencia a clases y el resultado final de carrera. Paisey & Paisey (2004) en su estudio aplicado a los alumnos de tercer año del módulo de contabilidad financiera de una universidad escocesa determinan una relación positiva entre la asistencia a clases y el rendimiento académico. Halpern (2007) en su estudio realizado a 127 estudiantes que completaron los cursos de un módulo de gestión de negocios de la Universidad Metropolitana de Londres concluye que existe una correlación positiva significativa entre la asistencia y el rendimiento académico. Stanca (2006), en un estudio realizado a los alumnos de una asignatura presencial de microeconomía, presenta un pequeño pero significativo impacto de las clases presenciales y conferencias en el rendimiento.

En el mismo contexto, la dedicación del tiempo como característica de un estudiante presencial que destina su jornada completa a los estudios; y, de un alumno a distancia que por razones laborales y/o familiares lo realiza únicamente en media jornada, es analizada en relación al rendimiento académico. Autores como Halpern (2007) manifiestan que los estudiantes con compromisos familiares y/o laborales no cuentan con el tiempo necesario para estudiar, por lo que el rendimiento académico de los estudiantes a tiempo parcial se ve afectado negativamente.

Hunt, Lincoln & Walker (2004) realizaron una investigación entre los estudiantes de pregrado de tiempo completo en la Universidad de Northumbria. Su estudio encontró que el impacto negativo del empleo en el logro académico es sustancial y estadísticamente significativo; y, Callender (2008), en su estudio realizado a 1.000 estudiantes de seis universidades del Reino Unido proporciona resultados que el periodo lectivo de trabajo tiene un efecto negativo en el rendimiento académico.

1.3.3. FACTORES INHERENTES AL DOCENTE

Los factores inherentes al docente analizados en esta investigación son experiencia docente, retribuciones, edad y sexo, pues la incidencia de los docentes en el rendimiento académico de los estudiantes, ha sido tema de otro sinnúmero de investigaciones; sin embargo, durante la búsqueda bibliográfica hemos comprobado que son minoría los que abordan la influencia específica sobre el rendimiento universitario. Se han encontrado importantes estudios en otros niveles educativos que destacan la influencia de las características de los profesores en las diferencias del éxito o fracaso del estudiante.

FORMACIÓN

La investigación sobre la influencia del profesor en el rendimiento académico, ha proporcionado resultados mixtos: algunos estudios muestran que si bien la educación adicional de maestros tiene una correlación positiva con el rendimiento de los estudiantes, en

algunos casos, otros encuentran que afecta negativamente el logro estudiantil (Wayne & Youngs, 2003).

Resultados de una asociación positiva entre la formación del docente y el rendimiento académico estudiantil fueron encontrados por Zhang (2008), quien determina en una gran escuela urbana de distrito en el estado de Utah, que un grado avanzado en ciencias o educación de la planta docente, tiene relación significativa y positiva sobre el rendimiento en ciencias de los estudiantes. Similares consecuencias son establecidas por Croninger, Rice, Rathbun & Nishio (2007), en su trabajo basado en datos del Estudio Longitudinal de la Niñez Temprana (ECLS) para analizar la relación entre las calificaciones de los maestros de la escuela primaria y el logro de primer grado en lectura y matemáticas, informando efectos positivos entre la preparación de los profesores y el rendimiento de los profesores.

Goldhaber & Brewer (1996) de igual forma en su investigación realizada en función de los datos del Estudio Longitudinal Nacional de la Educación de 1988 (NELS: 88), utilizado para estimar el impacto de los títulos de los maestros sobre el desempeño de los estudiantes en las áreas de matemáticas, ciencias, inglés e historia; y aplicado a 24000 estudiantes de octavo grado, encontraron que los maestros certificados en matemáticas y los que tienen licenciatura o maestrías en matemáticas y ciencias se asociaron con las puntuaciones más altas de rendimiento de los estudiantes, pero no se encuentran iguales resultados para los profesores de inglés o historia.

Sin embargo, Clotfelter, Ladd & Vigdor (2007), en el estudio realizado a los estudiantes de Carolina del Norte de los grados 3, 4 y 5 en los años 1995-2004, en donde se incluye una variable para indicar si un maestro tiene un título de postgrado de cualquier tipo, tales como el de una maestría, un doctorado o un grado más "avanzado", encuentran que tener un título de posgrado no es un predictivo de un mayor rendimiento en comparación con tener un maestro sin un título de posgrado. Los resultados indican un pequeño efecto positivo de tener un docente con una maestría y un efecto negativo inesperado y sorprendente de tener un profesor con un Ph.D. El último hallazgo se basa en un número muy reducido de profesores y puede decir más acerca de las características de los maestros que pertenecen a esta muestra en particular.

Monk (1994), al utilizar los datos extraídos de la encuesta longitudinal de la juventud americana, determina que el nivel de grado del maestro tiende a una relación negativa con el mejor rendimiento de los alumnos. Jepsen (2005), por su parte en su estudio realizado a más de 200 escuelas y cerca de 10.000 estudiantes, en dos cohortes de primer y tercer grado vinculado a su panel de docentes, encuentra que la formación del profesorado, no tiene un impacto significativo en el rendimiento de los estudiantes. Los resultados de los análisis de Zuzovsky (2003), también encuentra que los grados académicos avanzados, como una especialización en la materia que se enseña el profesor, sólo tenían relaciones positivas estadísticamente no significativas con el rendimiento estudiantil. Clotfelter, Ladd & Vigdor (2006), en su estudio realizado a los estudiantes de quinto grado de Carolina del Norte, indican que los profesores con grados avanzados muestran una tendencia leve pero no significativa en los estudiantes con mejores resultados en los exámenes.

EXPERIENCIA DOCENTE

Estudios recientes informan que en general la experiencia docente tiene un efecto positivo en los resultados académicos de los estudiantes. Croninger, Rice, Rathbun & Nishio (2007), en su trabajo también informan que los maestros que tienen más de 2 años de experiencia en la enseñanza de primer grado se asocian con un mayor rendimiento de los estudiantes en la lectura. Clotfelter, Ladd & Vigdor (2007), en su estudio encuentra evidencia clara de que los maestros con más experiencia son más eficaces para elevar el rendimiento estudiantil que aquellos con menos experiencia. Goldhaber (2007), dentro de su estudio, también señala que los estudiantes de maestros con uno o dos años de experiencia superan a los alumnos de los profesores noveles.

Clotfelter, Ladd & Vigdor (2006), utilizando datos administrativos sobre los estudiantes de quinto grado en Carolina del Norte, en su investigación, muestra que la experiencia del profesor está consistentemente asociada con el rendimiento estudiantil. Jepsen (2005), en su investigación también encontró que para la cohorte de tercer grado, hay algunas pruebas de que la experiencia docente se asocia positivamente con la efectividad del aula. Rockoff (2004), en su estudio realizado a los estudiantes y profesores de dos distritos contiguos de New

Jersey, determina que existe relación positiva entre la experiencia docente y el rendimiento en los exámenes, infiriendo similares resultados, Rivkin, Hanushek & Kain (2005), al utilizar un panel de datos coincidentes únicos del Plan de Escuelas UTD de Texas.

En el mismo contexto, Hanushek (1986), citado en Goldhaber & Anthony (2003), encontró que menos de la mitad de los 109 estudios previos sobre los efectos estimados de la experiencia docente en el rendimiento estudiantil, tuvieron relaciones estadísticamente significativas; de ellos, 33 estudios encontraron que los años adicionales de experiencia tuvieron un efecto positivo significativo, pero siete encontraron que más experiencia realmente tienen un impacto negativo en el rendimiento estudiantil.

A pesar de varios resultados bastante consistentes, no todos los estudios encontraron una asociación entre la experiencia y logros de los estudiantes, Zhang (2008) en los resultados de su estudio presenta que no existen evidencias suficientes para indicar que los años de docencia, influyen fuertemente en el rendimiento en ciencias, así como tampoco del comportamiento en la enseñanza. Zuzovsky (2003), también presenta en su estudio que los años de experiencia en la enseñanza, no tienen relación significativa con el rendimiento estudiantil. Monk (1994), de igual forma, encuentran que la experiencia del maestro tiende a no estar relacionado con el mejor rendimiento académico.

REMUNERACIONES

La investigación ha demostrado claramente que no es una tarea sencilla aislar la influencia de las remuneraciones de los profesores en el logro del estudiante. Goldhaber D. (2006). La idea de desarrollar incentivos de rendimiento más fuertes centradas directamente en los logros de los estudiantes tiene gran atractivo y es un tema de discusión frecuente. Pero los incentivos rara vez se han probado en la práctica y los experimentos con ellos son aún más limitados. Estudios y análisis de los pocos ejemplos disponibles carecen de muchos de los detalles cruciales para juzgar su aplicabilidad más general. Lavy (2002).

Autores como Hanushek (2011), señalan que salarios mucho más altos de los maestros estarían económicamente justificados si los salarios reflejan la efectividad del maestro, así en la década de los años noventa, países como México y Chile ponen en práctica reformas de remuneración docente que las vinculan al logro académico de los alumnos, sin embargo, en ambos casos, el impacto es limitado. Vegas (2006).

Lavy (2002), en su trabajo sobre el efecto causal de un programa en Israel que introdujo incentivos monetarios a las escuelas y a los maestros en función de la mejora relativa de los logros de sus alumnos, determina que los incentivos tienen efectos positivos sobre el rendimiento de los estudiantes. El mismo autor Lavy (2007) al revisar la evidencia sobre los programas de incentivos individuales y escolares implementadas en los últimos años, tanto en Estados Unidos como en el extranjero, observa que las evidencias científicas sugieren, aunque no de forma concluyente, que los incentivos de pago por rendimiento pueden mejorar el desempeño de los docentes, aunque también pueden tener consecuencias imprevistas y no deseadas, tales como que los maestros dirijan sus esfuerzos exclusivamente a las actividades premiadas.

Springer *et al.* (2011), en su trabajo presentan los resultados de un experimento riguroso del impacto de pago por desempeño en el logro de los estudiantes y la práctica docente realizado por el Centro Nacional de Incentivo de rendimiento, en colaboración con la Corporación RAND. El Proyecto de Incentivos en la Enseñanza (PUNTO) se llevó a cabo en las escuelas secundarias del distrito Escuelas Públicas de Metro Nashville (MNPS) de los grados 5-8, durante tres años académicos de 2006-07 hasta 2008-09. Los autores no encuentran diferencias significativas en general entre los estudiantes cuyos maestros fueron asignados al grupo de tratamiento y aquellos cuyos maestros fueron asignados al grupo de control. Además, no hubo diferencias significativas en un año dado, ni tampoco hubo diferencias significativas para los estudiantes en los grados 6-8, cuando se estimaron los efectos independientes para cada nivel de grado.

Figlio & Kenny (2007), en su trabajo documentan sistemáticamente la relación entre los incentivos de desempeño individual de profesores y logros de los estudiantes que se llevan a cabo en los Estados Unidos, contienen datos únicos sobre la frecuencia y magnitud de los

aumentos por mérito y bonos, evaluación de los docentes, y la experiencia del profesorado; en el mismo, encuentran que las puntuaciones de las pruebas son más altas en las escuelas que ofrecen incentivos financieros individuales para un buen rendimiento. Glewwe, Ilias & Kremer (2003), en su estudio informan sobre una evaluación aleatoria de un programa que proporciona a los maestros de primaria, en zonas rurales de Kenia, incentivos basados en resultados de los exámenes de los estudiantes, determinando que los estudiantes de las escuelas del programa tuvieron resultados más altos durante el tiempo que el programa estaba en vigor.

Muralidharan & Sundararaman (2009), presentan los resultados de una evaluación aleatoria de un programa de incentivos para los maestros en práctica a través de una amplia muestra representativa de escuelas primarias rurales administradas por el gobierno en el estado indio de Andhra Pradesh. El programa proporciona pagos de bonos a los maestros basados en el promedio de mejora de resultados de los exámenes de sus alumnos en las evaluaciones del aprendizaje administrados de forma independiente (con una bonificación media de un 3% del sueldo anual). Al cabo de dos años del programa, los alumnos de las escuelas de incentivos se desempeñaron significativamente mejor que los de las escuelas de control.

Fryer (2011), en su trabajo describe un ensayo aleatorio basado en más de doscientas escuelas públicas de la Ciudad de Nueva York, diseñado para comprender mejor el impacto de los incentivos docentes en el rendimiento estudiantil. En el mismo no encontró ninguna evidencia de que los incentivos docentes aumentan el rendimiento del estudiante, en todo caso, los incentivos docentes pueden disminuir el rendimiento estudiantil, especialmente en las escuelas más grandes.

SEXO, EDAD

Durante la búsqueda de información no se han encontrado investigaciones que estudien la influencia conjunta del sexo y edad de los docentes en el rendimiento académico de los estudiantes universitarios, a diferencia de lo que ocurre con los niveles no universitarios y en especial en la enseñanza primaria y primeros años de la enseñanza secundaria, niveles para

los que existe una extensa literatura científica. Los resultados de la presente investigación contribuyen a cubrir esta carencia.

1.4. MODELOS Y TÉCNICAS ESTADÍSTICAS UTILIZADOS EN LA MEDICIÓN DEL RENDIMIENTO ACADÉMICO

Para estudiar la relación entre el rendimiento académico y las variables que influyen en el mismo, los investigadores han utilizado diversos modelos y técnicas estadísticas. En la Tabla 9 se presentan las técnicas utilizadas en una selección de los trabajos revisados:

Capítulo 2. MODELOS MULTINIVEL

2.1. INTRODUCCIÓN A LOS MODELOS MULTINIVEL

En todas las áreas científicas, existen situaciones prácticas en las que los datos obtenidos por la observación directa del fenómeno bajo estudio tienen una estructura jerárquica o agrupada, es decir, una estructura *multinivel*. Los individuos de la población en general se agrupan en estructuras cuyas características pueden influir en las características de los propios individuos. Con objeto de abordar el estudio de este tipo de fenómenos surgieron los modelos multinivel.

La idea básica de estos modelos es recoger la existencia de tales jerarquías de datos permitiendo componentes residuales en cada nivel de la jerarquía. Por ejemplo, un modelo de dos niveles que permite la agrupación de los individuos en grupos incluye la posibilidad de residuos en los dos niveles individuo-grupo. Por lo tanto la varianza residual se divide en una componente entre-grupo (la varianza de los residuos a nivel de grupo) y una componente intra-grupo, dentro del grupo (la varianza de los residuos de nivel secundario). Los residuos del grupo representan características del grupo no observadas que afectan a las características o variables consideradas sobre los individuos.

Situaciones experimentales en las que se aplican estos modelos y que son las más utilizadas para ilustrar el funcionamiento de los mismos son: alumnos agrupados en centros escolares, médicos agrupados en hospitales o centros de salud, etc.

La cuestión clave que define una variable como un *nivel* es que sus unidades pueden considerarse como una muestra aleatoria de una población más amplia de unidades. Por ejemplo, si tenemos una estructura de datos en dos niveles, estudiantes dentro de centros escolares, los estudiantes son una muestra aleatoria de la población de estudiantes, igualmente que los centros escolares incluidos en los datos pueden ser considerados como

una muestra de la población de centros escolares. Algunos autores (Centre for Multilevel Modelling, University of Bristol) aconsejan que se debe tener al menos 20 unidades de nivel superior para que pueda considerarse como nivel.

En general, para la modelización multinivel es necesario determinar adecuadamente el tipo de la variable objetivo, la estructura de datos y la estructura de la varianza (Rasbash *et al.* 2009).

La variable objetivo o respuesta puede ser tanto continua como discreta, aunque el caso más abundante en la literatura científica es una variable respuesta continua normal. Así, se han desarrollado los modelos multinivel para respuestas de tipo Poisson, Bernoulli, categóricas, ordinales, tiempo de supervivencia e incluso respuestas multivariantes (por ejemplo, para medidas repetidas o estudios longitudinales).

Por otra parte, la estructura de datos más simple y más común es la *jerárquica*. En caso de una estructura de dos niveles, cada unidad de nivel inferior se anida dentro de una unidad de nivel superior. Estudios típicos de esta estructura son alumnos-colegios, enfermos-hospitales, etc. No obstante, además de la estructura jerárquica existen otras posibilidades tales como:

- Estructura de clasificación cruzada no jerárquica, en la que los individuos pueden pertenecer a más de una agrupación en un nivel jerárquico dado. Por ejemplo, en un estudio sobre el rendimiento escolar de los alumnos, los centros escolares y el barrio o entorno socio-económico de la residencia del alumno no están anidados, dando lugar a una estructura de clasificación cruzada donde los alumnos se *anidan* dentro de una clasificación cruzada de escuela y vecindario (Rasbash & Goldstein, 1994).
- Estructura no jerárquica de pertenencia múltiple, en la que los individuos de Nivel 1 pueden pertenecer a dos o más unidades de nivel superior. Por ejemplo, en un estudio longitudinal, las unidades de Nivel 1 pueden cambiar en el tiempo su relación de pertenencia con las unidades de Nivel 2 (Browne, Goldstein y Rasbash 2001)

En cuanto a la estructura de la varianza, el modelo multinivel más simple tiene un solo término residual para cada nivel, y es conocido como “modelo de componentes de la varianza”, o bien como “modelo de *intercepts* aleatorios”. Así, se divide la varianza residual en

una componente entre unidades de Nivel 1 y otra componente dentro de las unidades de nivel 2. El modelo supone que sólo el término de intersección (*intercept*) en la ecuación de regresión varía aleatoriamente según las unidades de Nivel 2, mientras que los efectos de las variables explicativas se mantienen fijos en todas las unidades de Nivel 2.

Este último supuesto de que los efectos de las variables explicativas son constantes a través de unidades de nivel superior puede ser poco realista en muchas situaciones. Una modelización más general se consigue con la especificación de dos residuos en el nivel superior, incluyendo interceptos aleatorios y *pendientes aleatorias*. Se trata del modelo conocido como modelo de pendientes aleatorias o de coeficientes aleatorios.

Una generalización del modelo anterior es incluir la posibilidad de heterocedasticidad en las unidades de primer nivel, modelizando dicha heterocedasticidad en términos de alguna variable (cuantitativa o categórica) definida sobre tales unidades. Otras extensiones o generalizaciones de estos modelos se han analizado y han sido incluidas en diversas monografías sobre el tema, entre las que pueden citarse: Goldstein (2011), West, Welch y Galecki (2007) y Smith (2011).

2.2. MODELOS CON DOS NIVELES: INTERCEPTS ALEATORIOS

Considérese una población de unidades de Nivel 1 que se “agrupa” en unidades de Nivel 2. Sobre las primeras se considera una variable respuesta Y que se pretende analizar. El modelo conocido como “modelo de componentes de la varianza” expresa la variable respuesta a través de

- un efecto error aleatorio asociado a la unidad de Nivel 2, y
- un efecto error aleatorio asociado a la unidad de Nivel 1

Así, el modelo puede ser representado por:

$$Y = \gamma_0 + u_0 + e$$

donde

- γ_0 : es un parámetro intercept fijo
- u_0 : error aleatorio asociado a la unidad de Nivel 2, verificando $u_0 \sim N(0, \sigma_u^2)$
- e : error aleatorio asociado a la unidad de Nivel 1, verificando $e \sim N(0, \sigma_e^2)$
- u_0 y e son independientes (o incorreladas)

Dada una muestra aleatoria $\{y_{ij} : i = 1, \dots, n_j, j = 1, \dots, J\}$, el modelo muestral queda representado como sigue:

$$y_{ij} = \gamma_0 + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_u^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{0j}, e_{ij}] = 0 \quad \forall (i, j)$$

El modelo es denominado modelo de componentes de la varianza debido a que la varianza de la variable respuesta se puede descomponer de la forma

$$Var(y_{ij}) = Var(u_{0j} + e_{ij}) = \sigma_u^2 + \sigma_e^2 \quad \forall (i, j)$$

es decir, la suma de las varianzas de Nivel 1 y de Nivel 2. Así, la covarianza entre dos unidades de Nivel 1 pertenecientes a la misma unidad de Nivel 2 es:

$$Cov[y_{ij}, y_{hj}] = Cov[u_{0j} + e_{ij}, u_{0j} + e_{hj}] = \sigma_u^2 \quad \forall i, h = 1 \dots n_j \quad \forall j = 1 \dots J$$

Y la correlación

$$\rho = \rho[y_{ij}, y_{hj}] = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \quad \forall i, h = 1 \dots n_j \quad \forall j = 1 \dots J$$

conocida como correlación intraclases o correlación intra-unidades de Nivel 2, que representa la proporción de varianza total que está “entre” unidades de Nivel 2.

Así, la estructura de varianzas y covarianzas del modelo muestral puede representarse matricialmente en los siguientes términos. Considérese el subvector de observaciones de la variable respuesta en la j-ésima unidad de Nivel 2

$$\mathbf{y}_j = [y_{1j} \dots y_{n_j j}]^T$$

$$Var[\mathbf{y}_j] = \begin{bmatrix} \sigma_u^2 + \sigma_e^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_e^2 & \dots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 + \sigma_e^2 \end{bmatrix} = \sigma_u^2 \mathbf{J}_{(n_j)} + \sigma_e^2 \mathbf{I}_{(n_j)}$$

donde $\mathbf{J}_{(n_j)}$ es la matriz con todos sus elementos la unidad de dimensiones $(n_j \times n_j)$ e $\mathbf{I}_{(n_j)}$ es la matriz identidad de orden n_j . Finalmente, considerando el vector conjunto de las observaciones de la variable respuesta $\mathbf{y} = [\mathbf{y}_1^T \mathbf{y}_2^T \dots \mathbf{y}_J^T]^T$ se tiene que

$$Var[\mathbf{y}] = \begin{bmatrix} \sigma_u^2 \mathbf{J}_{(n_1)} + \sigma_e^2 \mathbf{I}_{(n_1)} & 0 & \dots & 0 \\ 0 & \sigma_u^2 \mathbf{J}_{(n_2)} + \sigma_e^2 \mathbf{I}_{(n_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 \mathbf{J}_{(n_j)} + \sigma_e^2 \mathbf{I}_{(n_j)} \end{bmatrix}$$

es decir, una matriz diagonal por bloques en la que los bloques diagonales tienen todos la misma estructura $(\sigma_u^2 \mathbf{J}_{(n_2)} + \sigma_e^2 \mathbf{I}_{(n_2)})$ y sólo cambian las dimensiones de los mismos en función del tamaño muestral obtenido para cada unidad de Nivel 2.

Obviamente, con esta estructura de varianzas y covarianzas, no es adecuada la aplicación de métodos clásicos de estimación como mínimos cuadrados ordinarios.

El modelo anterior se puede extender a un caso más general conocido como “modelo de intercepts aleatorios” en el que se pretende explicar la variable respuesta a través de la variabilidad de las unidades de Nivel 2 y de un conjunto de variables explicativas o predictivas medidas sobre las unidades de Nivel 1. Considérense las variables X_1, X_2, \dots, X_p como variables predictivas, que pueden representarse de forma vectorial por

$$\mathbf{x}_{ij} = [(x_1)_{ij} \quad (x_2)_{ij} \quad \dots \quad (x_p)_{ij}]^T \text{ variables explicativas en la unidad } (i, j)$$

Así, el modelo puede ser representado por:

$$Y = \gamma_0 + u_0 + \gamma_1 X_1 + \gamma_2 X_2 + \dots + \gamma_p X_p + e$$

donde

- γ_0 : es el parámetro intercepto fijo
- $\gamma_1, \dots, \gamma_p$: son los parámetros fijos de regresión asociados a las variables predictoras
- u_0 : error aleatorio asociado a la unidad de Nivel 2, verificando $u_0 \sim N(0, \sigma_u^2)$
- e : error aleatorio asociado a la unidad de Nivel 1, verificando $e \sim N(0, \sigma_e^2)$
- u_0 y e son independientes (o incorreladas)

Si se denota el vector de coeficientes de regresión por $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_p]^T$ entonces el modelo muestral queda representado como sigue:

$$y_{ij} = \gamma_0 + u_{0j} + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_u^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{0j}, e_{ij}] = 0 \quad \forall (i, j)$$

En consecuencia, mantiene la misma estructura de varianzas y covarianzas que el modelo de componentes de la varianza.

Obsérvese que el modelo es un modelo lineal mixto, en tanto que tiene una parte “fija” y una parte aleatoria, además del error o componente residual asociada a las unidades de nivel 1 o individuos:

- Parte fija: $\gamma_0 + \boldsymbol{\gamma}^T \mathbf{x}_{ij}$
- Parte aleatoria: u_{0j}
- Error asociado a los individuos: e_{ij}

Se denomina modelo de “intercepts aleatorios” dado que se trata de un modelo de regresión múltiple clásico definido sobre grupos de individuos en el que los coeficientes de regresión asociados a las variables explicativas, $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_p]^T$, son constantes para todos los grupos o unidades de Nivel 2, mientras que el intercept ($\gamma_0 + u_{0j}$) tiene asociada la variabilidad inducida por estas unidades de Nivel 2.

El modelo de intercepts aleatorios se puede extender al incluir variables predictivas medidas sobre las unidades de Nivel 2 que traten de explicar la variabilidad de los intercepts. En concreto, los términos que definen los intercepts en el modelo anterior son

$$\gamma_0 + u_{0j} \quad \text{con} \quad u_{0j} \sim N(0, \sigma_u^2) \quad \forall j \text{ e independientes entre sí,}$$

es decir, vienen expresados en términos de un parámetro fijo más un error aleatorio asociado a la unidad de Nivel 2. Si la variabilidad (σ_u^2) de este error es muy grande, se puede tratar de explicar a través de variables medidas sobre estas unidades. Considérense las variables Z_1, Z_2, \dots, Z_q como variables predictivas que pueden representarse de forma vectorial por

$$\mathbf{z}_j = [(z_1)_j \quad (z_2)_j \quad \dots \quad (z_q)_j]^T \text{ variables explicativas en la unidad } j \text{ de Nivel 2}$$

Y el intercept se puede representar por

$$\gamma_0 + u_{0j} = \gamma_{00} + \gamma_{01}(z_1)_j + \dots + \gamma_{0q}(z_q)_j + u_{0j} = \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + u_{0j}$$

con

$$\boldsymbol{\gamma}_0 = [\gamma_{01}, \dots, \gamma_{0q}]^T$$

$$u_{0j} \sim N(0, \sigma_u^2) \quad \forall j \text{ e independientes entre sí,}$$

Si expresamos conjuntamente todo el modelo, se tiene

$$y_{ij} = \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + u_{0j} + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + e_{ij} = \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + u_{0j} + e_{ij}$$

$$y_{ij} = \gamma_{00} + \sum_{h=1}^q \gamma_{0h} (z_h)_j + \sum_{k=1}^p \gamma_k (x_k)_{ij} + u_{0j} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_u^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$\text{Cov}[u_{0j}, e_{ij}] = 0 \quad \forall (i, j)$$

Es decir, el modelo se compone de

- Parte fija: $\gamma_{00} + \boldsymbol{\gamma}_0^\top \mathbf{z}_j + \boldsymbol{\gamma}^\top \mathbf{x}_{ij}$. Esta parte contiene tanto los efectos de las variables predictivas medidas sobre unidades de Nivel 1 ($\boldsymbol{\gamma}^\top \mathbf{x}_{ij}$), como los efectos de las variables predictivas medidas sobre las unidades de Nivel 2 ($\boldsymbol{\gamma}_0^\top \mathbf{z}_j$).
- Parte aleatoria: u_{0j}
- Error asociado a los individuos: e_{ij}

En consecuencia, mantiene la misma estructura de varianzas y covarianzas que el modelo de componentes de la varianza.

2.3. MODELOS CON DOS NIVELES: COEFICIENTES ALEATORIOS

Una extensión del modelo anterior se obtiene al considerar que existen algunos coeficientes de regresión que pueden “variar” en función de las unidades de Nivel 2. Así, se puede plantear la cuestión: ¿Los coeficientes del modelo asociados a las variables predictivas medidas sobre las unidades de Nivel 1 “varían” en función de las unidades de Nivel 2?

Para resolver esta cuestión el modelo anterior se ha de extender al conocido como “modelo de coeficientes aleatorios”.

Como caso particular, considérese el modelo anterior con coeficiente aleatorio asociado a la variable X_1

$$y_{ij} = \gamma_{00} + \sum_{h=1}^q \gamma_{0h}(z_h)_j + u_{0j} + [\gamma_1 + u_{1j}](x_1)_{ij} + \sum_{k=2}^p \gamma_k (x_k)_{ij} + e_{ij}$$

$$\begin{aligned} y_{ij} &= \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + u_{0j} + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + u_{1j} (x_1)_{ij} + e_{ij} \\ &= \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + u_{0j} + u_{1j} (x_1)_{ij} + e_{ij} \end{aligned}$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u_1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$Cov(u_{0j}, u_{1j}) = \sigma_{01} \quad \forall j$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

Para este modelo, la varianza de la variable respuesta se puede descomponer de la forma

$$\begin{aligned} Var[y_{ij}] &= Var[u_{0j} + u_{1j} (x_1)_{ij} + e_{ij}] = \\ &= Var[u_{0j}] + (x_1)_{ij}^2 Var[u_{1j}] + Var[e_{ij}] + 2Cov(u_{0j}, u_{1j}) \\ &= \sigma_{u_0}^2 + (x_1)_{ij}^2 \sigma_{u_1}^2 + 2(x_1)_{ij} \sigma_{01} + \sigma_e^2 \end{aligned}$$

Y la covarianza entre dos unidades de Nivel 1 pertenecientes a la misma unidad de Nivel 2 es:

$$\begin{aligned} Cov[y_{ij}, y_{hj}] &= Cov[u_{0j} + u_{1j}(x_1)_{ij} + e_{ij}, u_{0j} + u_{1j}(x_1)_{hj} + e_{hj}] \\ &= \sigma_{u_0}^2 + [(x_1)_{ij} + (x_1)_{hj}] \sigma_{01} + [(x_1)_{ij} \cdot (x_1)_{hj}] \sigma_{u_1}^2 \end{aligned}$$

$$\forall i, h = 1 \dots n_j \quad \forall j = 1 \dots J$$

En consecuencia, si consideramos:

$$\mathbf{y}_j = [y_{1j} \dots y_{n_j j}]^T ; \mathbf{X}_{(1)j} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ (x_1)_{1j} & (x_1)_{2j} & \dots & (x_1)_{n_j j} \end{bmatrix}^T ; \mathbf{\Omega} = \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{01} \\ \sigma_{01} & \sigma_{u_1}^2 \end{bmatrix}$$

entonces

$$Var[\mathbf{y}_j] = \mathbf{X}_{(1)j} \mathbf{\Omega} \mathbf{X}_{(1)j}^T + \sigma_e^2 \mathbf{I}_{(n_j)}$$

De forma análoga, este modelo se puede extender al incluir variables predictivas medidas sobre las unidades de Nivel 2 que traten de explicar la variabilidad de los coeficientes aleatorios. En concreto, los términos que definen los coeficientes asociados a la primera variable predictiva X_1 en el modelo anterior son

$$[\gamma_1 + u_{1j}] \quad \text{con} \quad u_{1j} \sim N(0, \sigma_{u_1}^2) \quad \forall j \text{ e independientes entre sí,}$$

es decir, vienen expresados en términos de un parámetro fijo más un error aleatorio asociado a la unidad de Nivel 2. Si la variabilidad ($\sigma_{u_1}^2$) de este error es muy grande, se puede tratar de explicar a través de variables medidas sobre estas unidades. Considérense las variables Z_1, Z_2, \dots, Z_q como variables predictivas que pueden representarse de forma vectorial por

$$\mathbf{z}_j = [(z_1)_j \quad (z_2)_j \quad (z_q)_j]^T \text{ variables explicativas en la unidad } j \text{ de Nivel 2}$$

Y el coeficiente asociado a la variable predictiva X_1 se puede representar por

$$\gamma_1 + \gamma_{11}(z_1)_j + \dots + \gamma_{1q}(z_{1q})_j + u_{1j} = \gamma_{10} + \sum_{h=1}^q \gamma_{1h}(z_h)_j + u_{1j} = \gamma_1 + \boldsymbol{\gamma}_1^T \mathbf{z}_j + u_{1j}$$

con

$$\boldsymbol{\gamma}_1 = [\gamma_{11}, \dots, \gamma_{1q}]^T$$

$$u_{1j} \sim N(0, \sigma_{u_1}^2) \quad \forall j \text{ e independientes entre sí,}$$

Si expresamos conjuntamente todo el modelo, se tiene

$$y_{ij} = \gamma_{00} + \sum_{h=1}^q \gamma_{0h} (z_h)_j + u_{0j} + [\gamma_1 + \sum_{h=1}^q \gamma_{1h} (z_h)_j + u_{1j}] (x_1)_{ij} + \sum_{k=2}^p \gamma_k (x_k)_{ij} + e_{ij} = \\ \gamma_{00} + \sum_{h=1}^q \gamma_{0h} (z_h)_j + u_{0j} + \sum_{h=1}^q \gamma_{1h} (z_h)_j (x_1)_{ij} + \sum_{k=1}^p \gamma_k (x_k)_{ij} + (x_1)_{ij} u_{1j} + e_{ij}$$

O bien, en términos vectoriales

$$y_{ij} = \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + u_{0j} + \boldsymbol{\gamma}_1^T (\mathbf{z}\mathbf{x}_1)_{ij} + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + (x_1)_{ij} u_{1j} + e_{ij}$$

donde

$$\boldsymbol{\gamma}_1 = [\gamma_{11}, \dots, \gamma_{1q}]^T \quad \text{y} \quad (\mathbf{z}\mathbf{x}_1)_{ij} = \left[(z_1)_j (x_1)_{ij} \quad (z_2)_j (x_1)_{ij} \quad \dots \quad (z_q)_j (x_1)_{ij} \right]^T$$

Es decir, se tiene un modelo lineal mixto con las componentes:

- Parte fija: $\gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + \boldsymbol{\gamma}_1^T (\mathbf{z}\mathbf{x}_1)_{ij} + \boldsymbol{\gamma}^T \mathbf{x}_{ij}$. Esta parte contiene tanto los efectos de las variables predictivas medidas sobre unidades de Nivel 1 ($\boldsymbol{\gamma}^T \mathbf{x}_{ij}$), como los efectos de las variables predictivas medidas sobre las unidades de Nivel 2 ($\boldsymbol{\gamma}_0^T \mathbf{z}_j$), así como los efectos de la interacción entre las variables predictivas medidas sobre las unidades de Nivel 2 con la variable predictiva X_1
- Parte aleatoria: $u_{0j} + (x_1)_{ij} u_{1j}$
- Error asociado a los individuos: e_{ij}

En consecuencia, mantiene la misma estructura de varianzas y covarianzas que el modelo anterior.

El modelo puede extenderse incluyendo efectos aleatorios en todas y cada una de los coeficientes asociados a las restantes variables predictivas medidas sobre unidades de Nivel 1, es decir, X_2, X_3, \dots, X_p . Ello supone incluir las interacciones de las mismas con respecto a las variables predictivas medidas sobre unidades de Nivel 2 (Z_1, Z_2, \dots, Z_q) en la parte fija del

modelo y añadir en la parte aleatoria los términos $(x_k)_{ij} u_{kj}$, para $k = 2, \dots, p$, imponiendo las hipótesis

$$u_{kj} \sim N(0, \sigma_{uk}^2) \quad \forall j = 1, 2, \dots, q \text{ e independientes entre sí}$$

$$\text{Cov}(u_{kj}, u_{lj}) = \sigma_{kl} \quad \forall j = 1, 2, \dots, q; \forall k, l = 0, 1, 2, \dots, p$$

2.4. INFERENCIA SOBRE EL MODELO

2.4.1. ESTIMADORES DE MÁXIMA VEROSIMILITUD

Como se ha recogido anteriormente, el modelo se puede representar como la suma de una parte que contiene efectos fijos y otra que contiene los efectos aleatorios. En concreto, un modelo con efectos aleatorios en el intercept o término independiente y en el coeficiente de una de las variables explicativas, queda expresado vectorialmente de la forma:

$$y_{ij} = \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + u_{0j} + \boldsymbol{\gamma}_1^T (\mathbf{z}\mathbf{x}_1)_{ij} + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + (x_1)_{ij} u_{1j} + e_{ij} \quad i = 1, \dots, n_j; j = 1, \dots, J,$$

Y reordenando las componentes:

$$y_{ij} = \gamma_{00} + \boldsymbol{\gamma}_0^T \mathbf{z}_j + \boldsymbol{\gamma}_1^T (\mathbf{z}\mathbf{x}_1)_{ij} + \boldsymbol{\gamma}^T \mathbf{x}_{ij} + u_{0j} + (x_1)_{ij} u_{1j} + e_{ij} \quad i = 1, \dots, n_j; j = 1, \dots, J,$$

De forma conjunta, se puede expresar

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\gamma} + \mathbf{W}_j \mathbf{u}_j + \mathbf{e}_j \quad j = 1, \dots, J$$

con

- \mathbf{y}_j n_j -vector de observaciones en la unidad j de Nivel 2,
- \mathbf{X}_j $(n_j \times p)$ -matriz de observaciones de variables explicativas (y/o productos de ellas), también conocida como matriz de diseño de los efectos fijos (contiene una primera columna con todos sus elementos unitarios),
- $\boldsymbol{\gamma}$ p -vector de parámetros que representan los efectos fijos
- \mathbf{W}_j $(n_j \times q)$ -matriz de observaciones de variables explicativas asociadas a efectos aleatorios
- \mathbf{u}_j q -vector de efectos aleatorios tales que $\mathbf{u}_j \sim N_q[\mathbf{0}, \boldsymbol{\Sigma}_u]$
- \mathbf{e}_j n_j -vector de residuos aleatorios tales que $\mathbf{e}_j \sim N_{n_j}[\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_j}]$

Esta expresión también permite representar otros modelos con coeficientes aleatorios en diversas variables explicativas. Por otra parte, la matriz Σ_u es constante para cada una de las J unidades de Nivel 2, las cuales son observadas de forma independiente, por lo que

$$Cov(\mathbf{u}_j, \mathbf{u}_h) = \mathbf{0}, Cov(\mathbf{e}_j, \mathbf{e}_h) = \mathbf{0}, Cov(\mathbf{u}_j, \mathbf{e}_h) = \mathbf{0}$$

Para facilitar el procedimiento de estimación se considera la descomposición de Cholesky de dicha matriz:

$$\Sigma_u^{-1} = \sigma_e^2 \Psi^T \Psi, \text{ con } \Psi \text{ matriz triangular}$$

La matriz Ψ se conoce como “factor de precisión relativa” (Pinheiro y Bates 2000). Así, la función de verosimilitud del modelo viene dada por:

$$L(\boldsymbol{\gamma}, \Psi, \sigma_e^2; \mathbf{y}) = \prod_{j=1}^J f(\mathbf{y}_j; \boldsymbol{\gamma}, \Psi, \sigma_e^2) = \prod_{j=1}^J \int f_{Y/u_j}(\mathbf{y}_j; \boldsymbol{\gamma}, \Psi, \sigma_e^2) f_u(\mathbf{u}_j; \Psi, \sigma_e^2) d\mathbf{u}_j$$

Dado que $\mathbf{y}_{j/u_j} \sim N_{n_j}[\mathbf{X}_j \boldsymbol{\gamma} + \mathbf{W}_j \mathbf{u}_j, \sigma_e^2 \mathbf{I}_{n_j}]$ y $\mathbf{u}_j \sim N_q[\mathbf{0}, \Sigma_u]$ entonces

$$\begin{aligned} L(\boldsymbol{\gamma}, \Psi, \sigma_e^2; \mathbf{y}) &= \prod_{j=1}^J \frac{|\Psi|}{(2\pi\sigma_e^2)^{n_j/2}} \int \frac{1}{(2\pi\sigma_e^2)^{q/2}} \exp\left\{\frac{-1}{2\sigma_e^2} \left[\|\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\gamma} + \mathbf{W}_j \mathbf{u}_j\|^2 \right. \right. \\ &\quad \left. \left. + \|\Psi \mathbf{u}_j\|^2 \right] \right\} d\mathbf{u}_j \\ &= \prod_{j=1}^J \frac{|\Psi|}{(2\pi\sigma_e^2)^{n_j/2}} \int \frac{1}{(2\pi\sigma_e^2)^{q/2}} \exp\left\{\frac{-1}{2\sigma_e^2} \|\tilde{\mathbf{y}}_j - \tilde{\mathbf{X}}_j \boldsymbol{\gamma} + \tilde{\mathbf{W}}_j \mathbf{u}_j\|^2\right\} d\mathbf{u}_j \end{aligned}$$

siendo

$$\tilde{\mathbf{y}}_j = \begin{bmatrix} \mathbf{y}_j \\ \mathbf{0} \end{bmatrix}, \tilde{\mathbf{X}}_j = \begin{bmatrix} \mathbf{X}_j \\ \mathbf{0} \end{bmatrix} \text{ y } \tilde{\mathbf{W}}_j = \begin{bmatrix} \mathbf{W}_j \\ \Psi \end{bmatrix}$$

Considérese ahora la descomposición matricial conocida como “ortogonal-triangular”, según la cual dada una matriz \mathbf{X} de dimensiones $(n \times p)$ de rango p , existe una $(n \times n)$ -matriz ortogonal \mathbf{Q} y una $(p \times p)$ -matriz triangular superior \mathbf{R} tales que

$$\mathbf{X} = \mathbf{Q} \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

Esta descomposición permite expresar el problema de mínimos cuadrados ordinarios de forma fácil, dado que, al ser \mathbf{Q} ortogonal

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}\|^2 = \|\mathbf{Q}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\gamma})\|^2 = \left\| \mathbf{Q}^T\mathbf{y} - \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \boldsymbol{\gamma} \right\|^2 = \|\mathbf{c}_1 - \mathbf{R}\boldsymbol{\gamma}\|^2 + \|\mathbf{c}_2\|^2$$

siendo \mathbf{c}_1 y \mathbf{c}_2 vectores de dimensiones p y $(n - p)$, respectivamente, tales que

$$\mathbf{c} = \mathbf{Q}^T\mathbf{y} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$$

Así

$$\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}\|^2 = \|\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\gamma}}\|^2 : \quad \mathbf{R}\tilde{\boldsymbol{\gamma}} = \mathbf{c}_1$$

y la suma de cuadrados residual es $\|\mathbf{c}_2\|^2$.

Aplicando esta descomposición a la matriz $\tilde{\mathbf{W}}_j$

$$\tilde{\mathbf{W}}_j = \mathbf{Q}_{(j)} \begin{bmatrix} \mathbf{R}_{11(j)} \\ \mathbf{0} \end{bmatrix}$$

se tiene que

$$\begin{aligned} \|\tilde{\mathbf{y}}_j - \tilde{\mathbf{X}}_j \boldsymbol{\gamma} + \tilde{\mathbf{W}}_j \mathbf{u}_j\|^2 &= \|\mathbf{Q}_{(j)}^T[\tilde{\mathbf{y}}_j - \tilde{\mathbf{X}}_j \boldsymbol{\gamma} + \tilde{\mathbf{W}}_j \mathbf{u}_j]\|^2 \\ &= \|\mathbf{c}_{1(j)} - \mathbf{R}_{10(j)}\boldsymbol{\gamma} - \mathbf{R}_{11(j)}\mathbf{u}_j\|^2 + \|\mathbf{c}_{0(j)} - \mathbf{R}_{00(j)}\boldsymbol{\gamma}\|^2 \end{aligned}$$

donde $\mathbf{R}_{10(j)}$ y $\mathbf{R}_{00(j)}$ son matrices de dimensiones $(q \times p)$ y $(n_j \times p)$, respectivamente, tales que

$$\mathbf{Q}_{(j)}^T \tilde{\mathbf{X}}_j = \begin{bmatrix} \mathbf{R}_{10(j)} \\ \mathbf{R}_{00(j)} \end{bmatrix},$$

y los vectores $\mathbf{c}_{1(j)}$ y $\mathbf{c}_{0(j)}$, de dimensiones q y n_j , respectivamente, tales que

$$\mathbf{Q}_{(j)}^T \tilde{\mathbf{y}}_j = \begin{bmatrix} \mathbf{c}_{1(j)} \\ \mathbf{c}_{0(j)} \end{bmatrix}.$$

Considerando nuevamente la integral contenida en la expresión de la verosimilitud:

$$\int \frac{1}{(2\pi\sigma_e^2)^{\frac{q}{2}}} \exp\left\{\frac{-1}{2\sigma_e^2} \|\tilde{\mathbf{y}}_j - \tilde{\mathbf{X}}_j \boldsymbol{\gamma} + \tilde{\mathbf{W}}_j \mathbf{u}_j\|^2\right\} d\mathbf{u}_j =$$

$$\exp\left\{\frac{-1}{2\sigma_e^2} \|\mathbf{c}_{0(j)} - \mathbf{R}_{00(j)}\boldsymbol{\gamma}\|^2\right\} \int \frac{1}{(2\pi\sigma_e^2)^{\frac{q}{2}}} \exp\left\{\frac{-1}{2\sigma_e^2} \|\mathbf{c}_{1(j)} - \mathbf{R}_{10(j)}\boldsymbol{\gamma} - \mathbf{R}_{11(j)}\mathbf{u}_j\|^2\right\} d\mathbf{u}_j$$

como $\mathbf{R}_{11(j)}$ es no singular, se puede realizar el cambio de variable

$$\boldsymbol{\phi}_j = \frac{1}{\sigma_e} [\mathbf{c}_{1(j)} - \mathbf{R}_{10(j)}\boldsymbol{\gamma} - \mathbf{R}_{11(j)}\mathbf{u}_j]$$

en la integral, obteniéndose

$$\int \frac{1}{(2\pi\sigma_e^2)^{\frac{q}{2}}} \exp\left\{\frac{-1}{2\sigma_e^2} \|\mathbf{c}_{1(j)} - \mathbf{R}_{10(j)}\boldsymbol{\gamma} - \mathbf{R}_{11(j)}\mathbf{u}_j\|^2\right\} d\mathbf{u}_j = |\mathbf{R}_{11(j)}|^{-1}$$

en consecuencia, la verosimilitud del modelo puede ser expresada en los siguientes términos:

$$L(\boldsymbol{\gamma}, \boldsymbol{\Psi}, \sigma_e^2; \mathbf{y}) = \prod_{j=1}^J \frac{|\boldsymbol{\Psi}|}{|\mathbf{R}_{11(j)}| (2\pi\sigma_e^2)^{n_j/2}} \exp\left\{\frac{-1}{2\sigma_e^2} \|\mathbf{c}_{0(j)} - \mathbf{R}_{00(j)}\boldsymbol{\gamma}\|^2\right\}$$

$$= \frac{1}{(2\pi\sigma_e^2)^{\frac{n}{2}}} \exp\left\{\frac{-1}{2\sigma_e^2} \sum_{j=1}^J \|\mathbf{c}_{0(j)} - \mathbf{R}_{00(j)}\boldsymbol{\gamma}\|^2\right\} \prod_{j=1}^J \frac{|\boldsymbol{\Psi}|}{|\mathbf{R}_{11(j)}|}$$

dado que el término incluido en el exponente tiene una forma similar a una suma de cuadrados residual, se puede simplificar la expresión considerando la descomposición ortogonal-triangular

$$\begin{bmatrix} \mathbf{R}_{00(1)} & \mathbf{c}_{0(1)} \\ \vdots & \vdots \\ \mathbf{R}_{00(J)} & \mathbf{c}_{0(1)} \end{bmatrix} = \mathbf{Q}_0 \begin{bmatrix} \mathbf{R}_{00} & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{c}_d \end{bmatrix}$$

así, la verosimilitud puede ser expresada por

$$L(\boldsymbol{\gamma}, \boldsymbol{\Psi}, \sigma_e^2; \mathbf{y}) = \frac{1}{(2\pi\sigma_e^2)^{\frac{n}{2}}} \exp\left\{\frac{-1}{2\sigma_e^2} [\|\mathbf{c}_d\|^2 + \|\mathbf{c}_0 - \mathbf{R}_{00}\boldsymbol{\gamma}\|^2]\right\} \prod_{j=1}^J \frac{|\boldsymbol{\Psi}|}{|\mathbf{R}_{11(j)}|}$$

Fijados los parámetros Ψ , los valores de los parámetros γ y σ_e^2 que maximizan la verosimilitud son:

$$\hat{\gamma}(\Psi) = R_{00}^{-1}c_0, \quad \hat{\sigma}_e^2(\Psi) = \frac{1}{n}\|c_d\|^2$$

a través de estas expresiones, el procedimiento de optimización de la verosimilitud resulta más fácil si se considera en primer lugar la verosimilitud “concentrada” (profiled likelihood) como una función de Ψ :

$$L(\Psi; y) = L(\hat{\gamma}(\Psi), \Psi, \hat{\sigma}_e^2(\Psi); y) = \left[\frac{n}{2\pi\|c_d\|^2} \right]^{\frac{n}{2}} \exp\left\{ \frac{-n}{2} \right\} \prod_{j=1}^J \frac{|\Psi|}{|R_{11(j)}|}$$

tomando logaritmo

$$l(\Psi; y) = \frac{n}{2} [\log n - \log 2\pi - 1] - n \log \|c_d\| + \sum_{j=1}^J \log \frac{|\Psi|}{|R_{11(j)}|}$$

así, el procedimiento de obtención de los estimadores de máxima verosimilitud consiste en maximizar la expresión anterior en términos de la colección de parámetros Ψ , obteniéndose $\hat{\Psi}$, y a partir de estos, obtener los estimadores de los parámetros γ y σ_e^2 dados por $\hat{\gamma}(\hat{\Psi})$ y $\hat{\sigma}_e^2(\hat{\Psi})$, respectivamente.

2.4.2. ESTIMADORES DE MÁXIMA VEROSIMILITUD RESTRINGIDA

El procedimiento de máxima verosimilitud produce estimadores sesgados de los parámetros aleatorios, es decir, de los parámetros incluidos en la matriz Ψ y el parámetro σ_e^2 , dado que no tiene en cuenta la “variación” de los parámetros fijos. Este sesgo puede ser importante en pequeñas muestras, como afirman Goldstein (2011) y Pinheiro y Bates (2000). Se pueden obtener estimaciones insesgadas mediante una modificación conocida como máxima verosimilitud restringida (REML). Existen diversas formas de definir este procedimiento.

Como indican Pinheiro y Bates (2000), la siguiente es, posiblemente, la más adecuada desde el punto de vista computacional.

La verosimilitud restringida de los componentes de la varianza es

$$\begin{aligned}
 L_R(\Psi, \sigma_e^2; \mathbf{y}) &= \int L(\boldsymbol{\gamma}, \Psi, \sigma_e^2; \mathbf{y}) d\boldsymbol{\gamma} \\
 &= \left[\prod_{j=1}^J \frac{|\Psi|}{|\mathbf{R}_{11(j)}|} \right] \exp\left\{ \frac{-\|\mathbf{c}_d\|^2}{2\sigma_e^2} \right\} \int \frac{1}{(2\pi\sigma_e^2)^{\frac{n}{2}}} \exp\left\{ \frac{-1}{2\sigma_e^2} [\|\mathbf{c}_0 - \mathbf{R}_{00}\boldsymbol{\gamma}\|^2] \right\} d\boldsymbol{\gamma} \\
 &= \left[\prod_{j=1}^J \frac{|\Psi|}{|\mathbf{R}_{11(j)}|} \right] \exp\left\{ \frac{-\|\mathbf{c}_d\|^2}{2\sigma_e^2} \right\} \frac{1}{(2\pi\sigma_e^2)^{\frac{n-p}{2}}} |\mathbf{R}_{00}|^{-1}
 \end{aligned}$$

por tanto, la log-verosimilitud restringida

$$l_R(\Psi, \sigma_e^2; \mathbf{y}) = -\frac{n-p}{2} \log(2\pi\sigma_e^2) - \frac{\|\mathbf{c}_d\|^2}{2\sigma_e^2} - \log|\mathbf{R}_{00}| + \sum_{j=1}^J \log \frac{|\Psi|}{|\mathbf{R}_{11(j)}|}$$

así, se obtiene el estimador condicional

$$\hat{\sigma}_e^2(\Psi) = \frac{1}{n-p} \|\mathbf{c}_d\|^2,$$

que conduce a la log-verosimilitud restringida:

$$l_R(\Psi; \mathbf{y}) = l_R(\Psi, \tilde{\sigma}_e^2(\Psi); \mathbf{y}) = cte. - (n-p) \log \|\mathbf{c}_d\|^2 - \log|\mathbf{R}_{00}| + \sum_{j=1}^J \log \frac{|\Psi|}{|\mathbf{R}_{11(j)}|}$$

así, el procedimiento de obtención de los estimadores de máxima verosimilitud restringida consiste en maximizar la expresión anterior en términos de la colección de parámetros Ψ , obteniéndose $\hat{\Psi}_R$, y a partir de estos, obtener el estimador del parámetro σ_e^2 dado por $\tilde{\sigma}_e^2(\hat{\Psi}_R)$. Finalmente, se obtiene $\hat{\boldsymbol{\gamma}}(\hat{\Psi}_R)$.

2.4.3. DISTRIBUCIONES ASINTÓTICAS DE LOS ESTIMADORES

Los estimadores de máxima verosimilitud (EMV) del modelo multinivel son consistentes y asintóticamente normales (Pinheiro y Bates 2000). Además, los EMV de los parámetros que determinan los efectos fijos están asintóticamente incorrelados con los estimadores de los parámetros asociados a los efectos aleatorios. Así, asintóticamente,

$$\hat{\boldsymbol{\gamma}} \sim N_p \left[\boldsymbol{\gamma}, \sigma_e^2 \left[\mathbf{R}_{00}^{-1} (\mathbf{R}_{00}^{-1})^\top \right] \right]$$

por otra parte, considerando las $f = q(q + 1)/2$ componentes distintas de $\boldsymbol{\Psi}$ como el vector $\boldsymbol{\theta}$, los EMV de la estructura de varianzas y covarianzas verifican, asintóticamente

$$\begin{bmatrix} \hat{\boldsymbol{\theta}} \\ \log \hat{\sigma}_e^2 \end{bmatrix} \sim N_{f+1} \left[\begin{bmatrix} \boldsymbol{\theta} \\ \log \sigma_e^2 \end{bmatrix}, \mathbf{F}^{-1}(\boldsymbol{\theta}, \sigma_e^2) \right]$$

siendo $\mathbf{F}^{-1}(\boldsymbol{\theta}, \sigma_e^2)$ la matriz de información empírica.

Asimismo, los estimadores de máxima-verosimilitud restringida son también consistentes y asintóticamente normales con distribuciones aproximadas idénticas a las recogidas para los EMV.

A partir de estas distribuciones se pueden obtener fácilmente regiones e intervalos de confianza para los parámetros.

2.4.4. CONTRASTES DE HIPÓTESIS SOBRE LOS PARÁMETROS

Considerando las propiedades genéricas del test de razón de verosimilitudes, dado un modelo $M(\boldsymbol{\omega})$ dependiente de un vector k -dimensional de parámetros $\boldsymbol{\omega}$ que puede descomponerse en dos subvectores

$$\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_2 \end{bmatrix}$$

de dimensiones k_1 y k_2 respectivamente, y se pretende contrastar la hipótesis nula

$$H_0: \boldsymbol{\omega}_2 = \mathbf{0},$$

el estadístico del test de razón de verosimilitudes viene dado por $2[\log L(\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2) - \log L(\tilde{\boldsymbol{\omega}}_1)]$, donde $L(\hat{\boldsymbol{\omega}}_1, \hat{\boldsymbol{\omega}}_2)$ es la máxima verosimilitud alcanzada por el modelo general $M(\boldsymbol{\omega})$ y $L(\tilde{\boldsymbol{\omega}}_1)$ es la máxima verosimilitud del modelo restringido por la hipótesis nula $M_0(\boldsymbol{\omega})$. Además, la distribución asintótica, bajo hipótesis nula, de dicho estadístico es una distribución χ^2 con $k - k_1$ grados de libertad.

Ello permite realizar contrastes sobre los parámetros asociados a efectos fijos $\boldsymbol{\gamma}$, así como sobre los parámetros que determinan los efectos aleatorios, aunque especialmente se utiliza en este segundo caso, es decir, para evaluar la significación de los términos incluidos en la estructura de covarianzas. Además, se pueden utilizar los estimadores de máxima verosimilitud restringida en caso de que el ajuste del modelo se haya realizado a través de este método.

Cuando el interés se centra en los parámetros asociados a efectos fijos (parámetros incluidos en el vector $\boldsymbol{\gamma}$) no debe utilizarse la máxima verosimilitud restringida (Pinheiro y Bates 2000). Incluso, ni siquiera el criterio de máxima verosimilitud es recomendable. El enfoque que debe utilizarse para los contrastes sobre la significación de algún parámetro γ_j es la distribución condicional sobre los estimadores de los parámetros asociados a los efectos aleatorios. Para unos valores fijados $\boldsymbol{\Psi}$, los estimadores condicionados de los efectos fijos $\boldsymbol{\gamma}$ y de la varianza σ_e^2 se determinan a través del método estándar de mínimos cuadrados:

$$\hat{\boldsymbol{\gamma}}(\boldsymbol{\Psi}) = \mathbf{R}_{00}^{-1} \mathbf{c}_0, \quad \hat{\sigma}_e^2(\boldsymbol{\Psi}) = \frac{1}{n - p} \|\mathbf{c}_d\|^2$$

Y, en consecuencia, la distribución aproximada de estos estimadores de máxima verosimilitud y de máxima verosimilitud restringida es exacta. Por ello, los tests condicionados para la significación de un término o parámetro asociado a efectos fijos son los usuales estadísticos F de Snedecor y t de Student del modelo de regresión lineal. En la práctica, los parámetros $\boldsymbol{\Psi}$ desconocidos se sustituyen por los estimadores de máxima verosimilitud restringida, por lo que los tests, finalmente, quedan como aproximaciones. Es decir, para el contraste de

significación de un parámetro asociado a efectos fijos ($H_0: \gamma_j = 0$) se tendrá un F -test o t -test de forma similar al modelo de regresión lineal.

Estas aproximaciones también pueden ser utilizadas para la obtención de intervalos de confianza de los parámetros:

$$\hat{\gamma}_j \pm t_{f_j, 1-\alpha/2} \hat{\sigma}_e \left[\left\{ \mathbf{R}_{00}^{-1} (\mathbf{R}_{00}^{-1})^\top \right\}_{jj} \right]^{1/2}$$

donde $t_{f_j, 1-\alpha/2}$ es el cuantil de orden $(1 - \alpha/2)$ de una distribución t de Student con f_j grados de libertad y $\left\{ \mathbf{R}_{00}^{-1} (\mathbf{R}_{00}^{-1})^\top \right\}_{jj}$ es el elemento j -ésimo diagonal de la matriz.

Análogamente, se pueden obtener intervalos de confianza para las componentes de la estructura de varianzas y covarianzas, aunque desde el punto de vista práctico tengan un menor interés.

2.4.5. AJUSTE DEL MODELO

Como medida de la calidad relativa del modelo se puede utilizar el criterio de información de Akaike (AIC). Esta medida, *trade-off* entre la bondad de ajuste del modelo y la complejidad del mismo, se define como (Akaike 1974)

$$AIC = -2 \log L(\hat{\omega}) + 2d$$

siendo $L(\hat{\omega})$ la máxima verosimilitud del modelo y d el número de parámetros que intervienen en el mismo. Así, para comparar modelos con el mismo conjunto de datos, el modelo más adecuado según dicho criterio es aquel que alcanza menor valor AIC.

Los valores ajustados por el modelo para los casos incluidos en el mismo y la predicción para nuevos casos resultan de interés en la gran mayoría de los estudios en los que se aplican estos modelos multinivel. Para un caso o individuo, el modelo trata de ajustar el valor Y de la variable objetivo en dicho individuo según la igualdad

$$Y = \mathbf{x}^T \boldsymbol{\gamma} + \mathbf{w}^T \mathbf{u} + \mathbf{e} ,$$

siendo \mathbf{x} el vector de los valores de las variables explicativas, y/o productos de ellas, asociadas a los efectos fijos (contiene una primera columna con todos sus elementos unitarios); $\boldsymbol{\gamma}$ el p -vector de parámetros que representan los efectos fijos; \mathbf{w} el vector de valores de las variables explicativas asociadas a efectos aleatorios; \mathbf{u} el q -vector de efectos aleatorios y, finalmente, \mathbf{e} es el vector de residuos o perturbaciones aleatorias. En consecuencia, el valor esperado de esta variable es $E[Y] = \mathbf{x}^T \boldsymbol{\gamma}$.

Así, el mejor predictor lineal insesgado (BLUP) del valor esperado de la población es

$$\hat{y} = \mathbf{x}^T \hat{\boldsymbol{\gamma}}$$

Obviamente, el residuo asociado será

$$r = y - \hat{y} = y - \mathbf{x}^T \hat{\boldsymbol{\gamma}}$$

El estudio de estos residuos permite profundizar en el análisis del ajuste del modelo.

Capítulo 3. MODELIZACIÓN DEL RENDIMIENTO ACADÉMICO UNIVERSITARIO

3.1. POBLACIÓN

En esta investigación se consideró el total de estudiantes de la Universidad Técnica Particular de Loja - Ecuador que desarrollan sus estudios en las dos modalidades de estudio: modalidad “presencial” y modalidad “abierta y a distancia”. Los datos están referidos a los periodos académicos: octubre 2010 – febrero 2011, abril 2011 – agosto 2011, octubre 2011 – febrero 2012 y abril 2012 – agosto 2012.

La población total durante los cuatro periodos se encuentra referida en la Tabla 10, desglosada por sexo. Debe tenerse en cuenta que las variaciones de un periodo a otro están provocadas por los egresados, abandonos y nuevos ingresos.

Tabla 10. NÚMERO DE ESTUDIANTES POR PERIODO ACADÉMICO Y SEXO

AÑO	PERIODO ACADÉMICO	NÚMERO DE ESTUDIANTES	
		MASCULINO	FEMENINO
2011	Octubre 2010 – Febrero 2011	11609	17254
	Abril 2011 – Agosto 2011	11556	17248
2012	Octubre 2011 – Febrero 2012	12594	18200
	Abril 2012 – Agosto 2012	11755	17739

Los estudiantes se distribuyen en 4 áreas de formación académica: Administrativa, Biológica y Biomédica, Socio-Humanística y Técnica y, en 44 titulaciones de las dos modalidades de estudio. El porcentaje de estudiantes por área de formación, modalidad de estudio y sexo se muestra en la Tabla 11.

Tabla 11. PORCENTAJE DE ESTUDIANTES POR ÁREA DE FORMACIÓN, MODALIDAD Y SEXO

ÁREA DE FORMACIÓN	MODALIDAD				TOTAL ÁREA
	PRESENCIAL		DISTANCIA		
	MASC	FEM	MASC	FEM	
ADMINISTRATIVA	7,4%	15,5%	11,8%	22,4%	32,9%
BIOLÓGICA Y BIOMÉDICA	11,2%	16,1%	3,7%	2,4%	8,4%
SOCIO-HUMANÍSTICA	5,1%	7,2%	21,0%	34,1%	50,5%
TÉCNICA	25,7%	11,9%	3,7%	0,9%	8,2%
TOTAL SEXO POR MODALIDAD	49,3%	50,7%	40,2%	59,8%	
TOTAL MODALIDAD	10,7%		89,3%		

El porcentaje de estudiantes distribuidos por área de formación, hábitat y sexo lo visualizamos en la Tabla 12.

Tabla 12. PORCENTAJE DE ESTUDIANTES POR ÁREA DE FORMACIÓN, HÁBITAT Y SEXO

ÁREA DE FORMACIÓN	HÁBITAT					
	URBANO		RURAL		EXTRANJERO	
	MASC	FEM	MASC	FEM	MASC	FEM
ADMINISTRATIVA	11,9%	21,1%	9,2%	22,8%	10,6%	21,3%
BIOLÓGICA Y BIOMÉDICA	4,8%	4,5%	3,9%	2,5%	2,0%	0,8%
SOCIO-HUMANÍSTICA	19,3%	28,9%	19,5%	38,3%	21,6%	37,0%
TÉCNICA	7,2%	2,4%	2,9%	0,9%	5,1%	1,6%
TOTAL SEXO POR HÁBITAT	43,1%	56,9%	35,5%	64,5%	39,3%	60,7%
TOTAL HÁBITAT	72,2%		24,8%		3,0%	

3.2. BASES DE DATOS Y TRATAMIENTO DE LA INFORMACIÓN

En el desarrollo de este estudio se utilizaron tres bases de datos con registros de alumnos, becas y profesores respectivamente. Las mismas tuvieron que ser tratadas previamente a través del software para análisis estadísticos y gráficos “R”, en virtud de su naturaleza doble de programa y lenguaje de programación (Paradis, 2003). A continuación se describe el procedimiento realizado en cada base de datos para llegar a obtener la base final.

3.2.1. BASE DE DATOS ALUMNOS

La base de datos de alumnos fue solicitada a la Unidad de Gestión de Tecnologías de Información (UGTI) de la Universidad Técnica Particular de Loja, que es el custodio del sistema académico de la Universidad. La información proporcionó el total de estudiantes durante los años 2011 – 2012 y consideró 601.560 registros que mostraron 22 campos o variables por cada estudiante – asignatura como se señala en la Tabla 13.

Tabla 13. BASE DE DATOS DE ALUMNOS

CAMPOS/VARIABLES		
Cédula	Centro universitario	Estado nota
Nombres	Carrera	Matrícula
Sexo	Código carrera	Crédito
Fecha de nacimiento	Modalidad estudios	Nivel del componente
Provincia de nacimiento	Período académico	Nacionalidad
Cantón de nacimiento	Nombre asignatura o componente	Identificación de matrícula
Ciudad de nacimiento	Código asignatura	
Dirección	Nota asignatura	

Para el análisis de esta base de datos y para la obtención de la base de datos final de alumnos se desarrollaron varios pasos que se señalan a continuación:

1. Las bases de datos fueron entregadas en formato excel, en archivos separados por ciclo académico; y debido al número de registros, cada archivo con varias hojas de datos. De esta forma, a través de “R”, se procedió a unir las hojas separadas en cada archivo; crear la columna de año académico (2011 – 2012), como identificador de los períodos académicos que cada año engloba y a unir los archivos por años. (Anexo 1)

2. Para los periodos de estudio, el sistema académico de la universidad registra la evolución de las mallas académicas de cada titulación con nombres diferenciados; por lo que es necesario establecer una sola nominación y codificación a través de tablas de equivalencia manuales. Las áreas de formación y titulaciones para la modalidad presencial y abierta a distancia se anotan en las Tablas 14, 15, 16 y 17.

Tabla 14. ÁREA ADMINISTRATIVA

TITULACIONES	
MODALIDAD PRESENCIAL	MODALIDAD ABIERTA Y A DISTANCIA
Administración de Empresas	Administración de Empresas
Administración en Banca y Finanzas	Administración en Banca y Finanzas
Asistencia Gerencial y Relaciones Públicas	Administración en Gestión Pública
Economía	Asistencia Gerencial y Relaciones Públicas
Ingeniería en Contabilidad y Auditoría	Ingeniería en Administración de Empresas Turísticas y Hoteleras
Ingeniería en Hotelería y Turismo	Ingeniería en Contabilidad y Auditoría Economía

Tabla 15. ÁREA BIOLÓGICA Y BIOMÉDICA

TITULACIONES	
MODALIDAD PRESENCIAL	MODALIDAD ABIERTA Y A DISTANCIA
Biología	Gestión Ambiental
Bioquímica y Farmacia	
Gestión Ambiental	
Ingeniería Agropecuaria	
Ingeniería en Industrias Agropecuarias	
Ingeniería Química	
Medicina	

Tabla 16. ÁREA TÉCNICA

TITULACIONES	
MODALIDAD PRESENCIAL	MODALIDAD ABIERTA Y A DISTANCIA
Arquitectura	Informática
Artes Plásticas y Diseño	
Electrónica y Telecomunicaciones	
Geología y Minas	
Ingeniería Civil	
Sistemas Informáticos y Computación	

Tabla 17. ÁREA SOCIO - HUMANÍSTICA

TITULACIONES	
MODALIDAD PRESENCIAL	MODALIDAD ABIERTA Y A DISTANCIA
Comunicación Social	Ciencias de la Educación Mención Educación Básica
Derecho	Ciencias de la Educación Mención Educación Infantil
Inglés	Ciencias de la Educación Mención Lengua y Literatura
Relaciones Públicas	Ciencias de la Educación Mención Física y Matemática
Psicología	Ciencias de la Educación Mención Química y Biología
	Ciencias de la Educación Mención Inglés
	Ciencias Humanas y Religiosas
	Comunicación Social
	Derecho
	Psicología

Luego de normalizar las denominaciones correspondientes a cada carrera, se procedió a través de “R” a establecer las equivalencias de las *áreas de formación académica, carreras y códigos* en la base de datos 2011-2012 (Anexo 2).

3. La denominación de centros universitarios presentan inconsistencias al encontrarse registrados nombres similares con tilde y sin tilde, nombres equivalentes en centros ya señalados. Así mismo, la columna de *modalidad* presenta incompatibilidades al encontrarse registros con mayúsculas y minúsculas; a través de “R” se procedió a corregir estas inconsistencias. (Anexo 3)

4. La base de datos contenía campos vacíos en las variables ciudad, provincia y cantón, por lo que a partir de la información levantada, se procede a completar los registros faltantes en caso de que consten en otros periodos académicos. (Anexo 4)

5. A través de los 93 centros universitarios registrados en los años de estudio, como apoyo a la gestión académica y administrativa de la UTPL alrededor del Ecuador y 3 centros en el extranjero, se procedió a establecer la variable *hábitat*. El hábitat urbano se crea considerando a los 27 centros universitarios ubicados en las capitales de provincia del país; el hábitat rural contemplando a 66 centros ubicados en los cantones que no fueron nominados como capital de provincia y el hábitat extranjero que corresponde a los 3 centros universitarios ubicados fuera del país. La relación de centros y hábitat, se puede observar en las Tablas 18, 19 y 20.

Tabla 18. CENTROS UNIVERSITARIOS - HÁBITAT URBANO

PROVINCIAS		
Ambato	Latacunga	Quito – Tumbaco
Azogues	Loja	Quito – Turubamba
Coca	Macas	Quito – Villaflora
Cuenca	Machala	Riobamba
Esmeraldas	Nueva Loja	San Cristóbal
Guaranda	Portoviejo	Santo Domingo
Guayaquil	Puyo	Tena
Guayaquil – Centenario	Quito – Carcelén	Tulcán
Ibarra	Quito	Zamora

Tabla 19. CENTROS UNIVERSITARIOS - HÁBITAT RURAL

PROVINCIAS		
Alamor	Jipijapa	Quevedo
Alausí	Joya De Los Sachas	Quinindé
Arenillas	La Concordia	Quito - San Rafael
Bahía de Caráquez	Isabela	Salinas
Balsas	La Troncal	Samborondon
Calceta	Limón Indanza	San Gabriel
Cañar	Limonos	San Lorenzo
Cariamanga	Macará	San Miguel de los Bancos
Catacocha	Machachi	Santa Cruz
Cayambe	Manta	Santa Isabel
Celica	Méndez	Santa Rosa
Chone	Milagro	Santuario de Olón
Daule	Montalvo	Saraguro
Durán	Otavalo	Sevilla de Oro
El Ángel	Pacto	Shushufindi
El Chaco	Pasaje	Sucúa
Gonzanamá	Paute	Tabacundo
Guadalupe	Pedernales	Valle de los Quijos
Gualaceo	Pedro Vicente Maldonado	Vinces
Gualaquiza	Piñas	Yanzatza
Guaysimi	Portovelo	Zaruma
Huaquillas	Pucará	Zumba

Tabla 20. CENTROS UNIVERSITARIOS - HÁBITAT EXTRANJERO

CIUDADES		
Madrid	New York	Roma

A partir de esta clasificación, a través de “R” se procedió a crear la columna *hábitat* con sus equivalencias correspondientes. (Anexo 5)

6. La columna fecha de nacimiento presenta inconsistencias al determinarse que varios campos fueron rellenos con fechas fuera del rango normal de vida estudiantil de una persona, por lo que se establece considerar a todos los alumnos nacidos a partir de 1930 hasta 1945, esto considerando las características propias de los estudiantes de modalidad presencial y abierta y a distancia. (Anexo 6)

7. En función de la columna anterior, se establece la variable *edad*, considerando como fecha límite el año 2012. (Anexo 7)

8. La columna de créditos de cada asignatura presenta campos vacíos, créditos superiores a los establecidos en la normativa académica; y, materias con cero créditos por lo que se solicita una base adicional de asignaturas y créditos para contrastar las bases y modificar los errores encontrados. (Anexo 8)

9. A partir de la columna créditos se procede a calcular la *Nota Media Ponderad*, *Porcentaje de Créditos Aprobados* y *Porcentaje de Asignaturas*. La nota media para cada alumno y periodo se obtuvo ponderando la calificación obtenida en cada asignatura con sus créditos correspondientes y dividiendo por el total de créditos cursados por el alumno. (Anexo 9)

3.2.2. BASE DE DATOS BECAS

La segunda base de datos utilizada en la presente investigación, corresponde a la base de becas, la misma sostuvo información de becas 2011 – 2012, reportada en formato excel por la Universidad Técnica Particular de Loja al Sistema Nacional de Información de la Educación Superior del Ecuador (SNIESE) de la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESYT), en el año 2013.

La base de datos de becas estuvo compuesta de 87.603 registros que especificaron 60 variables por cada estudiante de grado y postgrado, las variables se muestran en la Tabla 21.

Tabla 21. BASE DE DATOS DE BECAS

CAMPOS/VARIABLES		
Código campus	Monto beca total	Email institucional
Campus	Beca parcial	Documento identificación
Provincia	Monto beca parcial	Cédula
Cantón	Ayuda económica	Primer apellido
Parroquia	Monto ayuda económica	Segundo Apellido
Dirección	Crédito económico	Nombres
Código oferta	Monto crédito económico	Provincia nacimiento
Oferta académica	Ayuda arancel	Cantón nacimiento
Sub área UNESCO	Monto arancel	Fecha nacimiento
Nivel	Manutención	Tipo discapacidad
Modalidad	Monto manutención	Carnet CONADIS
Organización malla curricular	Beca socioeconómico	Número carnet CONADIS
Jornada, carrera, programa	Beca excelencia académica	Porcentaje de discapacidad
Prácticas preprofesionales	Beca discapacidad	Número matrícula

CAMPOS/VARIABLES

Horas prácticas preprofesionales	Beca deportiva	Nacionalidad
Nombre, entidad, pasantías	Sistema registro consultas académicas	Sexo
Número créditos	Correo electrónico institucional	Autoidentificación etnia
Fecha inicio primer nivel	Sistema matrícula en línea	Ingreso familiar
Fecha inicio convalidación	Servicio biblioteca virtual	Email personal
Beca total	Plataforma enseñanza virtual	Observaciones

Con la base de datos de becas, se procedió a obtener la base de datos de alumnos – becas, con los pasos siguientes:

1. Utilizando “R”, se realiza la selección de la información de interés correspondiente a seis variables de los estudiantes de grado: cédula de identidad del estudiante, beca deportiva, beca por discapacidad, beca de excelencia académica, beca socioeconómica; y, ayuda económica de los años 2011 – 2012. (Anexo 10)
2. Sobre el archivo de estudiantes 2011-2012, se procedió a identificar a los alumnos del año 2011 y 2012, para realizar la unión de los archivos alumnos – becas. (Anexo 11)

3.2.3. BASE DE DATOS PROFESORES

La tercera base de datos utilizada para este estudio es la de profesores, la misma al igual que la base de datos de becas, correspondió a la información completa de profesores durante los años 2011 - 2012, reportada por la UTPL al Sistema Nacional de Información de la Educación Superior del Ecuador (SNIESE) de la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESYT), en el año 2013.

La base de datos de profesores estuvo compuesta de 3.712 registros que incluyen 144 variables por cada docente, las variables se muestran en la Tabla 22.

Tabla 22. BASE DE DATOS DE DOCENTES

CAMPOS/VARIABLES			
Código campus	Nivel penúltimo	título cuarto nivel	Horas administrativas
Campus	Penúltimo	título cuarto nivel	Horas vinculación sociedad
Tipo personal IES	Fecha penúltimo	título cuarto nivel	Horas otras actividades
Provincia	Universidad_ cuarto nivel	penúltimo título	Otras actividades
Cantón	País penúltimo	título cuarto nivel	Autoevaluación
Parroquia	Registro SENESCYT título cuarto nivel	penúltimo	Coevaluación pares
Dirección	Subárea penúltimo	título cuarto nivel	Coevaluación directivos
Documento identificación	Nivel antepenúltimo	título cuarto nivel	Heteroevaluación
Cédula	Título cuarto nivel	antepenúltimo título	Número cursos nacionales
Primer apellido	Fecha cuarto nivel	antepenúltimo título	Número cursos internacionales
Segundo apellido	Universidad cuarto nivel	antepenúltimo título	Número cursos metodología
Nombres	País antepenúltimo	título cuarto nivel	Cursando estudios
Sexo	Registro SENESCYT antepenúltimo	título cuarto nivel	Financiamiento estudios PHD
Nacionalidad	Subárea título	profe 3	Periodo sabático
Fecha nacimiento	Número nombramiento	contrato	Fecha inicio sabático
Tipo discapacidad	Fecha ingreso IES		Fecha fin sabático

CAMPOS/VARIABLES		
Carnet CONADIS	Relación trabajo IES	Nombre libro publicado personal académico 1
Número carnet CONADIS	Ingreso concurso	Año publicación libro 1
Porcentaje_discapacidad	Cargo contrato nombramiento	Participación libro 1
Email personal	Horas laboradas anuales	Revisión libro pares 1
Nivel último título tercer nivel	Remuneración anual	Numero ISBN 1
Ultimo título tercer nivel	Fecha salida IES	Nombre libro publicado personal académico 2
Fecha último título tercer nivel	Docencia tercer nivel	Año publicación libro 2
Universidad último título tercer nivel	Docencia cuarto nivel	Participación libro 2
País último título tercer nivel 1	Código asignatura dictada 1	Revisión libro pares 2
Registro SENESCYT último título tercer nivel	Asignatura dictada 1	Número ISBN 2
Subárea último título tercer nivel	Número horas asignatura 1	Nombre artículo 1
Nivel penúltimo título tercer nivel	Código asignatura dictada 2	Nombre revista artículo publicado personal académico 1
Penúltimo título tercer nivel	Asignatura dictada 2	Año publicación artículo 1
Fecha penúltimo título tercer nivel	Número horas asignatura 2	Participación artículo 1
Universidad penúltimo título tercer nivel	Código asignatura dictada 3	Estado publicación 1
País penúltimo título tercer nivel	Asignatura dictada 3	Revista indexada publicación artículo 1
Registro SENESCYT penúltimo título tercer nivel	Número horas asignatura 3	Base datos revista indexada1
Subárea penúltimo título tercer nivel	Código asignatura dictada 4	Nombre artículo 2
Nivel antepenúltimo_ título tercer nivel	Asignatura dictada 4	Nombre revista artículo publicado personal académico 2

CAMPOS/VARIABLES				
Antepenúltimo nivel	título tercer	Numero horas asignatura 4		Año publicación artículo 2
Fecha tercer nivel	antepenúltimo título	Código asignatura dictada 5		Participación artículo 2
Universidad título tercer nivel	antepenúltimo_	Asignatura dictada 5		Estado publicación 2
País antepenúltimo nivel	título tercer	Número horas asignatura 5		Revista indexada publicación artículo 2
Registro antepenúltimo nivel	antepenúltimo título tercer	Cargo autoridad		Base datos revista indexada 2
Subárea tercer nivel	antepenúltimo título	Fecha inicio cargo autoridad		Movilidad IES extranjeras
Nivel último	título cuarto nivel	Fecha fin cargo autoridad		Movilidad IES nacionales
Último título cuarto nivel		Número documento cargo autoridad		Sistema registro consultas académicas
Fecha último título cuarto nivel		Categoría personal académico		Correo electrónico institucional
Universidad cuarto nivel	último_ título	Tiempo dedicación		Servicio biblioteca virtual
País último título cuarto nivel		Unidad académica		Plataforma enseñanza virtual
Registro título cuarto nivel	antepenúltimo_ último	Horas dedicación		Email institucional
Subárea último nivel	último título cuarto	Horas investigación		Observaciones

Después del análisis de la base de datos correspondiente se realiza el procedimiento para alcanzar la base final estudiantes – becas - profesores, que se describe seguidamente:

1. Se seleccionan las variables a ser utilizadas como fecha de ingreso a la institución, fecha de nacimiento, remuneración, categoría del personal académico, sexo, nivel de formación, tiempo de dedicación y asignaturas. (Anexo 12)

2. Por cada asignatura del grupo de profesores que la imparten se establece la *media de la experiencia docente, media de la edad, media de retribuciones, categoría máxima, porcentaje de mujeres docentes que han impartido una asignatura, nivel de formación máxima, tiempo de dedicación máxima; y, número de profesores por asignatura.* (Anexo 13)

3. Para unir los archivos base de profesores y estudiantes – becas, se procedió a identificar a cada asignatura por su código en las dos bases y obtener la base final. (Anexo 14)

De esta forma, la base final estudiantes – becas – profesores, a ser utilizada para establecer el modelo multinivel de rendimiento universitario, está compuesta por 557.675 registros con un total de 46 variables como se describe en la Tabla 23.

Tabla 23. BASE ALUMNOS - BECAS - PROFESORES

CAMPOS/VARIABLES		
Asignatura Código	Código carrera	Experiencia media
Cédula	Periodo	Edad media
Modalidad	Componente	Retribución media
Nombres	Nota	Categoría máxima
Sexo	Estado Nota	Porcentaje mujeres
Fecha de nacimiento	Matrícula	Formación máxima
Edad	Créditos	Dedicación máxima
Provincia	Nivel componente	Número profesores por asignatura
Cantón	Nacionalidad	Porcentaje aprobados por componente
Ciudad	Identificación de matrícula	Porcentaje aprobados por carrera
Dirección	Año	Porcentaje aprobados por área de formación
Hábitat	Ayuda Económica	Nota media
Centro Universitario	Beca Socioeconómica	Nota media ponderada
Área Formación	Beca excelencia académica	Porcentaje de créditos aprobados
Carrera	Beca discapacidad	Porcentaje de asignaturas aprobados

3.3. DATA CLEANING: ADAPTACIÓN DE DATOS CORRESPONDIENTES AL PRIMER PERIODO

A través del procedimiento aquí recogido se adaptan los datos correspondientes al primer periodo bajo estudio (periodo octubre de 2010 a febrero de 2011) con objeto de poder aplicar los modelos multinivel. Obviamente, de forma análoga se procede con los restantes tres periodos.

En primer lugar, se carga el espacio R que incluye los datos y la librería "multinivel".

```
library(multilevel)
load("L:/TRABAJOPERSONAL/INVESTIGACION/MaryMoroch/estudiantesbecasdocentes_ultimo.RData")
```

Las dimensiones del conjunto de datos inicial son las que se recogen a continuación, que corresponden al número de variables consideradas (46) y el número de registros (uno por cada alumno-asignatura, un total de 557.675 registros)

```
dim(estudiantesbecasdocentes_ultimo)
## [1] 557675 46
```

REORDENACIÓN DE LOS DATOS

Con objeto de ir agrupando y agregando valores dentro de variables nuevas para construir las variables de Nivel 2 (nivel de CARRERA o titulación cursada por el alumno) se procede inicialmente a la reordenación de los datos:

```
Temp0 <- data.frame(estudiantesbecasdocentes_ultimo[,1:18],
                    estudiantesbecasdocentes_ultimo[,21:31],
                    estudiantesbecasdocentes_ultimo[,20:19],
                    estudiantesbecasdocentes_ultimo[,43:46],
                    estudiantesbecasdocentes_ultimo[,32:34],
                    estudiantesbecasdocentes_ultimo[,36:35],
                    estudiantesbecasdocentes_ultimo[,37:39])
```

SELECCIÓN DE REGISTROS

Una vez reordenados se procede a "limpiar" los datos de posibles errores y seleccionar los correspondientes al primer periodo analizado. Así, se seleccionan los registros que verifican:

- Tienen la NOTA correcta, es decir, dentro del intervalo [0,40).
- Corresponden al primer periodo "Oct/2010 - Feb/2011"

```
datos0 <- Temp0[which(Temp0$NOTA<40 & Temp0$NOTA>-0.0001 &
                    Temp0$PERIODO=="Octubre/2010 - febrero/2011"),]
```

PROCESO DE UNIÓN DE REGISTROS

Dado que los registros son "alumno-asignatura", se deben agrupar por alumnos, que se encuentran identificados por la variable CEDULA, asociando los valores medios a las variables "NOTA", "EXPERIENCIAMEDIA", "EDADMEDIA", "RETRIBUCIONMEDIA", "PORCENTAJEMUJERES". Es decir, se ha de proceder a la "UNIÓN" de los registros con la misma CEDULA con la media de las variables indicadas a las que hay que añadir la calificación media obtenida (NOTAMEDIA), la calificación media ponderada por créditos (NOTAMEDIAPONDERADA), porcentaje de créditos superados (PORCREDITOSAPROBADOS), porcentaje de componentes superados (PORCOMPONENTEAPROBADOS)

En primer lugar, en TEMP0 se incluyen CEDULA, y los valores medios de NOTA, EXPERIENCIAMEDIA, EDADMEDIA, RETRIBUCIONMEDIA, PORCENTAJEMUJERES con la orden "aggregate"

```
Temp0 <- aggregate(datos0[, 31:39], list(datos0$CEDULA), mean, na.rm=T)
```

Posteriormente, se añaden el resto de variables y se renombran dentro del data.frame que se ha creado


```
Temp0 <- aggregate (datos0[,31:39],list(datos0$CEDULA,datos0$SEXO,datos0$HABITAT,
                                     datos0$AREAFORMACION,datos0$CARRERA,
                                     datos0$MODALIDAD,datos0$EDAD,
                                     datos0$AYUDA_ECONOMICA,
                                     datos0$BECA_SOCIO_ECONOMICA,
                                     datos0$BECA_EXCELENCIA_ACADEMICA,
                                     datos0$BECA_DISCAPACIDAD,datos0$BECA_DEPORTIVA),
                  mean,na.rm=TRUE)
names(Temp0) <-c("CEDULA","SEXO","HABITAT","AREAFORMACION","CARRERA","MODALIDAD","EDAD",
"AYUDA_ECONOMICA","BECA_SOCIO_ECONOMICA","BECA_EXCELENCIA_ACADEMICA","BECA_DISCAPACIDAD","BECA_DEPOR
TIVA","Notamed","NOTAMEDIA","MEDIAPOND","PORCREDAPROB","PORCOMPONENTEAPROBADOS","Experienciamedia",
"Edadmedia","Retribucionmedia","Porcentajemujeres")
```

SELECCIÓN DE REGISTROS

Existen algunos alumnos que tras matricularse no continúan sus estudios. Por tanto se han de seleccionar los registros (alumnos) que no abandonan, es decir,

$$\text{Notamed} > 0, \text{PORCREDAPROB} > 0, \text{MEDIAPOND} > 0.$$

En particular aquellos con un porcentaje de créditos superados mayor que 0.

```
datos0 <- Temp0[which(Temp0$Notamed>0),]
Temp0 <- datos0[which(datos0$PORCREDAPROB>0),]
datos0 <- Temp0[which(Temp0$MEDIAPOND>0),]
dim(datos0)
## [1] 23085    21
```

PREPARACIÓN DE LAS VARIABLES DE NIVEL 2

Se calculan las medias por CARRERA de las variables asociadas al profesorado y se incluyen en el data.frame.

```
TEMP<-aggregate(datos0[,18:21],list(datos0$CARRERA),mean,na.rm=T)
names(TEMP)<-c("CARRERA","C.EXP MED","C.EDAD MED","C.RETRMED","C.PORCMUJ")
datos.final<-merge(datos0,TEMP,by=c("CARRERA"))
```

DETERMINACIÓN VARIABLE BECA

Dado que se tienen distintos tipos de becas, tras el análisis de las mismas se ha considerado conveniente considerar sólo las denominadas "beca socio-económica" y "beca de excelencia académica". Las otras modalidades (ayuda económica, deportivas,...) no son consideradas dado que no están asociadas al rendimiento académico previo.

```
attach(datos.final)
datos.final$beca<-ifelse(datos.final$BECA_SOCIO_ECONOMICA=="SI", "SI", "NO")
datos.final$beca[datos.final$BECA_EXCELENCIA_ACADEMICA=="SI"]<- "SI"
detach(datos.final)
```

ALMACENAMIENTO DEL DATA.FRAME

Se procede a almacenar el data.frame creado, así como obtener un breve resumen del mismo.

```
write.table(datos.final,"NotasmedPer1")
dim(datos.final)
## [1] 23085    26
summary(datos.final)
##      CARRERA          CEDULA          SEXO
## Length:23085      Length:23085      Length:23085
## Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character
##
##
##
##      HABITAT          AREAFORMACION      MODALIDAD          EDAD
## Length:23085      Length:23085      Length:23085      Min.   :18.00
## Class :character  Class :character  Class :character  1st Qu.:23.00
## Mode  :character  Mode  :character  Mode  :character  Median :28.00
##
##
##
##
##
##
##
##
##      AYUDA_ECONOMICA  BECA_SOCIO_ECONOMICA  BECA_EXCELENCIA_ACADEMICA
## Length:23085      Length:23085      Length:23085
## Class :character  Class :character  Class :character
```

```

## Mode :character Mode :character Mode :character
##
##
##
##
## BECA_DISCAPACIDAD BECA_DEPORTIVA Notamed NOTAMEDIA
## Length:23085 Length:23085 Min. : 0.6667 Min. : 0.6667
## Class :character Class :character 1st Qu.:26.6667 1st Qu.:26.6667
## Mode :character Mode :character Median :30.3333 Median :30.3333
## Mean :29.4787 Mean :29.4847
## 3rd Qu.:33.0000 3rd Qu.:33.0000
## Max. :40.0000 Max. :40.0000
##
## MEDIAPOND PORCREDAPROB PORCOMPONENTEAPROBADOS
## Min. : 0.5926 Min. : 5.714 Min. : 11.11
## 1st Qu.:26.6364 1st Qu.: 57.143 1st Qu.: 60.00
## Median :30.2500 Median : 83.333 Median : 83.33
## Mean :29.4088 Mean : 75.732 Mean : 76.07
## 3rd Qu.:32.9259 3rd Qu.:100.000 3rd Qu.:100.00
## Max. :40.0000 Max. :100.000 Max. :100.00
##
## Experienciamedia Edadmedia Retribucionmedia Porcentajemujeres
## Min. : 1.000 Min. :25.00 Min. : 0 Min. : 0.00
## 1st Qu.: 4.500 1st Qu.:35.21 1st Qu.: 5411 1st Qu.: 50.00
## Median : 5.750 Median :37.74 Median : 8243 Median : 66.67
## Mean : 6.655 Mean :38.26 Mean : 9176 Mean : 63.85
## 3rd Qu.: 8.228 3rd Qu.:41.00 3rd Qu.:12265 3rd Qu.: 81.94
## Max. :34.000 Max. :70.00 Max. :69038 Max. :100.00
## NA's :51 NA's :51 NA's :51 NA's :51
## C.EXP MED C.EDAD MED C.RETR MED C.PORCMUJ
## Min. : 3.196 Min. :34.39 Min. : 5573 Min. :17.96
## 1st Qu.: 5.914 1st Qu.:36.16 1st Qu.: 7582 1st Qu.:54.92
## Median : 6.285 Median :38.01 Median : 9549 Median :62.43
## Mean : 6.655 Mean :38.27 Mean : 9177 Mean :63.85
## 3rd Qu.: 7.365 3rd Qu.:38.88 3rd Qu.:10527 3rd Qu.:75.28
## Max. :11.572 Max. :48.08 Max. :17226 Max. :83.65
##
## beca
## Length:23085
## Class :character
## Mode :character
##
##
##
##

```

3.4. ANÁLISIS MULTINIVEL DEL PERIODO OCTUBRE/2010 - FEBRERO/2011, INDICADOR NOTA MEDIA PONDERADA

En primer lugar, se carga el conjunto de datos correspondientes al primer periodo analizado, generado según se recoge en el apartado anterior, y la librería "multilevel".

```
library(multilevel)
datos.final<-read.table("NotasmedPer1")
```

Con objeto de comprobar la lectura de los datos, se pide la dimensión del data.frame.

```
dim(datos.final)
## [1] 23085 26
```

Inicialmente se realizan unos pasos previos que consisten en analizar la influencia de la titulación cursada (CARRERA) en el rendimiento, a través de un modelo de efectos aleatorios, y la influencia de las variables consideradas de Nivel 1 (variables del alumno o alumna), a través de un modelo de Regresión Múltiple. Ello permitirá comprobar si es necesario o no construir un modelo multinivel.

3.4.1. PASO 1A: INFLUENCIA DE LA CARRERA COMO EFECTO ALEATORIO

Modelo de componentes de la varianza, considerando como factor aleatorio la titulación, a través de la variable CARRERA. Así, el modelo muestral puede ser representado por:

$$y_{ij} = \gamma_0 + u_{0j} + e_{ij}$$

donde

$$u_{0j} \sim N(0, \sigma_{u_0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{0j}, e_{ij}] = 0 \quad \forall (i, j)$$

- γ_0 : es un parámetro intercept fijo
- u_{0j} : error aleatorio asociado a la unidad j de Nivel 2,
- e_{ij} : error aleatorio asociado a la unidad i de Nivel 1 de la unidad j de Nivel 2 ,
- u_{0j} y e son independientes (o incorreladas)

Es decir, se establece que el único predictor de la variable objetivo es un término "intercept", y que éste puede variar como función de los valores de "CARRERA". Dado que tan sólo se tiene el objetivo de analizar la existencia de los efectos aleatorios a través de las componentes de la varianza, se obtienen las estimaciones de los parámetros de la estructura de covarianzas. Las componentes de la varianza son: varianza asociada al INTERCEPT o efecto aleatorio ($\sigma_{u_0}^2$) y varianza asociada al error experimental o varianza residual (σ_e^2).

```
Null.Model1A <- lme(MEDIAPOND~1,random=~1|CARRERA,data=datos.final,control=list(opt="optim"))
VarCorr(Null.Model1A) # Estimaciones de Los parámetros de La estructura de covarianzas

## CARRERA = pdLogChol(1)
##           Variance StdDev
## (Intercept)  0.8634849 0.9292389
## Residual    23.8255739 4.8811447
```

Para analizar la significación de $\sigma_{u_0}^2$, es decir, la significación de los efectos aleatorios, se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis de que $\sigma_{u_0}^2 = 0$, es decir,

```
Null.Model.0<-gls(MEDIAPOND~1,data=datos.final,control=list(opt="optim"))
logLik(Null.Model.0)*-2

## 'log Lik.' 139205.2 (df=2)

logLik(Null.Model1A)*-2

## 'log Lik.' 138800.2 (df=3)

(logLik(Null.Model.0)*-2) - (logLik(Null.Model1A)*-2) # Diferencia

## 'log Lik.' 404.9585 (df=2)

anova(Null.Model1A,Null.Model.0)

##           Model df      AIC      BIC    logLik  Test  L.Ratio p-value
## Null.Model1A     1   3 138806.2 138830.4 -69400.12
## Null.Model.0     2   2 139209.2 139225.3 -69602.60 1 vs 2 404.9585 <.0001
```

En conclusión: el modelo resulta significativo, es decir, la variable Nota Media Ponderada toma valores medios distintos según la CARRERA que se esté cursando.

3.4.2. PASO 1B: MODELO LINEAL MÚLTIPLE RESPECTO A LAS VARIABLES DE NIVEL 1

El modelo que se pretende estudiar en este paso es el modelo lineal múltiple de la variable objetivo respecto de las variables medidas sobre los estudiantes, es decir, las variables de Nivel 1.

Si se denota el vector de coeficientes de regresión por $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_p]^T$ y por el subíndice (i, j) los valores observados en el estudiante (i, j) , entonces el modelo muestral queda representado como sigue:

$$y_{ij} = \gamma_0 + \gamma_1 \mathbf{Edad}_{ij} + \gamma_2 \mathbf{Sexo}_{ij} + \gamma_3 \mathbf{Mod}_{ij} + \gamma_4 \mathbf{Sex} * \mathbf{Mod}_{ij} + \gamma_5 \mathbf{Rur}_{ij} + \gamma_6 \mathbf{Urb}_{ij} + \gamma_7 \mathbf{Beca}_{ij} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

donde

- γ_0 : es el parámetro intercept fijo
- $\gamma_1, \dots, \gamma_p$: son los parámetros fijos de regresión asociados a las variables predictoras (SEXO, MODALIDAD, la interacción SEXO con MODALIDAD, EDAD, HABITAT y BECA)
- e_{ij} : error aleatorio asociado a la unidad (i, j) de Nivel 1, verificando $e_{ij} \sim N(0, \sigma_e^2)$, independientes entre sí.

Dado que las variables sexo, modalidad, hábitat y beca son cualitativas, se consideran las siguientes modalidades de referencia (con valor 0):

- Sexo: Femenino (referencia), Masculino (valor 1)
- Modalidad: A distancia (referencia), Presencial (valor 1)
- Hábitat: Extranjero (referencia), Rural (Rur valor 1), Urbano (Urb valor 1)
- Beca: No (referencia), Si (valor 1)

Se procede a estudiar este modelo con el comando R: `lm()`. Se han analizado modelos que incluyen las restantes interacciones entre variables (además de SEXO*MODALIDAD), pero no resultan significativas.

```
Null.ModelB<-lm(MEDIAPOND~ SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + HABITAT +
beca,data=datos.final)
summary(Null.ModelB)

##
## Call:
## lm(formula = MEDIAPOND ~ SEXO + MODALIDAD + SEXO * MODALIDAD +
##     EDAD + HABITAT + beca, data = datos.final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.2969  -2.7126   0.8463   3.4569  11.8340
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    28.346979   0.218631  129.656 < 2e-16 ***
## SEXOMASCULINO -0.244072   0.072956  -3.345 0.000823 ***
## MODALIDADPRESENCIAL 1.326552   0.121321  10.934 < 2e-16 ***
## EDAD           0.078407   0.003974  19.728 < 2e-16 ***
## HABITATRURAL   -1.670878   0.185479  -9.008 < 2e-16 ***
## HABITATURBANA  -1.487224   0.176807  -8.412 < 2e-16 ***
## becaSI         4.380065   0.182568  23.991 < 2e-16 ***
## SEXOMASCULINO:MODALIDADPRESENCIAL -1.579550   0.164410  -9.607 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.806 on 23077 degrees of freedom
## Multiple R-squared:  0.05134, Adjusted R-squared:  0.05105
## F-statistic: 178.4 on 7 and 23077 DF, p-value: < 2.2e-16
```

Como conclusión se obtiene que todas las variables explicativas incluidas resultan significativas, es decir, los valores medios de MEDIAPOND se ven afectadas por la variación de las mismas. De forma global, también puede ser considerado como significativo, el coeficiente de determinación aunque es muy bajo ($R^2=0.05134$).

Dado este resultado y el obtenido en el PASO 1A, se concluye la necesidad de profundizar en la aplicación de los modelos multinivel sobre el conjunto de datos, comenzando con un modelo lineal múltiple con intercepts aleatorios.

3.4.3. PASO 2: MODELIZACIÓN DE LA ALEATORIEDAD DEL INTERCEPT

2A. ESTUDIO DE LA ALEATORIEDAD DE LOS INTERCEPTS DEL MODELO

En este modelo se consideran las variables medidas sobre los alumnos (SEXO, MODALIDAD, SEXO*MODALIDAD, EDAD, HABITAT y beca) y la variable CARRERA como factor aleatorio afectando sobre el término independiente o intercept. El modelo se puede representar por:

$$y_{ij} = \gamma_0 + u_{0j} + \gamma_1 \mathbf{Edad}_{ij} + \gamma_2 \mathbf{Sexo}_{ij} + \gamma_3 \mathbf{Mod}_{ij} + \gamma_4 \mathbf{Sex} * \mathbf{Mod}_{ij} + \gamma_5 \mathbf{Rur}_{ij} + \gamma_6 \mathbf{Urb}_{ij} + \gamma_7 \mathbf{Beca}_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$\text{Cov}[u_{0j}, e_{ij}] = 0 \quad \forall (i, j)$$

donde

- γ_0 : es el parámetro intercept fijo
- $\gamma_1, \dots, \gamma_p$: son los parámetros fijos de regresión asociados a las predictoras de Nivel 1
- u_{0j} : error aleatorio asociado a la unidad j de Nivel 2, verificando $u_{0j} \sim N(0, \sigma_{u0}^2)$
- e_{ij} : error aleatorio asociado a la unidad de Nivel 1, verificando $e_{ij} \sim N(0, \sigma_e^2)$
- u_{0j} y e_{ij} son independientes (o incorreladas)

```
Model.1 <-lme(MEDIAPOND ~ SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + HABITAT + beca, random=~1|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.1)$tTable
```

##	Value	Std.Error	DF	t-value
## (Intercept)	28.0706107	0.282212486	23045	99.466225
## SEXOMASCULINO	-0.0843937	0.077430449	23045	-1.089929
## MODALIDADPRESENCIAL	1.8155927	0.158464827	23045	11.457386
## EDAD	0.0725349	0.004022291	23045	18.033229
## HABITATRURAL	-1.6983153	0.185436567	23045	-9.158470
## HABITATURBANA	-1.4813966	0.175902249	23045	-8.421704
## becaSI	4.2907962	0.181925362	23045	23.585476
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.4278683	0.172280620	23045	-8.288038
##	p-value			
## (Intercept)	0.000000e+00			
## SEXOMASCULINO	2.757557e-01			
## MODALIDADPRESENCIAL	2.608730e-30			


```
## EDAD 3.353362e-72
## HABITATRURAL 5.690523e-20
## HABITATURBANA 3.924151e-17
## becaSI 1.496357e-121
## SEXOMASCULINO:MODALIDADPRESENCIAL 1.213457e-16
```

Se observa en la tabla de estimaciones que todos los coeficientes del modelo son significativos, salvo el asociado a SEXO. Dado que la interacción de SEXO*MODALIDAD también resulta significativa, se opta por mantener la variable SEXO en el modelo.

Dado que se mantiene el objetivo de analizar la existencia de los efectos aleatorios, se obtienen las estimaciones de los parámetros de la estructura de covarianzas. Nuevamente en este modelo, las componentes de la varianza son: varianza asociada al INTERCEPT o efecto aleatorio ($\sigma_{u_0}^2$) y varianza asociada al error experimental o varianza residual (σ_e^2).

VarCorr(Model.1)

```
## CARRERA = pdLogChol(1)
##          Variance StdDev
## (Intercept) 0.8710358 0.933293
## Residual    22.7037946 4.764850
```

En relación a la significación de $\sigma_{u_0}^2$, se realiza el test de razón de verosimilitudes a través de la comparación de la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis nula de que $\sigma_{u_0}^2 = 0$. Se obtiene de dos formas diferentes con objeto de comprobar que realmente la orden "anova" realiza la operación deseada.

```
Null.Model.1<-gls(MEDIAPOND~SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + + HABITAT
+ beca,data=datos.final,control=list(opt="optim"))
logLik(Null.Model.1)*-2 # (-2) veces La Verosimilitud del modelo
## 'log Lik.' 138012.7 (df=9)
logLik(Model.1)*-2 # (-2) veces La Verosimilitud del modelo bajo hipótesis
nula
## 'log Lik.' 137704.5 (df=10)
(logLik(Null.Model.0)*-2) - (logLik(Null.Model.1)*-2) # Estadístico del TRV
## 'log Lik.' 404.9585 (df=2)
anova(Model.1,Null.Model.1)
```

```
##           Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## Model.1      1 10 137724.5 137805.0 -68852.26
## Null.Model.1  2  9 138030.7 138103.1 -69006.36 1 vs 2 308.1993 <.0001
```

A través del Test de Razón de Verosimilitud concluimos que la componente de la varianza σ_{u0}^2 debe ser considerada como significativa, no nula, y por tanto, el efecto aleatorio CARRERA sobre los valores medios de la variable objetivo debe ser considerado como significativo. Además, a través del criterio de Akaike se observa que se mejora el ajuste de los datos al considerar la aleatoriedad en los intercepts provocada por la variable CARRERA:

AIC(Model.1,Null.Model.1)

```
##           df      AIC
## Model.1    10 137724.5
## Null.Model.1  9 138030.7
```

Como conclusión general del modelo se obtiene: Todas las variables explicativas incluidas resultan significativas, salvo SEXO, aunque se opta por mantenerla en el modelo, es decir, los valores medios de MEDIAPOND se ven afectados por la variación de las mismas. Además, el efecto aleatorio CARRERA sobre los valores medios también resulta significativo.

Así, se puede plantear:

1. ¿El efecto CARRERA sobre las medias se puede explicar a través de algunas de las variables asociadas a las carreras (Nivel 2)?
2. ¿La relación de la variable objetivo con las variables de Nivel 1 (alumnos) están afectadas por un efecto aleatorio debido al factor CARRERA? Y en su caso, ¿este efecto aleatorio se puede explicar a través de algunas de las variables asociadas a las carreras (Nivel 2)?

2B. INCLUSIÓN DE VARIABLES DE NIVEL 2 EN LA ALEATORIEDAD DE LOS INTERCEPTS

El modelo que se pretende analizar en este paso es incluir inicialmente la variable de Nivel 2 Área de Formación en la explicación de la variabilidad de los intercepts asociada a la titulación

que cursa el estudiante. Recuérdese que las áreas de formación son: administrativa (modalidad de referencia en el modelo), biológica y biomédica ($AFbb=1$), socio-humanística ($AFsh=1$) y técnica ($AFtc=1$).

Así, el modelo bajo estudio será:

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{04}AFtc_j + u_{0j}] + \gamma_1Edad_{ij} + \gamma_2Sexo_{ij} + \gamma_3Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{0j}, e_{ij}] = 0 \quad \forall (i, j)$$

```
Model.2 <-lme(MEDIAPOND ~ SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + HABITAT +
beca + AREAFORMACION,
random=~1|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.2)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	28.00321017	0.354906586	23045
## SEXOMASCULINO	-0.07952057	0.077380058	23045
## MODALIDADPRESENCIAL	1.90399558	0.160052032	23045
## EDAD	0.07243308	0.004021786	23045
## HABITATRURAL	-1.70180695	0.185421640	23045
## HABITATURBANA	-1.48105225	0.175891631	23045
## becaSI	4.28480469	0.181916314	23045
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.23700844	0.435065209	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	0.76881063	0.362649550	29
## AREAFORMACIONTÉCNICA	-0.93761282	0.421444461	29
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.39736191	0.172446963	23045
##	t-value	p-value	
## (Intercept)	78.9030446	0.000000e+00	
## SEXOMASCULINO	-1.0276623	3.041195e-01	
## MODALIDADPRESENCIAL	11.8961038	1.544283e-32	
## EDAD	18.0101786	5.057365e-72	
## HABITATRURAL	-9.1780385	4.749163e-20	
## HABITATURBANA	-8.4202542	3.972853e-17	
## becaSI	23.5537132	3.112265e-121	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.5447653	5.900800e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	2.1199823	4.268633e-02	
## AREAFORMACIONTÉCNICA	-2.2247601	3.403880e-02	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-8.1031401	5.619734e-16	

En la tabla de efectos fijos se puede observar que todos los efectos fijos son significativos, salvo la variable SEXO, como ya se había indicado. En particular resultan significativos los efectos provocados por la variable AREAFORMACION, dado que las áreas Socio-humanística y Técnica ejercen un efecto distinto al ejercido por la categoría de referencia, área Administrativa.

Para la significación de σ_{u0}^2 , se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis de que $\sigma_{u0}^2 = 0$, es decir,

```
Null.Model.2<-gls(MEDIAPOND~SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + HABITAT +
beca + AREAFORMACION,data=datos.final,control=list(opt="optim"))
logLik(Model.2)*-2 # (-2) veces La Verosimilitud del modelo
## 'log Lik.' 137688.4 (df=13)
logLik(Null.Model.2)*-2 # (-2) veces La Verosimilitud del modelo bajo hipót
esis nula
## 'log Lik.' 137838.6 (df=12)
(logLik(Null.Model.2)*-2) - (logLik(Null.Model.2)*-2) # # Estadístico del TRV
## 'log Lik.' 0 (df=12)
anova(Model.2,Null.Model.2)
##           Model df      AIC      BIC   logLik   Test  L.Ratio p-value
## Model.2      1 13 137714.4 137819.0 -68844.19
## Null.Model.2  2 12 137862.6 137959.1 -68919.28 1 vs 2 150.1781 <.0001
```

Se comprueba que esta componente de la varianza sigue resultando significativa. Además, la comparación del modelo respecto al anterior a través del criterio de Akaike permite concluir la mejora alcanzada.

```
AIC(Model.2)
## [1] 137714.4
AIC(Model.1)
## [1] 137724.5
```

Las conclusiones de este modelo son:

- Todos los efectos fijos incluidos en el modelo son significativos, salvo SEXO.
- La varianza asociada a los intercepts cambia para cada Carrera dependiendo de la variable AREADEFORMACION, resultando significativa además la componente de la varianza asociada.

Modelos ampliados a éste que incluyen variables a nivel de CARRERA (c.EDADMEDIA,...) no permiten disminuir de forma significativa $\sigma_{u_0}^2$ y no resultan significativas las variables incluidas, por lo que no pueden reducir la variabilidad no explicada de los intercepts debido a la CARRERA.

3.4.4. PASO 3: MODELIZACIÓN DE LA VARIABILIDAD DEL COEFICIENTE DE EDAD

3A. ESTUDIO DE LA VARIABILIDAD DEL COEFICIENTE DE EDAD

Se analiza en primer lugar si las pendientes asociadas a la variable de Nivel 1 EDAD dependen del factor CARRERA, como efecto aleatorio, es decir, se analiza el modelo:

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + \gamma_3Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u_1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{0j}, e_{ij}] = 0 \quad \forall (i, j) ; Cov[u_{1j}, e_{ij}] = 0 \quad \forall (i, j)$$

$$Cov(u_{0j}, u_{1j}) = \sigma_{01} \quad \forall j$$

```
Model.3a <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + HABITAT + beca,
random=~EDAD|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.3a)$tTable
```

```

##                               Value   Std.Error   DF
## (Intercept)                   28.06296738 0.423766324 23045
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA -0.29192582 0.431670020 29
## AREAFORMACIONSOCIO-HUMANÍSTICA      0.75055997 0.343835740 29
## AREAFORMACIONTÉCNICA                -1.11847115 0.417075496 29
## SEXOMASCULINO                      -0.06511891 0.077477034 23045
## MODALIDADPRESENCIAL                 1.92682829 0.162427549 23045
## EDAD                                0.07201934 0.008793438 23045
## HABITATRURAL                        -1.70715130 0.185312697 23045
## HABITATURBANA                       -1.49517264 0.175798593 23045
## becaSI                               4.28592822 0.181816985 23045
## SEXOMASCULINO:MODALIDADPRESENCIAL -1.43182595 0.172605759 23045
##                               t-value   p-value
## (Intercept)                   66.2227406 0.000000e+00
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA -0.6762708 5.042270e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA      2.1829027 3.728472e-02
## AREAFORMACIONTÉCNICA                -2.6816995 1.196139e-02
## SEXOMASCULINO                      -0.8404931 4.006407e-01
## MODALIDADPRESENCIAL                 11.8626939 2.297427e-32
## EDAD                                8.1901231 2.743830e-16
## HABITATRURAL                        -9.2122738 3.458013e-20
## HABITATURBANA                       -8.5050319 1.924255e-17
## becaSI                               23.5727603 2.006340e-121
## SEXOMASCULINO:MODALIDADPRESENCIAL -8.2953544 1.141258e-16

```

En la tabla de efectos fijos se puede observar que todos los efectos fijos son significativos, salvo la variable SEXO, como ya se había indicado.

Se obtienen las estimaciones de los parámetros de la estructura de covarianzas, que en este caso son:

- σ_{u0}^2 : varianza asociada al INTERCEPT o efecto aleatorio
- σ_{u1}^2 : varianza asociada al coeficiente de la variable EDAD
- σ_e^2 : varianza asociada al error experimental o varianza residual.
- σ_{01} : covarianza entre los efectos aleatorios de EDAD e intercept.

VarCorr(Model.3a)

```

## CARRERA = pdLogChol(EDAD)
##           Variance   StdDev   Corr
## (Intercept) 1.645271607 1.28268141 (Intr)
## EDAD        0.001083258 0.03291289 -0.851
## Residual    22.658436231 4.76008784

```

Significación de σ_{u1}^2 : Se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis nula, es decir, el modelo que se ha denominado Model.2:

```
anova(Model.3a,Model.2)
##           Model df      AIC      BIC   logLik   Test  L.Ratio p-value
## Model.3a      1 15 137695.2 137815.9 -68832.59
## Model.2       2 13 137714.4 137819.0 -68844.19 1 vs 2 23.21308 <.0001
```

Se comprueba que esta componente de la varianza resulta significativa, por lo que se puede concluir que:

- Todos los efectos fijos incluidos en el modelo son significativos, salvo SEXO.
- La varianza de las pendientes asociadas a EDAD cambian para cada Carrera resultando significativa la componente de la varianza asociada (σ_{u1}^2).

En consecuencia, se debe proceder con la inclusión de variables de Nivel 2 con objeto de tratar de explicar esta variabilidad asociada a los coeficientes de la variable EDAD de los alumnos.

3B. INCLUSIÓN DE VARIABLES DE NIVEL 2 EN LA ALEATORIEDAD DEL COEFICIENTE DE EDAD

Inicialmente se incluye en el modelo anterior la variable de Nivel 2 AREAFORMACION,

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + \gamma_3Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

En este caso, se mantiene la misma estructura de covarianzas.

```
Model.3b <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD + EDAD + HABITAT + beca + EDAD:AREAFORMACION,
random=~EDAD|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.3b)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	27.41421669	0.63854093	23042
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.43987408	1.08581467	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.52098038	0.78026378	29
## AREAFORMACIONTÉCNICA	0.82253339	1.02919097	29
## SEXOMASCULINO	-0.06272944	0.07750571	23042
## MODALIDADPRESENCIAL	1.93447868	0.16271620	23042
## EDAD	0.09297699	0.01852042	23042
## HABITATRURAL	-1.70695021	0.18530263	23042
## HABITATURBANA	-1.49552672	0.17579296	23042
## becaSI	4.28152796	0.18183194	23042
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.44039223	0.17263519	23042
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.02344388	0.03835188	23042
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.02474114	0.02339719	23042
## AREAFORMACIONTÉCNICA:EDAD	-0.07363607	0.03560527	23042
##	t-value	p-value	
## (Intercept)	42.9325914	0.000000e+00	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.4051097	6.883699e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.9493156	6.098820e-02	
## AREAFORMACIONTÉCNICA	0.7992039	4.306697e-01	
## SEXOMASCULINO	-0.8093525	4.183208e-01	
## MODALIDADPRESENCIAL	11.8886669	1.687256e-32	
## EDAD	5.0202432	5.199069e-07	
## HABITATRURAL	-9.2116888	3.476878e-20	
## HABITATURBANA	-8.5073187	1.886822e-17	
## becaSI	23.5466223	3.666119e-121	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-8.3435610	7.608391e-17	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.6112838	5.410178e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-1.0574408	2.903215e-01	
## AREAFORMACIONTÉCNICA:EDAD	-2.0681231	3.863959e-02	

Los efectos fijos se mantienen significativos, en particular los correspondientes al AREAFORMACION en las pendientes de la EDAD (resulta significativo al 5% el cambio de coeficiente del área TECNICA respecto del área de referencia ADMINISTRATIVA).

Se obtienen las estimaciones de los parámetros de la estructura de covarianzas:

VarCorr(Model.3b)

##	CARRERA = pdLogChol(EDAD)	Variance	StdDev	Corr
## (Intercept)		2.043343214	1.42945557	(Intr)
## EDAD		0.001679625	0.04098323	-0.887
## Residual		22.652850423	4.75950107	

Finalmente, se vuelve a contrastar la significación de la componente de la varianza σ_{u1}^2 , dado que en su caso se deberán incluir otras variables de Nivel 2 que puedan reducirla. Para ello, se compara la verosimilitud del modelo actual frente a la verosimilitud del modelo bajo la hipótesis de que $\sigma_{u1}^2 = 0$, es decir,

```
Null.Model.3b<-lme(MEDIAPOND~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD
+ EDAD + HABITAT + beca + EDAD*AREAFORMACION,
random=~1|CARRERA, data=datos.final, control=list(opt="optim"))

anova(Null.Model.3b,Model.3b)

##           Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## Null.Model.3b   1  16 137728.1 137856.8 -68848.05
## Model.3b       2  18 137713.0 137857.8 -68838.51 1 vs 2 19.08206 1e-04
```

Se comprueba que esta componente de la varianza resulta significativa, por lo que se debe proceder con la inclusión de otras variables de Nivel 2. En concreto, se considera la inclusión de la variable porcentaje de mujeres en el profesorado de la carrera cursada por el alumno: C.PORCMUJ (C.PM)

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + \gamma_{14}C.PM_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + \gamma_3Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

En este caso, se mantiene la misma estructura de covarianzas.

```
Model.3C <-lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + EDAD:C.PORCMUJ,
random=~EDAD|CARRERA, data=datos.final, control=list(opt="optim"))

summary(Model.3C)$tTable

##           Value      Std. Error      DF
## (Intercept) 27.4032535741 0.641339971 23041
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA 0.5127373946 1.079979433 29
## AREAFORMACIONSOCIO-HUMANÍSTICA 1.5228136879 0.783579689 29
## AREAFORMACIONTÉCNICA 0.8138850032 1.035004548 29
## SEXOMASCULINO -0.0614422957 0.077537235 23041
## MODALIDADPRESENCIAL 1.9397500179 0.163235378 23041
## EDAD 0.0780928226 0.031413254 23041
## HABITATRURAL -1.7065536636 0.185303946 23041
```

## HABITATURBANA	-1.4948552062	0.175791131	23041
## becaSI	4.2817105363	0.181832932	23041
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.4392589208	0.172701801	23041
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.0228561761	0.037373668	23041
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.0252140221	0.022650424	23041
## AREAFORMACIONTÉCNICA:EDAD	-0.0667108802	0.035021465	23041
## EDAD:C.PORCMUJ	0.0002243887	0.000385325	23041
##	t-value	p-value	
## (Intercept)	42.7281236	0.000000e+00	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.4747659	6.385098e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.9434063	6.172937e-02	
## AREAFORMACIONTÉCNICA	0.7863589	4.380361e-01	
## SEXOMASCULINO	-0.7924231	4.281222e-01	
## MODALIDADPRESENCIAL	11.8831471	1.801781e-32	
## EDAD	2.4859832	1.292644e-02	
## HABITATRURAL	-9.2094836	3.548795e-20	
## HABITATURBANA	-8.5035871	1.948310e-17	
## becaSI	23.5474976	3.593444e-121	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-8.3337806	8.262289e-17	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.6115583	5.408361e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-1.1131810	2.656423e-01	
## AREAFORMACIONTÉCNICA:EDAD	-1.9048569	5.681111e-02	
## EDAD:C.PORCMUJ	0.5823363	5.603459e-01	

En este caso se comprueba como el efecto correspondiente a la variable C.PORCMUJ en las pendientes de la EDAD no resulta significativo (la estimación del coeficiente es 0.000224 con p-valor igual a 0.5603). En consecuencia no se ha de incluir en el modelo. Repitiéndose esta acción con las variables de Nivel 2 (C.EDADMED, C.RETRMED, C.EXPMED), ninguna de ellas resultan significativas, por lo que se puede concluir que se debe mantener el modelo etiquetado como MODELO.3b, de forma que

- Todos los efectos fijos incluidos en el modelo son significativos, salvo SEXO.
- La varianza asociada a las pendientes asociadas a EDAD cambian para cada Carrera, dependiendo de la variable AREADEFORMACION, resultando significativa además la componente de la varianza asociada.

Model.3<-Model.3b

La comparación del modelo respecto al anterior a través del criterio de Akaike permite concluir la mejora alcanzada.

```
AIC(Model.3)
## [1] 137713
AIC(Model.2)
## [1] 137714.4
AIC(Model.1)
## [1] 137724.5
```

3.4.5. PASO 4: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE SEXO

4A. ESTUDIO DE LA VARIABILIDAD DEL COEFICIENTE DE SEXO

Se inicia el estudio incluyendo efectos aleatorios en el coeficiente de la variable SEXO.

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + [\gamma_{20} + u_{2j}]Sexo_{ij} + \gamma_3Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{2j} \sim N(0, \sigma_{u2}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0, 1, 2$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0, 1, 2$$

```
Model.4a <-lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION,
random=~EDAD+SEXO|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.4a)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	27.46565015	0.64620743	23042
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.13333229	1.10689565	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.45914872	0.79117157	29
## AREAFORMACIONTÉCNICA	0.93651562	1.04763896	29
## SEXOMASCULINO	-0.08574941	0.10230875	23042
## MODALIDADPRESENCIAL	1.92990844	0.16299999	23042
## EDAD	0.09528739	0.01917673	23042
## HABITATRURAL	-1.70217477	0.18527893	23042
## HABITATURBANA	-1.49222352	0.17578515	23042

```

## becaSI 4.27771399 0.18181712 23042
## SEXOMASCULINO:MODALIDADPRESENCIAL -1.42859545 0.18607727 23042
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -0.01311145 0.03954178 23042
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -0.02804351 0.02423087 23042
## AREAFORMACIONTÉCNICA:EDAD -0.08613324 0.03669183 23042
## t-value p-value
## (Intercept) 42.5028387 0.000000e+00
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA 0.1204561 9.049526e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA 1.8442886 7.537938e-02
## AREAFORMACIONTÉCNICA 0.8939297 3.787186e-01
## SEXOMASCULINO -0.8381435 4.019588e-01
## MODALIDADPRESENCIAL 11.8399299 3.009726e-32
## EDAD 4.9689082 6.781317e-07
## HABITATRURAL -9.1870929 4.367446e-20
## HABITATURBANA -8.4889055 2.209987e-17
## becaSI 23.5275637 5.685810e-121
## SEXOMASCULINO:MODALIDADPRESENCIAL -7.6774312 1.687484e-14
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -0.3315848 7.402058e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -1.1573465 2.471428e-01
## AREAFORMACIONTÉCNICA:EDAD -2.3474775 1.890943e-02

```

En la tabla de efectos fijos se puede observar que todos los efectos fijos son significativos, salvo la variable SEXO, como ya se había indicado. Se obtienen las estimaciones de los parámetros de la estructura de covarianzas, que en este caso son:

VarCorr(Model.4a)

```

## CARRERA = pdLogChol(EDAD + SEXO)
## Variance StdDev Corr
## (Intercept) 2.099000568 1.4487928 (Intr) EDAD
## EDAD 0.001840573 0.0429019 -0.909
## SEXOMASCULINO 0.060239470 0.2454373 0.081 0.113
## Residual 22.644338230 4.7586068

```

Significación de σ_{u2}^2 : Se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis de que $\sigma_{u2}^2 = 0$, es decir, el modelo que se ha denominado Model.3:

anova(Model.4a,Model.3)

```

## Model df AIC BIC logLik Test L.Ratio p-value
## Model.4a 1 21 137717.9 137886.8 -68837.94
## Model.3 2 18 137713.0 137857.8 -68838.51 1 vs 2 1.139873 0.7675

```

Se comprueba que la varianza de los coeficientes asociados a SEXO no cambian para cada Carrera, es decir, pueden ser considerados como constantes para todas las carreras, por lo que como conclusión de este paso se obtiene que debemos mantener el Modelo.3

Model.4<-Model.3

3.4.6. PASO 5: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE MODALIDAD

5A. ESTUDIO DE LA VARIABILIDAD DEL COEFICIENTE DE MODALIDAD

Se comienza incluyendo efectos aleatorios en el coeficiente de la variable MODALIDAD.

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + [\gamma_{30} + u_{3j}]Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{3j} \sim N(0, \sigma_{u3}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0,1,3$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0,1,3$$

```
Model.5a <-lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION,
random=~EDAD+MODALIDAD|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.5a)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	27.32629311	0.44889445	23042
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.76267983	0.88458340	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.77175639	0.52551463	29
## AREAFORMACIONTÉCNICA	0.27740141	0.89487488	29
## SEXOMASCULINO	-0.05950393	0.07737273	23042
## MODALIDADPRESENCIAL	1.73052097	0.28595054	23042
## EDAD	0.10068108	0.01279497	23042
## HABITATRURAL	-1.70877167	0.18512525	23042

```

## HABITATURBANA -1.49677311 0.17564427 23042
## becaSI 4.29837635 0.18200735 23042
## SEXOMASCULINO:MODALIDADPRESENCIAL -1.44884775 0.17430567 23042
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -0.03867011 0.02818668 23042
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -0.03737716 0.01589418 23042
## AREAFORMACIONTÉCNICA:EDAD -0.05849272 0.02924122 23042
## t-value p-value
## (Intercept) 60.8746511 0.000000e+00
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA 0.8621910 3.956563e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA 3.3714692 2.133145e-03
## AREAFORMACIONTÉCNICA 0.3099890 7.587860e-01
## SEXOMASCULINO -0.7690556 4.418682e-01
## MODALIDADPRESENCIAL 6.0518191 1.454307e-09
## EDAD 7.8688002 3.737363e-15
## HABITATRURAL -9.2303545 2.923201e-20
## HABITATURBANA -8.5216164 1.668473e-17
## becaSI 23.6164992 7.314517e-122
## SEXOMASCULINO:MODALIDADPRESENCIAL -8.3121092 9.915054e-17
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -1.3719285 1.700991e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -2.3516262 1.869992e-02
## AREAFORMACIONTÉCNICA:EDAD -2.0003512 4.547407e-02

```

En la tabla de efectos fijos se puede observar que todos los efectos fijos son significativos, salvo la variable SEXO, como ya se había indicado. Se obtienen las estimaciones de los parámetros de la estructura de covarianzas, que en este caso son:

VarCorr(Model.5a)

```

## CARRERA = pdLogChol(EDAD + MODALIDAD)
## Variance StdDev Corr
## (Intercept) 6.483626e-01 0.80520965 (Intr) EDAD
## EDAD 6.777989e-04 0.02603457 -0.841
## MODALIDADPRESENCIAL 1.125162e+00 1.06073668 0.359 -0.675
## Residual 2.262508e+01 4.75658288

```

Significación de σ_{u3}^2 : Se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis nula, es decir, el modelo que se ha denominado Model.4:

anova(Model.5a,Model.4)

```

## Model df AIC BIC logLik Test L.Ratio p-value
## Model.5a 1 21 137688.1 137857.1 -68823.04
## Model.4 2 18 137713.0 137857.8 -68838.51 1 vs 2 30.92128 <.0001

```

Se comprueba que la varianza de los coeficientes asociados a MODALIDAD cambian para cada Carrera. Se puede concluir que:

- Todos los efectos fijos incluidos en el modelo son significativos, salvo SEXO.
- La varianza asociada a las pendientes asociadas a EDAD cambian para cada Carrera dependiendo de la variable AREADEFORMACION.
- La varianza de las pendientes asociadas a MODALIDAD cambian para cada Carrera resultado significativa la componente de la varianza asociada (σ_{u3}^2).

En consecuencia, se debe proceder con la inclusión de variables de Nivel 2 con objeto de tratar de explicar esta variabilidad asociada a los coeficientes de la variable MODALIDAD elegida por los alumnos.

5B. INCLUSIÓN DE VARIABLES DE NIVEL 2 EN LA ALEATORIEDAD DEL COEFICIENTE DE MODALIDAD

Inicialmente se incluye la variable AREAFORMACION, resultando el siguiente modelo:

$$\begin{aligned}
 y_{ij} = & [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] \\
 & + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} \\
 & + [\gamma_{30} + \gamma_{31}AFbb_j + \gamma_{32}AFsh_j + \gamma_{33}AFtc_j + u_{3j}]Mod_{ij} + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} \\
 & + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}
 \end{aligned}$$

con idéntica estructura de varianzas y covarianzas

```

Model.5b <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:AREAFORMACION,
random=~EDAD+MODALIDAD|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.5b)$tTable
    
```

##	Value
## (Intercept)	27.39240312
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.58725852
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.68302851
## AREAFORMACIONTÉCNICA	-0.26279621
## SEXOMASCULINO	-0.05446379
## MODALIDADPRESENCIAL	1.89675498

## EDAD	0.09834488	
## HABITATRURAL	-1.70856225	
## HABITATURBANA	-1.49771416	
## becaSI	4.29546679	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.45556725	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.03607981	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.03250371	
## AREAFORMACIONTÉCNICA:EDAD	-0.05338359	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:MODALIDADPRESENCIAL	0.05876517	
## AREAFORMACIONSOCIO-HUMANÍSTICA:MODALIDADPRESENCIAL	-0.84986703	
## AREAFORMACIONTÉCNICA:MODALIDADPRESENCIAL	0.44084426	
##	Std. Error	DF
## (Intercept)	0.43266165	23040
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.93778479	28
## AREAFORMACIONSOCIO-HUMANÍSTICA	0.50279191	28
## AREAFORMACIONTÉCNICA	0.94707576	28
## SEXOMASCULINO	0.07752741	23040
## MODALIDADPRESENCIAL	0.53683415	23040
## EDAD	0.01329616	23040
## HABITATRURAL	0.18513541	23040
## HABITATURBANA	0.17566047	23040
## becaSI	0.18204880	23040
## SEXOMASCULINO:MODALIDADPRESENCIAL	0.17502189	23040
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	0.02906747	23040
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	0.01668556	23040
## AREAFORMACIONTÉCNICA:EDAD	0.02944557	23040
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:MODALIDADPRESENCIAL	0.86299841	23040
## AREAFORMACIONSOCIO-HUMANÍSTICA:MODALIDADPRESENCIAL	0.74371873	23040
## AREAFORMACIONTÉCNICA:MODALIDADPRESENCIAL	0.86251489	28
##	t-value	
## (Intercept)	63.31137247	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.62621885	
## AREAFORMACIONSOCIO-HUMANÍSTICA	3.34736592	
## AREAFORMACIONTÉCNICA	-0.27748171	
## SEXOMASCULINO	-0.70251016	
## MODALIDADPRESENCIAL	3.53322337	
## EDAD	7.39648544	
## HABITATRURAL	-9.22871674	
## HABITATURBANA	-8.52618771	
## becaSI	23.59513981	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-8.31648696	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-1.24124352	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-1.94801457	
## AREAFORMACIONTÉCNICA:EDAD	-1.81295866	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:MODALIDADPRESENCIAL	0.06809418	
## AREAFORMACIONSOCIO-HUMANÍSTICA:MODALIDADPRESENCIAL	-1.14272641	
## AREAFORMACIONTÉCNICA:MODALIDADPRESENCIAL	0.51111496	
##	p-value	
## (Intercept)	0.000000e+00	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	5.362436e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	2.338764e-03	

## AREAFORMACIONTÉCNICA	7.834487e-01
## SEXOMASCULINO	4.823682e-01
## MODALIDADPRESENCIAL	4.113287e-04
## EDAD	1.446146e-13
## HABITATRURAL	2.968092e-20
## HABITATURBANA	1.604079e-17
## becaSI	1.198103e-121
## SEXOMASCULINO:MODALIDADPRESENCIAL	9.556860e-17
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	2.145284e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	5.142535e-02
## AREAFORMACIONTÉCNICA:EDAD	6.985120e-02
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:MODALIDADPRESENCIAL	9.457113e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:MODALIDADPRESENCIAL	2.531641e-01
## AREAFORMACIONTÉCNICA:MODALIDADPRESENCIAL	6.132775e-01

En este caso se comprueba como el efecto correspondiente a la variable AREAFORMACION en los coeficientes de la MODALIDAD no resulta significativo, por lo que no se ha de incluir en el modelo.

Dado el resultado del modelo anterior, se repite el análisis sustituyendo la variable AREAFORMACION por C.EDADMEDIA (*C.Edm*):

$$\begin{aligned}
 y_{ij} = & [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] \\
 & + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} \\
 & + [\gamma_{30}Pres_{ij} + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + u_{3j}] + \gamma_4Sex * Mod_{ij} \\
 & + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}
 \end{aligned}$$

con idéntica estructura de varianzas y covarianzas

```

Model.5b <-lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED,
random=~EDAD+MODALIDAD|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.5b)$tTable

```

##	Value	Std.Error	DF
## (Intercept)	26.63090240	1.65394442	23040
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.76231390	0.84986457	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.51290873	0.51012529	29
## AREAFORMACIONTÉCNICA	-0.06644386	0.84396146	29
## SEXOMASCULINO	-0.05707024	0.07744128	23040
## MODALIDADPRESENCIAL	-4.35018565	3.80174817	23040
## EDAD	0.09793641	0.01291703	23040
## HABITATRURAL	-1.70917265	0.18516206	23040
## HABITATURBANA	-1.49776761	0.17565823	23040

```

## becaSI 4.29820979 0.18202154 23040
## SEXOMASCULINO:MODALIDADPRESENCIAL -1.44612617 0.17433977 23040
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -0.03752528 0.02911738 23040
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -0.03393577 0.01615737 23040
## AREAFORMACIONTÉCNICA:EDAD -0.05213230 0.02884609 23040
## MODALIDADDISTANCIA:C.EDADMED 0.02245283 0.04305906 23040
## MODALIDADPRESENCIAL:C.EDADMED 0.18546268 0.08924280 23040
## t-value p-value
## (Intercept) 16.10144946 5.168356e-58
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA 0.89698280 3.771144e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA 2.96575912 5.988577e-03
## AREAFORMACIONTÉCNICA -0.07872854 9.377893e-01
## SEXOMASCULINO -0.73694853 4.611612e-01
## MODALIDADPRESENCIAL -1.14425929 2.525280e-01
## EDAD 7.58196042 3.532140e-14
## HABITATRURAL -9.23068513 2.914245e-20
## HABITATURBANA -8.52660090 1.598380e-17
## becaSI 23.61374315 7.797513e-122
## SEXOMASCULINO:MODALIDADPRESENCIAL -8.29487232 1.145896e-16
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -1.28875883 1.974949e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -2.10032765 3.571087e-02
## AREAFORMACIONTÉCNICA:EDAD -1.80725706 7.073525e-02
## MODALIDADDISTANCIA:C.EDADMED 0.52144274 6.020634e-01
## MODALIDADPRESENCIAL:C.EDADMED 2.07818100 3.770373e-02

```

Se comprueba que el efecto de la variable de Nivel 2 C.EDADMEDIA (edad media del profesorado) en los coeficientes de la MODALIDAD resulta significativo (estimación del coeficiente asociado a C.EDADMEDIA en la Modalidad Presencial es 0.185463, con un p-valor asociado de 0.0377). Tras la inclusión de este efecto fijo en la explicación de la aleatoriedad del coeficiente de MODALIDAD, la estructura de covarianzas resultantes es la que a continuación se obtiene:

VarCorr(Model.5b)

```

## CARRERA = pdLogChol(EDAD + MODALIDAD)
## Variance StdDev Corr
## (Intercept) 5.272183e-01 0.72609797 (Intr) EDAD
## EDAD 6.006203e-04 0.02450756 -0.772
## MODALIDADPRESENCIAL 1.278762e+00 1.13082352 -0.045 -0.439
## Residual 2.262511e+01 4.75658548

```

Para comprobar si aún debe considerarse significativa la componente de la varianza asociada al coeficiente de modalidad, se procede a realizar el test de razón de verosimilitudes:

```
Null.Model.5b<-lme(MEDIAPOND~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD
+ EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED,
random=~EDAD|CARRERA, data=datos.final, control=list(opt="optim"))

anova(Null.Model.5b,Model.5b)

##           Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## Null.Model.5b     1 20 137720.0 137880.9 -68840.01
## Model.5b          2 23 137695.9 137881.0 -68824.94 1 vs 2 30.12212 <.0001
```

Dado que se concluye la significación de dicha componente de la varianza, se debe proceder con la inclusión de otras variables de Nivel 2 con objeto de tratar de explicar esta variabilidad. Así, se amplía el modelo anterior con la inclusión de la variable de Nivel 2 C.RETRMED en la explicación de la variabilidad de los coeficientes de MODALIDAD.

```
Model.5c <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED + MODALIDAD:C.
RETRMED,
random=~EDAD+MODALIDAD|CARRERA, data=datos.final, control=list(opt="optim"))

summary(Model.5c)$tTable

##           Value      Std. Error    DF
## (Intercept)      2.630941e+01  1.821647e+00 23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA      1.007857e+00  8.957093e-01    29
## AREAFORMACIONSOCIO-HUMANÍSTICA      1.451689e+00  5.558265e-01    29
## AREAFORMACIONTÉCNICA      3.011246e-01  8.983259e-01    29
## SEXOMASCULINO      -6.157704e-02  7.750468e-02 23038
## MODALIDADPRESENCIAL      -7.441388e-01  4.110063e+00 23038
## EDAD      9.582054e-02  1.282543e-02 23038
## HABITATRURAL      -1.708679e+00  1.851742e-01 23038
## HABITATURBANA      -1.499018e+00  1.756585e-01 23038
## becaSI      4.292809e+00  1.820275e-01 23038
## SEXOMASCULINO:MODALIDADPRESENCIAL      -1.432842e+00  1.745813e-01 23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD      -3.965452e-02  2.838296e-02 23038
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD      -3.213381e-02  1.594159e-02 23038
## AREAFORMACIONTÉCNICA:EDAD      -5.284642e-02  2.864411e-02 23038
## MODALIDADDISTANCIA:C.EDADMED      2.054811e-02  4.665771e-02 23038
## MODALIDADPRESENCIAL:C.EDADMED      1.340234e-01  8.815498e-02 23038
## MODALIDADDISTANCIA:C.RETRMED      4.745343e-05  5.958184e-05 23038
## MODALIDADPRESENCIAL:C.RETRMED      -1.283820e-04  8.210090e-05 23038
##           t-value      p-value
## (Intercept)      14.4426525  4.479435e-47
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA      1.1252050  2.697289e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA      2.6117664  1.411793e-02
## AREAFORMACIONTÉCNICA      0.3352064  7.398822e-01
## SEXOMASCULINO      -0.7944944  4.269158e-01
## MODALIDADPRESENCIAL      -0.1810529  8.563276e-01
## EDAD      7.4711375  8.233280e-14
```

```
## HABITATRURAL -9.2274142 3.004285e-20
## HABITATURBANA -8.5337064 1.503457e-17
## becaSI 23.5833056 1.574734e-121
## SEXOMASCULINO:MODALIDADPRESENCIAL -8.2073065 2.379443e-16
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -1.3971244 1.623896e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -2.0157223 4.384061e-02
## AREAFORMACIONTÉCNICA:EDAD -1.8449313 6.506036e-02
## MODALIDADDISTANCIA:C.EDADMED 0.4404011 6.596507e-01
## MODALIDADPRESENCIAL:C.EDADMED 1.5203156 1.284454e-01
## MODALIDADDISTANCIA:C.RETRMED 0.7964411 4.257839e-01
## MODALIDADPRESENCIAL:C.RETRMED -1.5637101 1.178994e-01
```

Dado que el coeficiente asociado a C.RETRMED dentro de la variable de Nivel 1 Modalidad no es significativo, se elimina esta variable del modelo y se sustituye por la variable C.PORCMUJ.

```
Model.5c <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED + MODALIDAD:C.
PORCMUJ,
random=~EDAD+MODALIDAD|CARRERA, data=datos.final, control=list(opt="optim"))
summary(Model.5c)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	27.87091925	1.94109138	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.38213177	0.74043385	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.48150027	0.40278573	29
## AREAFORMACIONTÉCNICA	-0.99103883	0.77153781	29
## SEXOMASCULINO	-0.06131503	0.07771399	23038
## MODALIDADPRESENCIAL	-5.51431388	3.93663637	23038
## EDAD	0.09670055	0.01243601	23038
## HABITATRURAL	-1.70517236	0.18515377	23038
## HABITATURBANA	-1.49640067	0.17567116	23038
## becaSI	4.29952740	0.18203687	23038
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.45243908	0.17489325	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.03600413	0.02889062	23038
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.03276180	0.01547547	23038
## AREAFORMACIONTÉCNICA:EDAD	-0.04555247	0.02818609	23038
## MODALIDADDISTANCIA:C.EDADMED	0.02700482	0.04446418	23038
## MODALIDADPRESENCIAL:C.EDADMED	0.23508437	0.09021371	23038
## MODALIDADDISTANCIA:C.PORCMUJ	-0.01995370	0.01201299	23038
## MODALIDADPRESENCIAL:C.PORCMUJ	-0.03024318	0.01501424	23038
##	t-value	p-value	
## (Intercept)	14.3583757	1.499992e-46	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.5160917	6.097033e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	3.6781349	9.513522e-04	
## AREAFORMACIONTÉCNICA	-1.2844981	2.091342e-01	
## SEXOMASCULINO	-0.7889833	4.301299e-01	
## MODALIDADPRESENCIAL	-1.4007679	1.612970e-01	
## EDAD	7.7758510	7.807083e-15	
## HABITATRURAL	-9.2094931	3.548520e-20	
## HABITATURBANA	-8.5181919	1.718400e-17	

```
## becaSI 23.6189924 6.909066e-122
## SEXOMASCULINO:MODALIDADPRESENCIAL -8.3047176 1.055031e-16
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -1.2462221 2.126955e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -2.1170143 3.426932e-02
## AREAFORMACIONTÉCNICA:EDAD -1.6161330 1.060792e-01
## MODALIDADDISTANCIA:C.EDADMED 0.6073387 5.436321e-01
## MODALIDADPRESENCIAL:C.EDADMED 2.6058609 9.170263e-03
## MODALIDADDISTANCIA:C.PORCMUJ -1.6610097 9.672510e-02
## MODALIDADPRESENCIAL:C.PORCMUJ -2.0143001 4.398963e-02
```

Se concluye que resulta significativa la variable C.PORCMUJ en el efecto aleatorio en el coeficiente de MODALIDAD, por lo que se ha de incluir en el modelo. Tras la inclusión de este efecto fijo, la estructura de covarianzas resultantes es la que a continuación se obtiene:

`VarCorr(Model.5c)`

```
## CARRERA = pdLogChol(EDAD + MODALIDAD)
##          Variance StdDev Corr
## (Intercept) 2.019893e-01 0.44943224 (Intr) EDAD
## EDAD 5.305722e-04 0.02303415 -0.572
## MODALIDADPRESENCIAL 1.418091e+00 1.19083625 -0.213 -0.442
## Residual 2.262427e+01 4.75649717
```

Para comprobar si aún debe considerarse significativa la componente de la varianza en cuestión, se procede a realizar el test de razón de verosimilitudes:

```
Null.Model.5c<-lme(MEDIAPOND~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD
+ EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED+MODALIDAD:C.
PORCMUJ,
random=~EDAD|CARRERA, data=datos.final, control=list(opt="optim"))
```

`anova(Null.Model.5c,Model.5c)`

```
##          Model df      AIC      BIC    logLik  Test L.Ratio p-value
## Null.Model.5c    1 22 137731.2 137908.2 -68843.59
## Model.5c        2 25 137709.8 137911.0 -68829.92 1 vs 2 27.3379 <.0001
```

Dado que se concluye la significación de dicha componente de la varianza, se debe proceder con la inclusión de otras variables de Nivel 2 con objeto de tratar de explicar esta variabilidad. Se amplía el modelo anterior con la inclusión de la variable de Nivel 2 C.EXP MED en la explicación de la variabilidad de los coeficientes de MODALIDAD.

```
Model.5d <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED + MODALIDAD:C.
PORCMUJ + MODALIDAD:C.EXPMED,
random=~EDAD+MODALIDAD|CARRERA, data=datos.final, control=list(opt="optim"))
```

```
summary(Model.5d)$tTable
```

	Value	Std.Error	DF
## (Intercept)	27.36987177	2.20796102	23036
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.38189474	0.73721543	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.42113398	0.40609018	29
## AREAFORMACIONTÉCNICA	-1.00369340	0.77625768	29
## SEXOMASCULINO	-0.06013301	0.07773661	23036
## MODALIDADPRESENCIAL	-4.44571608	4.75682441	23036
## EDAD	0.09588249	0.01212292	23036
## HABITATRURAL	-1.70464115	0.18516399	23036
## HABITATURBANA	-1.49589220	0.17567372	23036
## becaSI	4.30014742	0.18205185	23036
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.45337209	0.17493882	23036
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.03688606	0.02832650	23036
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.03104370	0.01513898	23036
## AREAFORMACIONTÉCNICA:EDAD	-0.04218652	0.02764076	23036
## MODALIDADDISTANCIA:C.EDADMED	0.05900340	0.06601261	23036
## MODALIDADPRESENCIAL:C.EDADMED	0.22881769	0.09642571	23036
## MODALIDADDISTANCIA:C.PORCMUJ	-0.02041840	0.01269378	23036
## MODALIDADPRESENCIAL:C.PORCMUJ	-0.02985471	0.01581560	23036
## MODALIDADDISTANCIA:C.EXPMED	-0.09573121	0.13547481	23036
## MODALIDADPRESENCIAL:C.EXPMED	-0.05317978	0.19959476	23036
##	t-value	p-value	
## (Intercept)	12.3959941	3.557295e-35	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.5180233	6.083719e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	3.4995526	1.526240e-03	
## AREAFORMACIONTÉCNICA	-1.2929900	2.062208e-01	
## SEXOMASCULINO	-0.7735482	4.392060e-01	
## MODALIDADPRESENCIAL	-0.9345975	3.500056e-01	
## EDAD	7.9091906	2.706502e-15	
## HABITATRURAL	-9.2061158	3.661520e-20	
## HABITATURBANA	-8.5151736	1.763619e-17	
## becaSI	23.6204542	6.681551e-122	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-8.3078878	1.027317e-16	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-1.3021750	1.928696e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-2.0505811	4.031905e-02	
## AREAFORMACIONTÉCNICA:EDAD	-1.5262430	1.269631e-01	
## MODALIDADDISTANCIA:C.EDADMED	0.8938201	3.714275e-01	
## MODALIDADPRESENCIAL:C.EDADMED	2.3729946	1.765268e-02	
## MODALIDADDISTANCIA:C.PORCMUJ	-1.6085355	1.077316e-01	
## MODALIDADPRESENCIAL:C.PORCMUJ	-1.8876744	5.908224e-02	
## MODALIDADDISTANCIA:C.EXPMED	-0.7066348	4.798006e-01	
## MODALIDADPRESENCIAL:C.EXPMED	-0.2664387	7.899037e-01	

El resultado del modelo anterior indica que la variable C.EXP MED no resulta significativa, por lo que se puede concluir este paso. El modelo resultante es:

$$\begin{aligned}
 y_{ij} = & [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] \\
 & + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} \\
 & + [\gamma_{30}Pres_{ij} \\
 & + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + \gamma_{33}(C.Pmu_j \times Pres_{ij}) \\
 & + \gamma_{34}(C.Pmu_j \times Dis_{ij}) + u_{3j}] + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} \\
 & + e_{ij}
 \end{aligned}$$

siendo las variables indicadoras de Modalidad:

- *Pres*=1 (Presencial), 0(Distancia)
- *Dis* =1 (Distancia), 0 (Presencial).

Las conclusiones son:

- Todos los efectos fijos incluidos en el modelo son significativos, salvo SEXO.
- La varianza asociada a las pendientes asociadas a EDAD cambian para cada Carrera dependiendo de la variable AREADEFORMACION.
- La varianza asociada a los coeficientes asociados a SEXO no cambian para cada Carrera.
- La varianza asociada a los coeficientes asociados a MODALIDAD cambian para cada Carrera dependiendo de las variables C.EDAD MEDIA y C.PORCMUJ.

Model.5<-Model.5C

La comparación del modelo respecto al anterior a través del criterio de Akaike permite concluir la mejora alcanzada.

```

AIC(Model.5)
## [1] 137709.8
AIC(Model.4)
## [1] 137713
AIC(Model.3)
## [1] 137713
AIC(Model.2)
## [1] 137714.4
AIC(Model.1)
## [1] 137724.5
    
```

3.4.7. PASO 6: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE SEXO*MOD

6A. ESTUDIO DE LA VARIABILIDAD DEL COEFICIENTE DE SEXO*MODALIDAD

Se inicia el paso incluyendo efectos aleatorios en el coeficiente de la variable interacción SEXO*MODALIDAD.

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + [\gamma_{30}Pres_{ij} + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + \gamma_{33}(C.Pmuj_j \times Pres_{ij}) + \gamma_{34}(C.Pmuj_j \times Dis_{ij}) + u_{3j}] + [\gamma_{40} + u_{4j}]Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{3j} \sim N(0, \sigma_{u3}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{4j} \sim N(0, \sigma_{u4}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0, 1, 3, 4$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0, 1, 3, 4$$

```
Model.6a <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED + MODALIDAD:C.
PORCMUJ, random=~EDAD+MODALIDAD+SEXO*MODALIDAD|CARRERA, data=datos.final, control
=list(opt="optim"))
summary(Model.6a)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	28.27641998	1.97993597	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.13367729	0.80172998	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.61335440	0.45106280	29
## AREAFORMACIONTÉCNICA	-0.73761476	0.81446927	29
## SEXOMASCULINO	-0.08945554	0.09701663	23038
## MODALIDADPRESENCIAL	-5.14918877	3.80144091	23038
## EDAD	0.09795287	0.01346125	23038
## HABITATRURAL	-1.70334565	0.18511393	23038
## HABITATURBANA	-1.49451704	0.17563767	23038

## becaSI	4.29435749	0.18205362	23038
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.51323250	0.22107034	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.02931360	0.03071812	23038
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.03604334	0.01680874	23038
## AREAFORMACIONTÉCNICA:EDAD	-0.05642837	0.02976348	23038
## MODALIDADDISTANCIA:C.EDADMED	0.01682986	0.04504973	23038
## MODALIDADPRESENCIAL:C.EDADMED	0.21133385	0.08549949	23038
## MODALIDADDISTANCIA:C.PORCMUJ	-0.02073921	0.01234502	23038
## MODALIDADPRESENCIAL:C.PORCMUJ	-0.02761113	0.01463457	23038
##	t-value	p-value	
## (Intercept)	14.2814820	4.491236e-46	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.1667360	8.687349e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	3.5767844	1.245056e-03	
## AREAFORMACIONTÉCNICA	-0.9056385	3.725904e-01	
## SEXOMASCULINO	-0.9220640	3.565049e-01	
## MODALIDADPRESENCIAL	-1.3545361	1.755787e-01	
## EDAD	7.2766546	3.531565e-13	
## HABITATRURAL	-9.2016071	3.817876e-20	
## HABITATURBANA	-8.5090916	1.858306e-17	
## becaSI	23.5884215	1.399419e-121	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-6.8450273	7.838102e-12	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.9542773	3.399533e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-2.1443216	3.201756e-02	
## AREAFORMACIONTÉCNICA:EDAD	-1.8958927	5.798673e-02	
## MODALIDADDISTANCIA:C.EDADMED	0.3735839	7.087173e-01	
## MODALIDADPRESENCIAL:C.EDADMED	2.4717559	1.345230e-02	
## MODALIDADDISTANCIA:C.PORCMUJ	-1.6799656	9.297756e-02	
## MODALIDADPRESENCIAL:C.PORCMUJ	-1.8867061	5.921243e-02	

Se obtienen las estimaciones de los parámetros de la estructura de covarianzas, que en este caso son:

- σ_{u0}^2 : varianza asociada al INTERCEPT o efecto aleatorio
- σ_{u1}^2 : varianza asociada al coeficiente de la variable EDAD
- σ_{u3}^2 : varianza asociada al coeficiente de la variable MODALIDAD
- σ_{u4}^2 : varianza asociada al coeficiente de la variable interacción SEXO*MODALIDAD
- σ_e^2 : varianza asociada al error experimental o varianza residual.
- σ_{kd} : covarianzas entre los errores aleatorios asociados a las variables

VarCorr(Model.6a)

```
## CARRERA = pdLogChol(EDAD + MODALIDAD + SEXO * MODALIDAD)
##
## Variance StdDev Corr
## (Intercept) 0.361361778 0.60113374 (Intr) EDAD
## EDAD 0.000696103 0.02638376 -0.701
## MODALIDADPRESENCIAL 1.159793725 1.07693720 -0.114 -0.420
## SEXOMASCULINO 0.040547054 0.20136299 -0.397 0.225
## MODALIDADPRESENCIAL:SEXOMASCULINO 0.356022610 0.59667630 0.466 -0.297
## Residual 22.609410517 4.75493539
##
## MODALIDADPRESENCIAL SEXOMASC
## EDAD
## MODALIDADPRESENCIAL
## SEXOMASCULINO -0.087
## MODALIDADPRESENCIAL:SEXOMASCULINO 0.103 -0.661
## Residual
```

Significación de σ_{u4}^2 : Se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis de que $\sigma_{u4}^2 = 0$.

anova(Model.6a,Model.5)

```
## Model df AIC BIC logLik Test L.Ratio p-value
## Model.6a 1 34 137725.7 137999.3 -68828.85
## Model.5 2 25 137709.8 137911.0 -68829.92 1 vs 2 2.139083 0.9891
```

Se comprueba que la varianza de los coeficientes asociados a SEXO*MODALIDAD no cambian para cada Carrera, pueden ser considerado como constantes para todas las carreras, por lo que como conclusión de este paso se obtiene que debemos mantener el Modelo.5

Model.6<-Model.5

3.4.8. PASO 7: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE HÁBITAT

7A. ESTUDIO DE LA VARIABILIDAD DEL COEFICIENTE DE HÁBITAT

Se inicia el estudio incluyendo efectos aleatorios en los coeficientes de la variable.

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + [\gamma_{30}Pres_{ij} + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + \gamma_{33}(C.Pmuj_j \times Pres_{ij}) + \gamma_{34}(C.Pmuj_j \times Dis_{ij}) + u_{3j}] + \gamma_4Sex * Mod_{ij} + [\gamma_{50} + u_{5j}]Rur_{ij} + [\gamma_{60} + u_{6j}]Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{3j} \sim N(0, \sigma_{u3}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{5j} \sim N(0, \sigma_{u5}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{6j} \sim N(0, \sigma_{u6}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0, 1, 3, 5, 6$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0, 1, 3, 5, 6$$

```
Model.7a <- lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED + MODALIDAD:C.
PORCMUJ, random=~EDAD+MODALIDAD+HABITAT|CARRERA, data=datos.final, control=list
(opt="optim"))
summary(Model.7a)$tTable
```

##	Value	Std.Error	DF
## (Intercept)	28.28532463	1.95185287	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.51279575	0.80076503	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.64521342	0.45527898	29
## AREAFORMACIONTÉCNICA	-0.67315435	0.82899941	29
## SEXOMASCULINO	-0.06135144	0.07770170	23038
## MODALIDADPRESENCIAL	-5.24473798	3.84732911	23038
## EDAD	0.09839882	0.01237807	23038
## HABITATRURAL	-1.73387474	0.20833674	23038
## HABITATURBANA	-1.51403258	0.20953459	23038
## becaSI	4.29865300	0.18201360	23038
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.45280336	0.17483691	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.03831687	0.02859025	23038
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.03463981	0.01527711	23038

```
## AREAFORMACIONTÉCNICA:EDAD -0.04762728 0.02795870 23038
## MODALIDADDISTANCIA:C.EDADMED 0.00910342 0.04397875 23038
## MODALIDADPRESENCIAL:C.EDADMED 0.20532902 0.08824602 23038
## MODALIDADDISTANCIA:C.PORCMUJ -0.01727596 0.01211195 23038
## MODALIDADPRESENCIAL:C.PORCMUJ -0.02455543 0.01464987 23038
## t-value p-value
## (Intercept) 14.4915250 2.215559e-47
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA 0.6403823 5.269520e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA 3.6136380 1.129279e-03
## AREAFORMACIONTÉCNICA -0.8120082 4.234024e-01
## SEXOMASCULINO -0.7895765 4.297832e-01
## MODALIDADPRESENCIAL -1.3632153 1.728280e-01
## EDAD 7.9494505 1.958883e-15
## HABITATRURAL -8.3224627 9.088417e-17
## HABITATURBANA -7.2256929 5.140749e-13
## becaSI 23.6172080 7.199818e-122
## SEXOMASCULINO:MODALIDADPRESENCIAL -8.3094774 1.013685e-16
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -1.3402076 1.801911e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD -2.2674322 2.337306e-02
## AREAFORMACIONTÉCNICA:EDAD -1.7034871 8.849044e-02
## MODALIDADDISTANCIA:C.EDADMED 0.2069959 8.360149e-01
## MODALIDADPRESENCIAL:C.EDADMED 2.3267794 1.998563e-02
## MODALIDADDISTANCIA:C.PORCMUJ -1.4263570 1.537789e-01
## MODALIDADPRESENCIAL:C.PORCMUJ -1.6761534 9.372172e-02
```

Se obtienen las estimaciones de los parámetros de la estructura de covarianzas:

VarCorr(Model.7a)

```
## CARRERA = pdLogChol(EDAD + MODALIDAD + HABITAT)
## Variance StdDev Corr
## (Intercept) 6.991524e-01 0.83615335 (Intr) EDAD MODALI HABITATR
## EDAD 5.385503e-04 0.02320669 -0.797
## MODALIDADPRESENCIAL 1.160898e+00 1.07744994 0.291 -0.457
## HABITATRURAL 1.146498e-01 0.33859979 -0.664 0.629 -0.650
## HABITATURBANA 1.737463e-01 0.41682888 -0.775 0.627 -0.590 0.887
## Residual 2.261723e+01 4.75575712
```

La significación de las componentes de varianza asociadas a la variable hábitat se comprueba a través de la comparación de la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis nula:

anova(Model.7a,Model.6)

```
## Model df AIC BIC logLik Test L.Ratio p-value
## Model.7a 1 34 137726.0 137999.6 -68829.02
## Model.6 2 25 137709.8 137911.0 -68829.92 1 vs 2 1.783198 0.9945
```

Dado que no resulta significativa, se mantiene el modelo anterior.

Model .7 < - Model .6

3.4.9. PASO 8: MODELIZACION DE LA VARIABILIDAD DEL COEFICIENTE DE LA VARIABLE BECA

8A. ESTUDIO DE LA VARIABILIDAD DEL COEFICIENTE DE BECA

Como en pasos anteriores, se inicia el estudio incluyendo efectos aleatorios en el coeficiente de la variable BECA

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_{2}Sexo_{ij} + [\gamma_{30}Pres_{ij} + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + \gamma_{33}(C.Pmuj_j \times Pres_{ij}) + \gamma_{34}(C.Pmuj_j \times Dis_{ij}) + u_{3j}] + \gamma_{40}Sex * Mod_{ij} + \gamma_{5}Rur_{ij} + \gamma_{6}Urb_{ij} + [\gamma_{70} + u_{7j}]Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{3j} \sim N(0, \sigma_{u3}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{7j} \sim N(0, \sigma_{u7}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0, 1, 3, 7$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0, 1, 3, 7$$

```
Model.8a <-lme(MEDIAPOND ~ AREAFORMACION + SEXO + MODALIDAD + SEXO*MODALIDAD +
EDAD + HABITAT + beca + EDAD:AREAFORMACION + MODALIDAD:C.EDADMED + MODALIDAD:C.
PORCMUJ, random=~EDAD+MODALIDAD+beca|CARRERA, data=datos.final, control=list(opt=
"optim"))
summary(Model.8a)$tTable
```

	Value	Std.Error	DF
## (Intercept)	28.49101588	1.97658608	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.18110207	0.87937524	29
## AREAFORMACIONSOCIO-HUMANÍSTICA	1.69521280	0.49714333	29
## AREAFORMACIONTÉCNICA	-0.49146471	0.87879640	29
## SEXOMASCULINO	-0.05934015	0.07769756	23038
## MODALIDADPRESENCIAL	-5.54795365	3.81982357	23038
## EDAD	0.09512917	0.01568298	23038
## HABITATRURAL	-1.71201466	0.18508384	23038
## HABITATURBANA	-1.50677916	0.17560334	23038
## becaSI	4.19243473	0.24833698	23038
## SEXOMASCULINO:MODALIDADPRESENCIAL	-1.45842589	0.17492442	23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.01526144	0.03500495	23038
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.03986991	0.01937559	23038
## AREAFORMACIONTÉCNICA:EDAD	-0.06737981	0.03301624	23038
## MODALIDADDISTANCIA:C. EDADMED	0.01011229	0.04482111	23038
## MODALIDADPRESENCIAL:C. EDADMED	0.22321207	0.08683717	23038
## MODALIDADDISTANCIA:C. PORCMUJ	-0.01918055	0.01229424	23038
## MODALIDADPRESENCIAL:C. PORCMUJ	-0.03204809	0.01462734	23038
##	t-value	p-value	
## (Intercept)	14.4142551	6.736167e-47	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.2059440	8.382736e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA	3.4099075	1.930009e-03	
## AREAFORMACIONTÉCNICA	-0.5592475	5.802866e-01	
## SEXOMASCULINO	-0.7637324	4.450345e-01	
## MODALIDADPRESENCIAL	-1.4524110	1.464010e-01	
## EDAD	6.0657594	1.333783e-09	
## HABITATRURAL	-9.2499412	2.435892e-20	
## HABITATURBANA	-8.5805839	1.002583e-17	
## becaSI	16.8820392	1.471224e-63	
## SEXOMASCULINO:MODALIDADPRESENCIAL	-8.3374633	8.009876e-17	
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.4359793	6.628558e-01	
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-2.0577389	3.962643e-02	
## AREAFORMACIONTÉCNICA:EDAD	-2.0408079	4.128130e-02	
## MODALIDADDISTANCIA:C. EDADMED	0.2256143	8.215034e-01	
## MODALIDADPRESENCIAL:C. EDADMED	2.5704668	1.016238e-02	
## MODALIDADDISTANCIA:C. PORCMUJ	-1.5601251	1.187441e-01	
## MODALIDADPRESENCIAL:C. PORCMUJ	-2.1909718	2.846381e-02	

Se obtienen las estimaciones de los parámetros de la estructura de covarianzas, que en este caso son:

VarCorr(Model.8a)

## CARRERA = pdLogChol(EDAD + MODALIDAD + beca)	Variance	StdDev	Corr		
## (Intercept)	0.595153735	0.77146208	(Intr)	EDAD	MODALI
## EDAD	0.001191351	0.03451595	-0.837		
## MODALIDADPRESENCIAL	1.282360810	1.13241371	-0.056	-0.305	

```
## becaSI          0.631166241 0.79445972 -0.561  0.646 -0.415
## Residual       22.600059590 4.75395200
```

Significación de σ_{u7}^2 : Se compara la verosimilitud del modelo frente a la verosimilitud del modelo bajo la hipótesis de que $\sigma_{u7}^2 = 0$.

```
anova(Model.8a,Model.7)
```

```
##          Model df      AIC      BIC    logLik  Test  L.Ratio p-value
## Model.8a     1 29 137710.6 137944 -68826.31
## Model.7      2 25 137709.8 137911 -68829.92 1 vs 2 7.214678  0.125
```

Dado que no resulta significativa la varianza σ_{u7}^2 , se mantiene el modelo anterior.

```
Model.8<-Model.7
```

```
summary(Model.8)$tTable
```

```
##                               Value Std.Error  DF
## (Intercept)                   27.87091925 1.94109137 23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA      0.38213177 0.74043384  29
## AREAFORMACIONSOCIO-HUMANÍSTICA          1.48150027 0.40278573  29
## AREAFORMACIONTÉCNICA                   -0.99103884 0.77153781  29
## SEXOMASCULINO                       -0.06131503 0.07771399 23038
## MODALIDADPRESENCIAL                   -5.51431389 3.93663638 23038
## EDAD                                  0.09670055 0.01243601 23038
## HABITATRURAL                          -1.70517236 0.18515377 23038
## HABITATURBANA                         -1.49640067 0.17567116 23038
## becaSI                                 4.29952740 0.18203687 23038
## SEXOMASCULINO:MODALIDADPRESENCIAL     -1.45243908 0.17489325 23038
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -0.03600413 0.02889062 23038
## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD   -0.03276180 0.01547547 23038
## AREAFORMACIONTÉCNICA:EDAD             -0.04555247 0.02818609 23038
## MODALIDADDISTANCIA:C.EDADMED           0.02700482 0.04446418 23038
## MODALIDADPRESENCIAL:C.EDADMED         0.23508437 0.09021371 23038
## MODALIDADDISTANCIA:C.PORCMUJ          -0.01995370 0.01201299 23038
## MODALIDADPRESENCIAL:C.PORCMUJ        -0.03024318 0.01501424 23038
##                               t-value      p-value
## (Intercept)                   14.3583757 1.499991e-46
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA      0.5160917 6.097033e-01
## AREAFORMACIONSOCIO-HUMANÍSTICA          3.6781349 9.513522e-04
## AREAFORMACIONTÉCNICA                   -1.2844981 2.091342e-01
## SEXOMASCULINO                       -0.7889833 4.301299e-01
## MODALIDADPRESENCIAL                   -1.4007679 1.612970e-01
## EDAD                                  7.7758511 7.807078e-15
## HABITATRURAL                          -9.2094931 3.548520e-20
## HABITATURBANA                         -8.5181919 1.718400e-17
## becaSI                                 23.6189924 6.909066e-122
## SEXOMASCULINO:MODALIDADPRESENCIAL     -8.3047176 1.055031e-16
## AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD -1.2462221 2.126955e-01
```

## AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-2.1170143	3.426932e-02
## AREAFORMACIONTÉCNICA:EDAD	-1.6161330	1.000006e-01
## MODALIDADDISTANCIA:C.EDADMED	0.6073387	5.436321e-01
## MODALIDADPRESENCIAL:C.EDADMED	2.6058609	9.170263e-03
## MODALIDADDISTANCIA:C.PORCMUJ	-1.6610097	9.672510e-02
## MODALIDADPRESENCIAL:C.PORCMUJ	-2.0143001	4.398964e-02

Es decir, el modelo multinivel ajustado para el indicador de rendimiento académico “nota media ponderada” es:

$$y_{ij} = [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + [\gamma_{30}Pres_{ij} + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + \gamma_{33}(C.Pmuj_j \times Pres_{ij}) + \gamma_{34}(C.Pmuj_j \times Dis_{ij}) + u_{3j}] + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \quad \forall j \text{ e independientes entre sí}$$

$$u_{3j} \sim N(0, \sigma_{u3}^2) \quad \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0, 1, 3$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0, 1, 3$$

Y se obtienen las siguientes estimaciones:

	Coef.	Estimación	EES	Df	t-student	p-valor
(Intercept)	γ_{00}	27.870919	1.941091	23038	14.358376	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	γ_{01}	0.382132	0.740434	29	0.516092	0.6097
AREAFORMACIONSOCIO-HUMANÍSTICA	γ_{02}	1.481500	0.402786	29	3.678135	0.0010***
AREAFORMACIONTÉCNICA	γ_{03}	-0.991039	0.771538	29	-1.284498	0.2091
SEXOMASCULINO	γ_2	-0.061315	0.077714	23038	-0.788983	0.4301
MODALIDADPRESENCIAL	γ_{30}	-5.514314	3.936636	23038	-1.400768	0.1613
EDAD	γ_{10}	0.096701	0.012436	23038	7.775851	0.0000***
HABITATRURAL	γ_5	-1.705172	0.185154	23038	-9.209493	0.0000***
HABITATURBANA	γ_6	-1.496401	0.175671	23038	-8.518192	0.0000***
becaSI	γ_7	4.299527	0.182037	23038	23.618992	0.0000***
SEXOMASCULINO:MODALIDADPRESENCIAL	γ_4	-1.452439	0.174893	23038	-8.304718	0.0000***
AREAFORMACION_BIOL Y BIOMÉD:EDAD	γ_{11}	-0.036004	0.028891	23038	-1.246222	0.2127
AREAFORMACION_SOCIO-HUMAN:EDAD	γ_{12}	-0.032762	0.015475	23038	-2.117014	0.0343**
AREAFORMACIONTÉCNICA:EDAD	γ_{13}	-0.045552	0.028186	23038	-1.616133	0.1000*
MODALIDADDISTANCIA:C.EDADMED	γ_{32}	0.027005	0.044464	23038	0.607339	0.5436
MODALIDADPRESENCIAL:C.EDADMED	γ_{31}	0.235084	0.090214	23038	2.605861	0.0092***
MODALIDADDISTANCIA:C.PORCMUJ	γ_{34}	-0.019954	0.012013	23038	-1.661010	0.0967*
MODALIDADPRESENCIAL:C.PORCMUJ	γ_{33}	-0.030243	0.015014	23038	-2.014300	0.0440**

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

	Estimación Varianza	Estimación Desv. Estándar	Matriz de correlaciones	
			(Intercept)	EDAD
(Intercept): u_{0j}	σ_{u0}^2	0.20199	0.44943224	
EDAD: u_{1j}	σ_{u1}^2	0.00053	0.02303415	-0.572
MODALIDAD: u_{3j}	σ_{u3}^2	1.41810	1.19083625	-0.213
Residual: e_{ij}	σ_e^2	22.6240	4.75649717	-0.442

3.5. RESULTADOS

En este apartado se incluyen los resultados de los modelos multinivel obtenidos siguiendo un procedimiento similar al descrito en la sección 3.4. para las dos variables objetivo, indicadores del rendimiento académico (nota media ponderada y porcentaje de créditos superados) en los cuatro periodos académicos considerados.

3.5.1. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO OCTUBRE 2010 - FEBRERO 2011

El modelo multinivel de dos niveles con coeficientes aleatorios, utilizándose el índice i para el estudiante, índice j para el grupo determinado por la carrera o titulación que está cursando el estudiante, obtenido del análisis de los datos en el primer periodo académico es:

$$\begin{aligned} \text{Mediapond}_{ij} = & [\gamma_{00} + \gamma_{01}\mathbf{AFbb}_j + \gamma_{02}\mathbf{AFsh}_j + \gamma_{03}\mathbf{AFtc}_j + u_{0j}] + [\gamma_{10} + \gamma_{11}\mathbf{AFbb}_j + \\ & \gamma_{12}\mathbf{AFsh}_j + \gamma_{13}\mathbf{AFtc}_j + u_{1j}]\mathbf{Edad}_{ij} + \gamma_2\mathbf{Sexo}_{ij} + \\ & [\gamma_{30}\mathbf{Pres}_{ij} + \gamma_{31}(\mathbf{C.Edm}_j \times \mathbf{Pres}_{ij}) + \gamma_{32}(\mathbf{C.Edm}_j \times \mathbf{Dis}_{ij}) + \gamma_{33}(\mathbf{C.Pmuj}_j \times \mathbf{Pres}_{ij}) + \\ & \gamma_{34}(\mathbf{C.Pmuj}_j \times \mathbf{Dis}_{ij}) + u_{3j}] + \gamma_4\mathbf{Sex} * \mathbf{Mod}_{ij} + \gamma_5\mathbf{Rur}_{ij} + \gamma_6\mathbf{Urb}_{ij} + \gamma_7\mathbf{Beca}_{ij} + e_{ij} \end{aligned}$$

donde:

Sexo = 0 (Femenino), 1 (Masculino)

Modalidad = 0 (Distancia), 1 (Presencial)

- Dis = 1 (Si Modalidad a Distancia), 0 (caso contrario)
- Pres = 1 (Si Modalidad es Presencial), 0 (caso contrario)

Hábitat (Extranjero, Rural, Urbano):

- Rural = 1 (Si Hábitat=Rural), 0 (en caso contrario)
- Urbano = 1 (Si Hábitat=Urbano), 0 (en caso contrario)

Beca = 0 (No recibe beca), 1 (Sí recibe beca)

Área de Formación (Administrativa; Biológica-Biomédica; Socio-Humanística;

Técnica):

- AreaBB = 1 (Biológica y Biomédica), 0 (en caso contrario)

- AreaSH = 1 (Socio-Humanística), 0 (en caso contrario)
- AreaTC = 1 (Técnica), 0 (en caso contrario)

$$\begin{aligned}
 u_{0j} &\sim N(0, \sigma_{u_0}^2) \quad \forall j \text{ e independientes entre sí} \\
 u_{1j} &\sim N(0, \sigma_{u_1}^2) \quad \forall j \text{ e independientes entre sí} \\
 u_{3j} &\sim N(0, \sigma_{u_3}^2) \quad \forall j \text{ e independientes entre sí} \\
 e_{ij} &\sim N(0, \sigma_e^2) \quad \forall (i, j) \text{ e independientes entre sí} \\
 Cov[u_{kj}, e_{ij}] &= 0 \quad \forall (i, j), \forall k = 0, 1, 3 \\
 Cov(u_{kj}, u_{dj}) &= \sigma_{kd} \quad \forall k, d = 0, 1, 3
 \end{aligned}$$

En la Tabla 24 se presentan las estimaciones y significación de los coeficientes γ para el modelo ajustado para nota media ponderada en el periodo octubre 2010 – febrero 2011.

**Tabla 24. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA NOTA MEDIA PONDERADA
PERIODO OCTUBRE 2010 - FEBRERO 2011**

	Estimación	EES	Df	t-student	p-valor
(Intercept)	27.870919	1.941091	23038	14.358376	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.382132	0.740434	29	0.516092	0.6097
AREAFORMACIONSOCIO-HUMANÍSTICA	1.481500	0.402786	29	3.678135	0.0010***
AREAFORMACIONTÉCNICA	-0.991039	0.771538	29	-1.284498	0.2091
SEXOMASCULINO	-0.061315	0.077714	23038	-0.788983	0.4301
MODALIDADPRESENCIAL	-5.514314	3.936636	23038	-1.400768	0.1613
EDAD	0.096701	0.012436	23038	7.775851	0.0000***
HABITATRURAL	-1.705172	0.185154	23038	-9.209493	0.0000***
HABITATURBANA	-1.496401	0.175671	23038	-8.518192	0.0000***
becaSI	4.299527	0.182037	23038	23.618992	0.0000***
SEXOMASCULINO:MODALIDADPRESENCIAL	-1.452439	0.174893	23038	-8.304718	0.0000***
AREAFORMACION_BIOL Y BIOMÉD:EDAD	-0.036004	0.028891	23038	-1.246222	0.2127
AREAFORMACION_SOCIO-HUMAN:EDAD	-0.032762	0.015475	23038	-2.117014	0.0343**
AREAFORMACIONTÉCNICA:EDAD	-0.045552	0.028186	23038	-1.616133	0.1000*
MODALIDADDISTANCIA:C. EDADMED	0.027005	0.044464	23038	0.607339	0.5436
MODALIDADPRESENCIAL:C. EDADMED	0.235084	0.090214	23038	2.605861	0.0092***
MODALIDADDISTANCIA:C. PORCMUJ	-0.019954	0.012013	23038	-1.661010	0.0967*
MODALIDADPRESENCIAL:C. PORCMUJ	-0.030243	0.015014	23038	-2.014300	0.0440**

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

		Estimación Varianza	Estimación Desv. Estándar	Matriz de correlaciones	
				(Intercept)	EDAD
(Intercept): u_{0j}	σ_{u0}^2	0.20199	0.44943224		
EDAD: u_{1j}	σ_{u1}^2	0.00053	0.02303415	-0.572	
MODALIDAD: u_{3j}	σ_{u3}^2	1.41810	1.19083625	-0.213	-0.442
Residual: e_{ij}	σ_e^2	22.6240	4.75649717		

3.5.2. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO ABRIL 2011 - AGOSTO 2011

En la Tabla 25 se muestran las estimaciones y significación de los coeficientes γ para el modelo ajustado para nota media ponderada en el periodo abril 2011 – agosto 2011. La expresión paramétrica del modelo es semejante al periodo anterior, con los parámetros incluidos en la tabla siguiente.

**Tabla 25. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES y:
MODELO AJUSTADO PARA NOTA MEDIA PONDERADA
PERIODO ABRIL 2011 - AGOSTO 2011**

	Estimación	EES	DF	t-student	p-valor
(Intercept)	26.437615	0.4715701	23296	56.06295	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-1.161184	0.9312417	29	-1.24692	0.2224
AREAFORMACIONSOCIO-HUMANÍSTICA	1.967399	0.5649216	29	3.48261	0.0016***
AREAFORMACIONTÉCNICA	-0.940360	1.0962116	29	-0.85783	0.3980
SEXOMASCULINO	-0.089698	0.0751662	23296	-1.19332	0.2328
MODALIDADPRESENCIAL	2.267112	0.3174026	23296	7.14270	0.0000***
EDAD	0.097728	0.0120617	23296	8.10233	0.0000***
HABITATRURAL	-1.087037	0.1798867	23296	-6.04290	0.0000***
HABITATURBANA	-0.792245	0.1711410	23296	-4.62919	0.0000***
becaSI	3.311491	0.1858570	23296	17.81741	0.0000***
SEXOMASCULINO:MODALIDADPRESENCIAL	-1.828704	0.1796449	23296	-10.17955	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	0.005508	0.0269209	23296	0.20458	0.8379
AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.034049	0.0152280	23296	-2.23596	0.0254**
AREAFORMACIONTÉCNICA:EDAD	-0.000615	0.0280796	23296	-0.02190	0.9825
AREAFORMACIONBIOL. BIOM. :MOD. PRESENCIAL	1.295371	0.5776835	23296	2.24235	0.0249**
AREAFORMACIONSOCIO-HUM. :MOD. PRESENCIAL	-0.000176	0.4711795	23296	-0.00037	0.9997
AREAFORMACIONTÉCNICA:MOD. PRESENCIAL	0.449813	0.7659233	29	0.58728	0.5616

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

	Variance	StdDev	Corr
(Intercept)	0.9056682	0.95166603	(Intr) EDAD
EDAD	0.0005234	0.02287805	-0.848
MODALIDADPRESENCIAL	0.2637601	0.51357577	0.435 -0.558
Residual	22.2056800	4.71229020	

3.5.3. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO OCTUBRE 2011 – FEBRERO 2012

En la Tabla 26 se indican las estimaciones y significación de los coeficientes γ para el modelo ajustado para nota media ponderada en el periodo octubre 2011 – febrero 2012.

**Tabla 26. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA NOTA MEDIA PONDERADA
PERIODO OCTUBRE 2011 - FEBRERO 2012**

	Estimación	EES	DF	t-student	p-valor
(Intercept)	26.193996	1.5598434	24426	16.792709	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	0.078944	0.6326181	30	0.124790	0.9015
AREAFORMACIONSOCIO-HUMANÍSTICA	2.094202	0.4162811	30	5.030739	0.0000***
AREAFORMACIONTÉCNICA	-0.119769	0.6631231	30	-0.180614	0.8579
SEXOMASCULINO	-0.108416	0.1036349	24426	-1.046137	0.2955
MODALIDADPRESENCIAL	-4.658481	2.9496252	24426	-1.579347	0.1143
EDAD	0.102581	0.0085235	24426	12.035115	0.0000***
HABITATRURAL	-1.202540	0.1672658	24426	-7.189395	0.0000***
HABITATURBANA	-1.036272	0.1589996	24426	-6.517447	0.0000***
becaSI	2.499593	0.9708074	24426	2.574756	0.0100***
SEXOMASCULINO:MODALIDADPRESENCIAL	-1.624404	0.2522773	24426	-6.438962	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA:EDAD	-0.012795	0.0186351	24426	-0.686581	0.4924
AREAFORMACIONSOCIO-HUMANÍSTICA:EDAD	-0.033215	0.0104824	24426	-3.168612	0.0015***
AREAFORMACIONTÉCNICA:EDAD	-0.011966	0.0197387	24426	-0.606224	0.5444
MODALIDADDISTANCIA:C.EDADMED	0.003459	0.0353335	24426	0.097882	0.9220
MODALIDADPRESENCIAL:C.EDADMED	0.172007	0.0683680	24426	2.515895	0.0119***
becaNO:C.PORCMUJ	0.002953	0.0074625	24426	0.395718	0.6923
becaSI:C.PORCMUJ	0.038937	0.0186022	24426	2.093153	0.0363***

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

	Variance	StdDev	Corr
(Intercept)	0.595396	0.77161934	(Intr) EDAD MODALIDADPRESENCIAL SEXOMA becaSI
EDAD	0.000171	0.01310864	-0.840
MODALIDADPRESENCIAL	0.930462	0.96460477	-0.438 0.426
SEXOMASCULINO	0.086198	0.29359580	-0.430 0.171 0.566
becaSI	1.004985	1.00248940	0.605 -0.495 -0.502 -0.397
MODALIDADPRESENCIAL:SEXOMASCULINO	0.726762	0.85250346	-0.376 0.097 -0.234 -0.038 -0.433
Residual	19.54209	4.42064342	

3.5.4. MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO ABRIL 2012 – AGOSTO 2012

En la Tabla 27 se muestran las estimaciones y significación de los coeficientes γ para el modelo ajustado para nota media ponderada en el periodo abril 2012 – agosto 2012.

**Tabla 27. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA NOTA MEDIA PONDERADA
PERIODO ABRIL 2012 - AGOSTO 2012**

	Estimación	EES	Df	t-student	p-valor
(Intercept)	25.499087	0.7196080	23989	35.43469	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.124312	0.7536680	30	0.16494	0.8701
AREAFORMACIONSOCIO-HUMANÍSTICA	2.162339	0.5053993	30	4.27848	0.0002***
AREAFORMACIONTÉCNICA	-0.959855	0.7748396	30	-1.23878	0.2250
SEXOMASCULINO	-0.131027	0.0694518	23989	-1.88659	0.0592*
MODALIDADPRESENCIAL	2.226971	0.3224606	23989	6.90618	0.0000***
EDAD	0.109360	0.0096517	23989	11.33064	0.0000***
HABITATRURAL	-1.026447	0.1719314	23989	-5.97010	0.0000***
HABITATURBANA	-0.867034	0.1641427	23989	-5.28219	0.0000***
becaSI	6.820847	1.1661375	23989	5.84909	0.0000***
SEXOMASCULINO:MODALIDADPRESENCIAL	-1.071638	0.1623071	23989	-6.60253	0.0000***
AREAFORMACION_BIOL Y BIOMÉD:EDAD	-0.013663	0.0207663	23989	-0.65796	0.5106
AREAFORMACION_SOCIO-HUMAN:EDAD	-0.030888	0.0119443	23989	-2.58600	0.0097***
AREAFORMACIONTÉCNICA:EDAD	-0.022047	0.0222837	23989	0.98939	0.3225
becaNO:C.RETRMED	0.000079	0.0000630	23989	1.24830	0.2119
becaNO:C.RETRMED	-0.000190	0.0000846	23989	-2.24070	0.0251**

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

	Variance	StdDev	Corr
(Intercept)	0.982835	0.99138018	(Intr) EDAD MODALI
EDAD	0.000324	0.01799855	-0.795
MODALIDADPRESENCIAL	1.277993	1.13048354	-0.498 0.556
becaSI	1.154452	1.07445432	-0.384 -0.101 -0.317
Residual	19.16932	4.37827872	

3.5.5. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011

Considerando ahora como variable objetivo el indicador porcentaje de créditos superados, se ha aplicado un procedimiento análogo al realizado sobre la variable nota media ponderada. A continuación se muestra el modelo multinivel de dos niveles con coeficientes aleatorios, utilizándose el índice i para el estudiante, índice j para el grupo determinado por la carrera o titulación que está cursando el estudiante.

$$\begin{aligned}
 \text{PorcCredSuperados}_{ij} &= (\gamma_{00} + \gamma_{01} \text{CEdadmed} + u_{0j}) \\
 &+ (\gamma_{10} + \gamma_{11} \text{AFbb}_j + \gamma_{12} \text{AFsh}_j + \gamma_{13} \text{AFtc}_j + \gamma_{14} \text{CEdadmed} \\
 &+ u_{1j}) \text{Edad}_{ij} \\
 &+ \gamma_2 \text{Sexo}_{ij} \\
 &+ \gamma_3 * \text{Modalidad}_{ij} + \gamma_4 (\text{Sexo}_{ij} * \text{Mod}_{ij}) \\
 &+ \gamma_5 \text{Rur}_{ij} + \gamma_6 \text{Urb}_{ij} \\
 &+ \gamma_7 \text{Beca}_{ij} + e_{ij}
 \end{aligned}$$

donde:

Sexo = 0 (Femenino), 1 (Masculino)

Modalidad = 0 (Distancia), 1 (Presencial)

Hábitat (Extranjero, Rural, Urbano):

- Rural = 1 (Si Hábitat=Rural), 0 (en caso contrario)
- Urbano = 1 (Si Hábitat=Urbano), 0 (en caso contrario)

Beca = 0 (No recibe beca), 1 (Sí recibe beca)

Área de Formación (Administrativa; Biológica-Biomédica; Socio-Humanística; Técnica):

- AreaBB = 1 (Biológica y Biomédica), 0 (en caso contrario)
- AreaSH = 1 (Socio-Humanística), 0 (en caso contrario)
- AreaT = 1 (Técnica), 0 (en caso contrario)

$$u_{0j} \sim N(0, \sigma_{u0}^2) \forall j \text{ e independientes entre sí}$$

$$u_{1j} \sim N(0, \sigma_{u1}^2) \forall j \text{ e independientes entre sí}$$

$$e_{ij} \sim N(0, \sigma_e^2) \forall (i, j) \text{ e independientes entre sí}$$

$$Cov[u_{kj}, e_{ij}] = 0 \quad \forall (i, j), \forall k = 0, 1$$

$$Cov(u_{kj}, u_{dj}) = \sigma_{kd} \quad \forall k, d = 0, 1$$

En la Tabla 28 se muestran las estimaciones y significación de los coeficientes γ para el modelo ajustado para porcentaje de créditos superados en el periodo octubre 2010 – febrero 2011.

**Tabla 28. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS
PERIODO OCTUBRE 2010 - FEBRERO 2011**

	Estimación	EES	DF	t-Student	p-valor
(Intercept)	1.677210	14.552911	23041	0.115249	0.9082
SEXOMASCULINO	-0.514573	0.431274	23041	-1.193145	0.2328
MODALIDADPRESENCIAL	15.484929	0.852023	23041	18.174311	0.0000***
EDAD	2.005621	0.418660	23041	4.790576	0.0000***
HABITATRURAL	-8.365796	1.038092	23041	-8.058822	0.0000***
HABITATURBANA	-6.734037	0.984856	23041	-6.837587	0.0000***
becaSI	15.888036	1.018475	23041	15.599831	0.0000***
C. EDADMED	1.611715	0.373979	31	4.309639	0.0002***
SEXOMASCULINO:MODALIDADPRESENCIAL	-5.247020	0.957966	23041	-5.477250	0.0000***
EDAD:AREAFORMACIONBIOLÓG Y BIOM	-0.157925	0.050448	23041	-3.130474	0.0017***
EDAD:AREAFORMACIONSOCIO-HUMAN	-0.015953	0.031393	23041	-0.508156	0.6113
EDAD:AREAFORMACIONTÉCNICA	-0.119569	0.048800	23041	-2.450198	0.0143**
EDAD:C. EDADMED	-0.037646	0.010846	23041	-3.470944	0.0005***

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

Variance	Stddev	Corr
(Intercept)	16.29956	4.03727158 (Intr)
EDAD	0.006996	0.08364121 -0.89
Residual	712.3919	26.69067148

3.5.6. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO ABRIL 2011 – AGOSTO 2011

En la Tabla 29 se muestran las estimaciones y significación de los coeficientes γ para el modelo ajustado para porcentaje de créditos superados en el periodo abril 2011 – agosto 2011.

**Tabla 29. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS
PERIODO ABRIL 2011 - AGOSTO 2011**

	Estimación	EES	DF	t-Student	p-valor
(Intercept)	53.80382	2.422560	23298	22.209489	0.0000***
SEXOMASCULINO	-0.29312	0.425798	23298	-0.688396	0.4912
MODALIDADPRESENCIAL	19.28284	0.906963	23298	21.260879	0.0000***
EDAD	0.67055	0.056820	23298	11.801342	0.0000***
HABITATRURAL	-6.17408	1.027445	23298	-6.009159	0.0000***
HABITATURBANA	-4.66607	0.977642	23298	-4.772784	0.0000***
BecaSI	14.63899	1.060370	23298	13.805551	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	5.06395	3.943630	30	1.284082	0.2089
AREAFORMACIONSOCIO-HUMANÍSTICA	9.77942	2.833821	30	3.450967	0.0017***
AREAFORMACIONTÉCNICA	0.74382	3.930402	30	0.189248	0.8512
SEXOMASCULINO:MODALIDADPRESENCIAL	-5.86766	1.008605	23298	-5.817597	0.0000***
EDAD:AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.17467	0.120674	23298	-1.447496	0.1478
EDAD:AREAFORMACIONSOCIO-HUMANÍSTICA	-0.19154	0.070645	23298	-2.711341	0.0067***
EDAD:AREAFORMACIONTÉCNICA	-0.04576	0.123455	23298	-0.370677	0.7109

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

Variance	StdDev	Corr	
(Intercept)	20.19088	4.6806850	(Intr)
EDAD	0.008896	0.0943191	-0.944
Residual	725.7731	26.9401759	

3.5.7. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO OCTUBRE 2011 – FEBRERO 2012

En la Tabla 30 se muestran las estimaciones y significación de los coeficientes γ para el modelo ajustado para porcentaje de créditos superados en el periodo octubre 2011 – febrero 2012.

**Tabla 30. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS
PERIODO OCTUBRE 2011 - FEBRERO 2012**

	Estimación	EES	DF	t-Student	p-valor
(Intercept)	60.19806	2.249119	24430	26.765170	0.0000***
SEXOMASCULINO	-0.81502	0.410174	24430	-1.987017	0.0469**
MODALIDADPRESENCIAL	14.28928	0.814883	24430	17.535389	0.0000***
EDAD	0.63337	0.054647	24430	11.590071	0.0000***
HABITATRURAL	-8.49927	1.334863	24430	-6.367147	0.0000***
HABITATURBANA	-7.14393	1.111258	24430	-6.428686	0.0000***
BecaSI	18.23506	1.076397	24430	16.940821	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	4.66518	3.444801	30	1.354268	0.1858
AREAFORMACIONSOCIO-HUMANÍSTICA	8.68852	2.536097	30	3.425942	0.0018***
AREAFORMACIONTÉCNICA	-3.90075	3.503836	30	-1.113278	0.2744
SEXOMASCULINO:MODALIDADPRESENCIAL	-4.81398	0.910829	24430	-5.285280	0.0000***
EDAD:AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.19861	0.110837	24430	-1.791882	0.0732*
EDAD:AREAFORMACIONSOCIO-HUMANÍSTICA	-0.15875	0.066489	24430	-2.387666	0.0170**
EDAD:AREAFORMACIONTÉCNICA	0.12359	0.115626	24430	1.068903	0.2851

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

	Variance	StdDev	Corr
(Intercept)	19.57518570	4.4243853	(Intr)
EDAD	0.01026476	0.1013151	-0.757
HABITATRURAL	11.14734643	3.3387642	0.095 -0.642
HABITATURBANA	3.87078682	1.9674315	-0.287 -0.336 0.758
Residual	687.50314966	26.2202813	

3.5.8. MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO ABRIL 2012 – AGOSTO 2012

En la Tabla 31 se muestran las estimaciones y significación de los coeficientes γ para el modelo ajustado para porcentaje de créditos superados en el periodo abril 2012 – agosto 2012.

**Tabla 31. ESTIMACIONES Y SIGNIFICACIÓN DE LOS COEFICIENTES γ :
MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS
PERIODO ABRIL 2012 - AGOSTO 2012**

	Estimación	EES	DF	t-Student	p-valor
(Intercept)	55.34552	2.356730	23991	23.484036	0.0000***
SEXOMASCULINO	-0.89562	0.417796	23991	-2.143674	0.0321**
MODALIDADPRESENCIAL	17.91842	1.711393	23991	10.470079	0.0000***
EDAD	0.66105	0.057831	23991	11.430638	0.0000***
HABITATRURAL	-5.85603	1.201253	23991	-4.874938	0.0000***
HABITATURBANA	-4.19549	1.364293	23991	-3.075216	0.0021***
BecaSI	13.78214	1.122047	23991	12.283036	0.0000***
AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	5.27722	3.402148	30	1.551143	0.1314
AREAFORMACIONSOCIO-HUMANÍSTICA	7.95914	2.356133	30	3.378053	0.0020***
AREAFORMACIONTÉCNICA	0.50516	3.694255	30	0.136743	0.8921
SEXOMASCULINO:MODALIDADPRESENCIAL	-3.24646	0.971071	23991	-3.343171	0.0008***
EDAD:AREAFORMACIONBIOLÓGICA Y BIOMÉDICA	-0.11536	0.122958	23991	-0.938214	0.3481
EDAD:AREAFORMACIONSOCIO-HUMANÍSTICA	-0.14777	0.068720	23991	-2.150301	0.0315**
EDAD:AREAFORMACIONTÉCNICA	0.01927	0.128766	23991	0.149675	0.8810

(Significación de los coeficientes: *** 1%, ** 5%, * 10%)

Y las estimaciones de las componentes de la varianza:

	Variance	StdDev	Corr
(Intercept)	34.93917593	5.9109370	(Intr) EDAD MODALI HABITATR
EDAD	0.01549012	0.1244593	-0.840
MODALIDADPRESENCIAL	49.27228144	7.0194217	-0.626 0.728
HABITATRURAL	4.88396680	2.2099699	0.078 -0.354 -0.711
HABITATURBANA	13.26363211	3.6419270	-0.154 -0.196 -0.609 0.859
Residual	694.89314630	26.3608260	

Capítulo 4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

En este capítulo se incluye el análisis e interpretación de los resultados obtenidos en el capítulo anterior. Así se realiza en primer lugar un análisis pormenorizado del modelo ajustado para el indicador nota media ponderada en el primer periodo académico estudiado. En segundo lugar se procede de forma análoga con el indicador porcentaje de créditos superados. Posteriormente se realiza un análisis comparativo de los modelos ajustados en los cuatro periodos académicos. Y finalmente se realiza un análisis comparativo de los modelos ajustados para los dos indicadores de rendimiento académico.

4.1. ANÁLISIS MODELO AJUSTADO PARA NOTA MEDIA PONDERADA EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011

En el modelo multinivel señalado en el punto 3.5.1, podemos observar que en el rendimiento académico universitario medido por la nota media ponderada influyen variables de primer nivel como sexo, edad, modalidad de estudios, hábitat y becas de los estudiantes; y, variables de segundo nivel como área de formación académica, edad media del profesorado y porcentaje de mujeres docentes.

Considerando el análisis de las variables que resultan significativas para un nivel de significación $\alpha \leq 0.1$, a continuación se destacan en negrillas los términos en los que interviene parámetros para los que se rechaza la hipótesis nula $\gamma_{ij}=0$

$$\begin{aligned} \text{Mediapond}_{ij} = & [\gamma_{00} + \gamma_{01}AFbb_j + \gamma_{02}AFsh_j + \gamma_{03}AFtc_j + u_{0j}] + [\gamma_{10} + \gamma_{11}AFbb_j + \\ & \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij} + \gamma_2Sexo_{ij} + \\ & [\gamma_{30}Pres_{ij} + \gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{32}(C.Edm_j \times Dis_{ij}) + \gamma_{33}(C.Pmuj_j \times Pres_{ij}) + \\ & \gamma_{34}(C.Pmuj_j \times Dis_{ij}) + u_{3j}] + \gamma_4Sex * Mod_{ij} + \gamma_5Rur_{ij} + \gamma_6Urb_{ij} + \gamma_7Beca_{ij} + e_{ij} \end{aligned}$$

Para ver la relación de cada uno de estos términos con el rendimiento académico siempre supondremos que los alumnos tienen iguales valores para todas las variables incluidas en el modelo excepto para las comentadas en cada epígrafe.

SEXO

De los relacionados con el sexo, el único parámetro que resulta significativo es el que interviene en el término $\gamma_4Sex * Mod_{ij}$ con una estimación $\hat{\gamma}_4 = -1.452439$.

Por tanto el modelo estima que en la modalidad presencial los hombres tendrán una nota media ponderada inferior en 1.452439 puntos a la de las mujeres que tengan iguales valores en las demás variables, es decir que tengan la misma edad, estudien la misma rama, etc.

EDAD

De los relacionados con la edad los parámetros que resultan significativos son los que aparecen en los términos

$$[\gamma_{10} + \gamma_{12}AFsh_j + \gamma_{13}AFtc_j + u_{1j}]Edad_{ij}$$

Siendo las estimaciones de estos tres parámetros:

$$\hat{\gamma}_{10} = 0.096701 \quad \hat{\gamma}_{12} = -0.032762 \quad \hat{\gamma}_{13} = -0.045552$$

Por tanto los valores que se obtienen como multiplicadores de la edad de los estudiantes de cada área de formación son los que se muestran en la Tabla 32.

Tabla 32. EDAD * ÁREA DE FORMACIÓN

ÁREA DE FORMACIÓN				
	ADMINISTRATIVA	BIOLÓGICA Y BIOMÉDICA	SOCIO-HUMANÍSTICA	TÉCNICA
EDAD	0.096701	0.096701	0.096701 + (-0.032762) = 0.063939	0.096701 + (-0.045552) = 0.051149

El modelo prevé que en todas las áreas de formación un incremento de edad repercute en un incremento de la nota media ponderada, siendo este incremento igual en el área administrativa y biológica-biomédica, y menor en las otras dos áreas. El menor aumento en relación a la edad se da en el área técnica con 0.051149 por año de edad. El Gráfico 1, ilustra lo anotado en este párrafo.

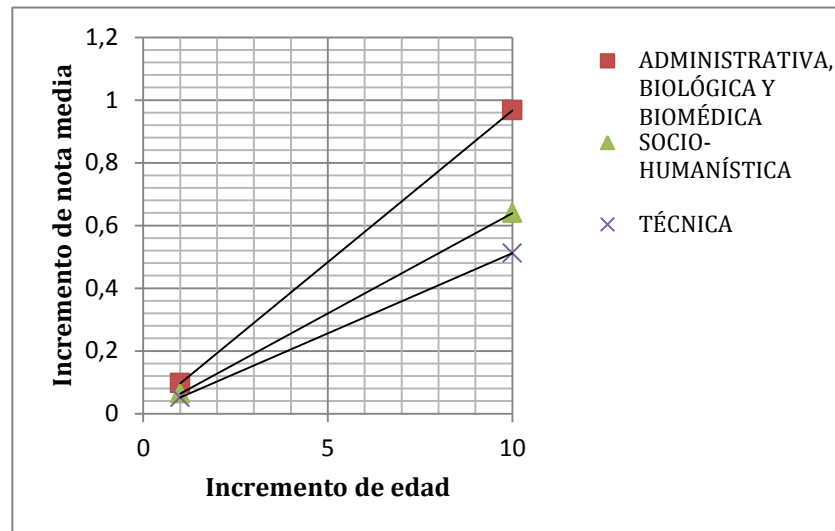


Gráfico 1. EDAD * ÁREA DE FORMACIÓN

MODALIDAD

Los parámetros relacionados con la modalidad son los que aparecen en la siguiente parte del modelo:

$$[\gamma_{31}(C.Edm_j \times Pres_{ij}) + \gamma_{33}(C.Pmuj_j \times Pres_{ij}) + \gamma_{34}(C.Pmuj_j \times Dis_{ij}) + u_{3j}] \\ + \gamma_4 Sex * Mod_{ij}$$

Y sus estimaciones son

$$\hat{\gamma}_{31} = 0.235084 \quad \hat{\gamma}_{33} = -0.030243 \quad \hat{\gamma}_{34} = -0.019954 \quad \hat{\gamma}_4 = -1.452439$$

Por tanto el incremento en un año en la edad media del profesorado supone un incremento de 0.235084 en la nota media de los estudiantes de modalidad presencial.

El incremento de un punto en el porcentaje de profesoras supone un decremento de 0.030243 en la nota media de los estudiantes de modalidad presencial, y un decremento de 0.019954 en la nota media de los estudiantes de modalidad a distancia.

En la modalidad presencial los estudiantes de sexo masculino tienen una nota media de 1.452439 puntos, menor que los de sexo femenino.

En la Tabla 33, se muestra, como ejemplo, la contribución de la modalidad para alumnos de sexo masculino y femenino si reciben clases de profesores con edad media de 39 años y de los cuales el 62% son mujeres.

Tabla 33. MODALIDAD*SEXO*EDAD MEDIA DOCENTES*PORCENTAJE DOCENTES MUJERES

MODALIDAD	SEXO	C.EDADMED (39 AÑOS)	C.PORMUJ (62%)	TOTAL
MODALIDAD PRESENCIAL	MASCULINO	0.235084 x 39	-0.030243 x 62	5.840771
		-1.452439	= 9.168276	= -1.875066
	FEMENINO	0.235084 x 39	-0.030243 x 62	7.29321
		-	= 9.168276	= -1.875066
MODALIDAD A DISTANCIA	MASCULINO		-0.019954 x 62	-1.237148
			= -1.237148	
	FEMENINO		-0.019954 x 62	-1.237148
			= -1.237148	

HÁBITAT

Los parámetros relacionados con el hábitat son

$$\gamma_5 Rur_{ij} + \gamma_6 Urb_{ij}$$

La puntuación media los estudiantes de hábitat rural con respecto a los de los centros del extranjero disminuye en -1.705172 y los de hábitat urbano en -1.496401.

BECA

El parámetro relacionado con la beca es

$$\gamma_7 Beca_{ij}$$

Los estudiantes que poseen beca presentan una puntuación media superior en 4.299527 en relación a los estudiantes que no poseen beca.

ÁREA DE FORMACIÓN

Además de la interacción con la edad, en el modelo interviene el sumando

$$\gamma_{02}AFsh_j$$

lo que supone que los alumnos del área socio – humanística tienen una nota media inferior en 1.4815 puntos a los del área administrativa que es la que se emplea como referencia en este modelo.

EDAD MEDIA DOCENTE

Como hemos visto anteriormente, para un incremento de 1 año de edad media de los docentes la puntuación media superior de los estudiantes de modalidad presencial varía en 0.235084 en relación a los alumnos de la modalidad a distancia.

PORCENTAJE DE MUJERES

Para un incremento del 1% de docentes mujeres la puntuación media del rendimiento disminuye en -0.030243 para los estudiantes de la modalidad presencial y en -0.019954 para los alumnos de la modalidad a distancia.

4.2. ANÁLISIS MODELO AJUSTADO PARA PORCENTAJE DE CRÉDITOS SUPERADOS EN EL PERIODO OCTUBRE 2010 – FEBRERO 2011

En el modelo multinivel señalado en el punto 3.5.5, podemos observar que en el rendimiento académico universitario medido por el porcentaje de créditos superados intervienen variables de primer nivel como sexo, edad, modalidad de estudios, hábitat y becas de los estudiantes; y, variables de segundo nivel como área de formación académica y edad media del profesorado.

Considerando el análisis de las variables que resultan significativas para un nivel de significación $\alpha \leq 0.1$, a continuación se destacan en negrillas los términos en los que intervienen parámetros para los que se rechaza la hipótesis nula $\gamma_{ij}=0$

$$\begin{aligned}
 \text{PorcCredSuperados}_{ij} &= (\gamma_{00} + \gamma_{01} \text{CEdadmed} + u_{0j}) \\
 &+ (\gamma_{10} + \gamma_{11} \text{AFbb}_j + \gamma_{12} \text{AFsh}_j + \gamma_{13} \text{AFtc}_j + \gamma_{14} \text{CEdadmed} + u_{1j}) \text{Edad}_{ij} \\
 &+ \gamma_2 \text{Sexo}_{ij} \\
 &+ \gamma_3 * \text{Modalidad}_{ij} + \gamma_4 (\text{Sexo}_{ij} * \text{Modalidad}_{ij}) \\
 &+ \gamma_5 \text{Rural}_{ij} + \gamma_6 \text{Urbano}_{ij} \\
 &+ \gamma_7 \text{Beca}_{ij} + e_{ij}
 \end{aligned}$$

Las estimaciones de los parámetros están recogidos en la Tabla 28. Para ver la relación de cada uno de estos términos con el rendimiento académico siempre supondremos que los alumnos tienen iguales valores para todas las variables incluidas en el modelo excepto para las comentadas en cada epígrafe.

SEXO

De los relacionados con el sexo, el único parámetro que resulta significativo es el que interviene en el término $\gamma_4 (\text{Sexo}_{ij} * \text{Mod}_{ij})$ con una estimación $\hat{\gamma}_4 = -5.247020$.

Por tanto el modelo estima que en la modalidad presencial los hombres tendrán un porcentaje de créditos superados menor en -5.247020 puntos porcentuales a la de las mujeres que tengan iguales valores en las demás variables, es decir que tengan la misma edad, estudien la misma rama, etc.

EDAD

De los relacionados con la edad, los parámetros que resultan significativos son los que aparecen en los términos

$$(\gamma_{10} + \gamma_{11} AFbb_j + \gamma_{13} AFtc_j + \gamma_{14} CEdadmed + u_{1j}) Edad_{ij}$$

Siendo las estimaciones de estos tres parámetros:

$$\hat{\gamma}_{10} = 2.005621 \quad \hat{\gamma}_{11} = -0.015953 \quad \hat{\gamma}_{13} = -0.119569$$

$$\hat{\gamma}_{14} = 0.037646$$

Por tanto los valores que se obtienen como multiplicadores de la edad de los estudiantes de cada área de formación son los que se muestran en la Tabla 34.

Tabla 34. EDAD*ÁREA DE FORMACIÓN*EDAD MEDIA DOCENTES

EDAD	ÁREA DE FORMACIÓN				C.EDADMED (Ej: 39 AÑOS)
	ADMINISTRATIVA	BIOLÓGICA Y BIOMÉDICA	SOCIO- HUMANÍSTICA	TÉCNICA	
2.005621	2.005621 + (-1,468194) = 0,537427	2.005621+ (-0.157925) + (-1.468194) = 1.847696	2.005621 + (-1.468194) = 0,537427	2.005621+ (-0.119569) + (-1,468194) = 0.379502	-0.037646 x 39 = -1.468194

El modelo prevé que en todas las áreas de formación un incremento de edad del estudiante repercute en un incremento del porcentaje de créditos superados, siendo este incremento igual en el área administrativa y socio - humanística, y mayor en el área biológica y biomédica. El menor aumento en relación a la edad se da en el área técnica con 0.379502 por año de edad. En el Gráfico 2, se ilustra lo recogido en este párrafo.

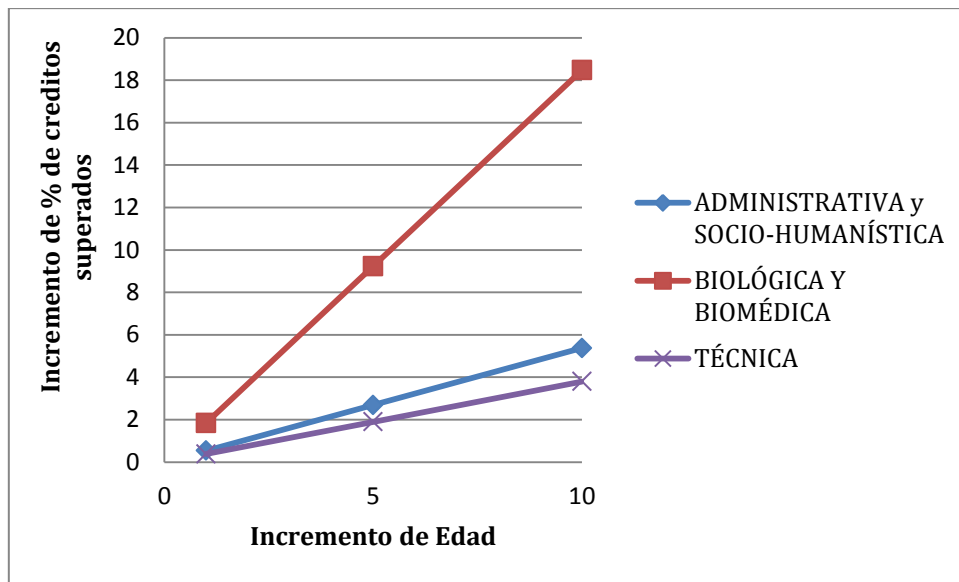


Gráfico 2. EDAD * ÁREA DE FORMACIÓN * EDAD MEDIA DOCENTE

MODALIDAD

Los parámetros relacionados con la modalidad son los que aparecen en la siguiente parte del modelo:

$$\gamma_3 * Modalidad_{ij} + \gamma_4 (Sexo_{ij} * Modalidad_{ij})$$

Y sus estimaciones son

$$\hat{\gamma}_3 = 15.484929 \quad \hat{\gamma}_4 = -5.247020$$

En la Tabla 35, se analiza la variable modalidad para alumnos de sexo masculino y femenino.

Tabla 35. MODALIDAD * SEXO

MODALIDAD	SEXO	
	MASCULINO	FEMENINO
MODALIDAD PRESENCIAL	15.484929 + (-5.247020) = 10.237909	15.484929

De acuerdo a la tabla anterior, los estudiantes de la modalidad presencial de sexo femenino superan 15.484929 el porcentaje de créditos superados con respecto a las estudiantes de la modalidad a distancia y esa diferencia sólo alcanza el valor de 10.237909 en caso de los estudiantes de sexo masculino.

HÁBITAT

Los parámetros relacionados con el hábitat son

$$\gamma_5 \text{Rural}_{ij} + \gamma_6 \text{Urbano}_{ij}$$

El hábitat rural con respecto al extranjero disminuye la media en porcentaje de créditos superados en -8.365796 y el hábitat urbano en -6.734037 puntos porcentuales.

BECA

El parámetro relacionado con la beca es

$$\gamma_7 \text{Beca}_{ij}$$

Los estudiantes que poseen beca presentan una media en porcentaje de créditos superados superior de 15.888036 en relación a los estudiantes que no poseen beca.

ÁREA DE FORMACIÓN

Los parámetros relacionados con el área de formación son

$$(\gamma_{11} AFbb_j + \gamma_{13} AFtc_j) Edad_{ij}$$

Y sus estimaciones son

$$\hat{\gamma}_{11} = -0.157925 \quad \hat{\gamma}_{13} = -0.119569$$

En la Tabla 36, se analiza el área de formación para estudiantes de 23 y 30 años como referencia.

Tabla 36. ÁREA DE FORMACIÓN * EDAD

AREA DE FORMACIÓN	EDAD (23 AÑOS)	EDAD (30 AÑOS)
ADMINISTRATIVA	-	-
SOCIO-HUMANÍSTICA	-	-
BIOLÓGICA Y BIOMÉDICA	-0.157925 x 23 = -3,632275	-0.157925 x 30 = -4,73775
TÉCNICA	-0.119569 x 23 = -2,750087	-0.119569 x 30 = -3,58707

De acuerdo a la Tabla 36, en relación al área administrativa la media inferior en porcentaje de créditos superados la muestra el área biológica y biomédica para estudiantes con 30 años de edad con -4,73775 puntos porcentuales. El área socio – humanística tiene igual media en porcentaje de créditos superados que el área administrativa.

EDAD MEDIA DOCENTE

El parámetro relacionado con la edad media docente es

$$\gamma_{01} CEdadmed \quad (\gamma_{14} CEdadmed) Edad_{ij}$$

Y sus estimaciones son

$$\hat{\gamma}_{01} = 1.611715$$

$$\hat{\gamma}_{14} = -0.037646$$

En la Tabla 37, se analiza la edad media de los docentes para estudiantes con edades de 23 y 30 años como referencia.

Tabla 37. EDAD MEDIA DOCENTE * EDAD

C.EDADMED	EDAD (23 AÑOS)	TOTAL (23 AÑOS)	EDAD (30 AÑOS)	TOTAL (30 AÑOS)
1.611715	$(-0.037646) \times 23$ = -0.865858	0,7458570	$(- 0.037646) \times 30$ = -1,12938	0,482335

Para un incremento de 1 año de edad media de los docentes la mayor media en porcentaje de créditos superados corresponde a los estudiantes de 23 años de edad con 0,7458570 puntos porcentuales.

En el Gráfico 3 en 3D y en el Gráfico 4 de contorno podemos ver la variación de los créditos superados en función de la edad de los estudiantes y de la edad media de los docentes.

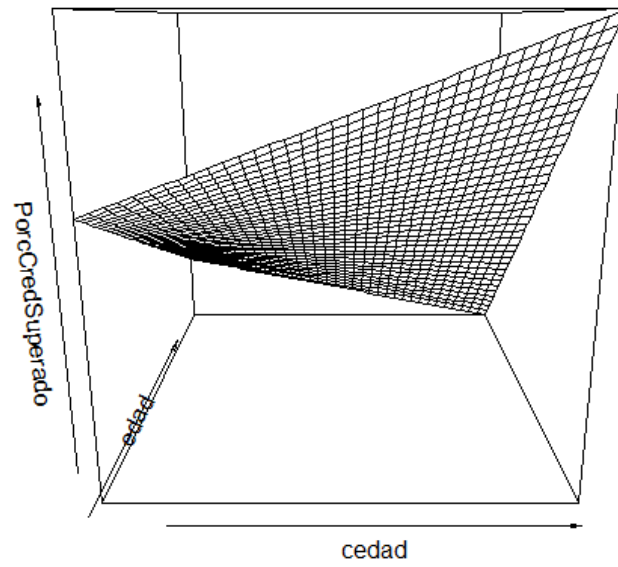


Gráfico 3. VARIACIÓN DE CRÉDITOS SUPERADOS - 3D

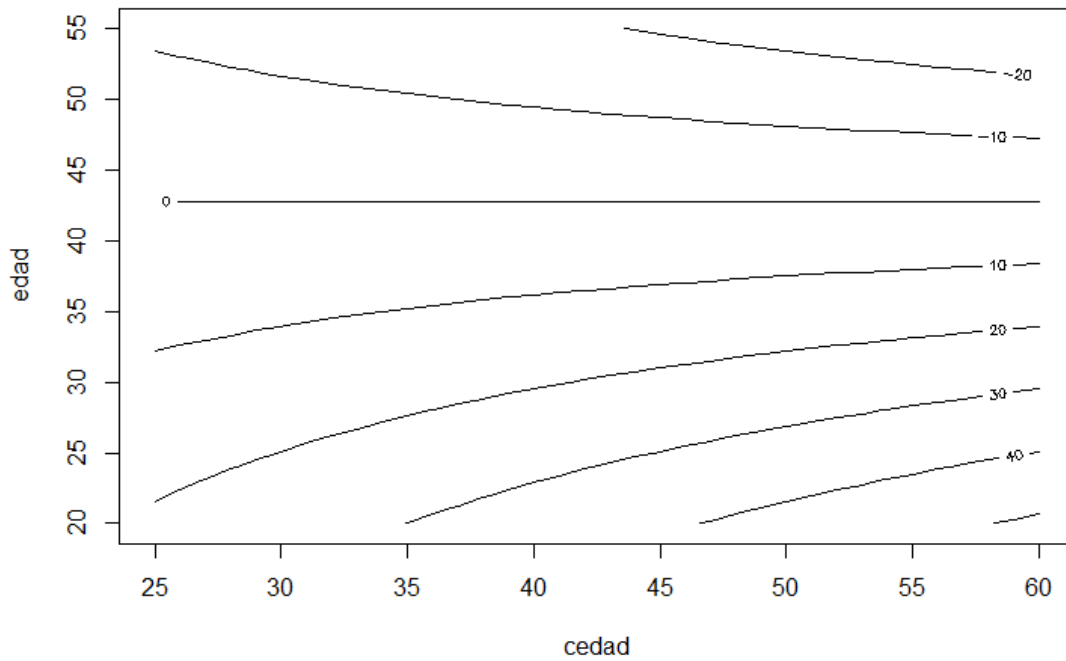


Gráfico 4. VARIACIÓN DE CRÉDITOS - CONTORNO

4.3. ANÁLISIS DEL MODELO AJUSTADO POR MEDIA PONDERADA POR PERIODOS

Los resultados del modelo multinivel del rendimiento académico universitario medido por la nota media ponderada para los cuatro periodos de estudio (Periodo 1 = Octubre 2010/Febrero 2011, Periodo 2 = Abril 2011/Febrero 2011, Periodo 3 = Octubre 2011/Febrero 2012, Periodo 4 = Abril 2012 – Agosto 21012) se resumen en la Tabla 38.

Con objeto de que resulte más fácil visualmente el análisis comparativo, la significación de cada coeficiente se recoge simbólicamente en la columna “p-valor”.

Debe interpretarse como sigue:

p-valor < 0.01 ***

p-valor < 0.05 **

p-valor < 0.1 *

Tabla 38. MODELO MEDIA PONDERADA POR PERIODOS

VARIABLES	γ	PERIODO 1		PERIODO 2		PERIODO 3		PERIODO 4	
		Estimación	p-valor	Estimación	p-valor	Estimación	p-valor	Estimación	p-valor
INTERCEPT	γ_{00}	27.870919	***	26.437615	***	26.193996	***	25.499087	***
ÁREA FORMACIÓN Administrativa	γ_{01}								
Biológica y Biomédica	γ_{02}	1.481500	***	1.967399	***	2.094202	***	2.162339	***
Socio – Humanística	γ_{03}								
Técnica	γ_2							-0.131027	*
SEXO Masculino	γ_{10}	0.096701	***	0.097728	***	0.102581	***	0.109360	***
EDAD * ÁREA FORMACIÓN Administrativa	γ_{11}								
Biológica y Biomédica	γ_{12}	-0.032762	**	-0.034049	**	-0.033215	***	-0.030888	***
Socio – Humanística	γ_{13}	-0.045552	*						
Técnica	γ_{30}			2.267112	***			2.226971	***
MODALIDAD Presencial	γ_{31}	0.235084	***			0.172007	**		
Presencial *PORCENTAJE MUJERES	γ_{33}	-0.030243	**						
Presencial *ÁREA FORMACIÓN Biológica y Biomédica				1.295371	**				
MODALIDAD a Distancia*EDAD MEDIA	γ_{32}								
a Distancia*PORCENTAJE MUJERES	γ_{33}	-0.019954	*						
SEXO Masculino*MODALIDAD Presencial	γ_4	-1.452439	***	-1.828704	***	-1.624404	***	-1.071638	***
HÁBITAT Extranjero	γ_5	-1.705172	***	-1.087037	***	-1.202540	***	-1.026447	***
Rural	γ_6	-1.496401	***	-0.792245	***	-1.036272	***	-0.867034	***
Urbano	γ_{70}	4.299527	***	3.311491	***	2.499593	***	6.820847	***
BECA SI	γ_{71}					0.038937	**		
SI**PORCENTAJE MUJERES	γ_{72}							-0.000190	**
SI***RETRIBUCIÓN MEDIA									

A partir de estas estimaciones podemos comprobar que el modelo es estable para los cuatro periodos con influencia significativa ($p < 0,01$) de las variables área de formación socio – humanística, edad, hábitat rural, hábitat urbano, beca, interacción de las variables sexo y modalidad presencial, interacción de las variables edad y área de formación socio – humanística ($p < 0,05$ periodos 1 y 2).

En el periodo 1 se observa una leve influencia de la interacción de las variables edad y área de formación técnica, modalidad presencial y porcentaje de mujeres, modalidad a distancia y porcentaje de mujeres, efecto que no se reproduce en el resto de periodos.

En los periodos 2 y 4, la modalidad presencial influye en el rendimiento académico universitario, esta influencia puede ser explicada por el desarrollo de los ciclos académicos pares (segundo, cuarto, sexto, octavo, décimo), y no se repite en el resto de periodos.

En los periodos 1 y 3 la interacción de la modalidad presencial y edad media incide en la nota media ponderada de los estudiantes universitarios y puede explicarse por la ejecución de los ciclos académicos impares (primero, tercero, quinto, séptimo, noveno), ésta influencia no se observa en los periodos restantes.

En el periodo 2 influye la interacción de la variable modalidad a distancia y porcentaje de mujeres docentes con $p < 0,05$.

4.4. ANÁLISIS DEL MODELO AJUSTADO POR PORCENTAJE DE CRÉDITOS SUPERADOS POR PERIODO

Los resultados del Modelo Multinivel del Rendimiento Académico Universitario medido por el porcentaje de créditos superados para los cuatro periodos de estudio (Periodo 1 = Octubre 2010/Febrero 2011, Periodo 2 = Abril 2011/Febrero 2011, Periodo 3 = Octubre 2011/Febrero 2012, Periodo 4 = Abril 2012 – Agosto 21012) se resumen en la Tabla 39.

Tabla 39. MODELO PORCENTAJE DE CRÉDITOS SUPERADOS

VARIABLES NIVEL 2	γ	PERIODO 1		PERIODO 2		PERIODO 3		PERIODO 4	
		Estimación	p-valor	Estimación	p-valor	Estimación	p-valor	Estimación	p-valor
INTERCEPT	γ_{00}			53.80382	***	60.19806	***	55.34552	***
EDAD MEDIA	γ_{01}	1.611716	***						
ÁREA FORMACIÓN Administrativa									
Biológica y Biomédica	γ_{01}								
Socio - Humanística	γ_{02}			9.77942	***	8.68852	***	7.95914	***
Técnica	γ_{03}								
SEXO Masculino	γ_2					-0.81502	***	-0.89562	**
EDAD	γ_{10}	2.005621	***	0.67055	***	0.63337	***	0.66105	***
EDAD * ÁREA FORMACIÓN Administrativa									
Biológica y Biomédica	γ_{11}	-0.157925	***			-0.19861	*		
Socio - Humanística	γ_{12}			0.0067	***	-0.15875	**	-0.14777	**
Técnica	γ_{13}	-0.119569	**						
EDAD * EDAD MEDIA	γ_{14}	-0.037646	***						
MODALIDAD A distancia									
Presencial	γ_3	15.484929	***	19.28284	***	14.28928	***	17.91842	***
SEXO Masculino*MODALIDAD Presencial	γ_4	-5.24702	***	-5.86766	***	-4.81398	***	-3.24646	***
HÁBITAT Extranjero									
Rural	γ_5	-8.365796	***	-6.17408	***	-8.49927	***	-5.85603	***
Urbano	γ_6	-6.734037	***	-4.66607	***	-7.14393	***	-4.19549	***
BECA SI	γ_7	15.888036	***	14.63899	***	18.23506	***	13.78214	***

A partir de estas estimaciones de los parámetros podemos verificar que el modelo es estable para los cuatro periodos con influencia significativa ($p < 0,01$) de las variables edad, modalidad presencial, interacción de las variables sexo y modalidad presencial, hábitat rural, hábitat urbano y beca.

En el periodo 1 con $p < 0,01$, las variables edad media del profesorado e interacción de las variables edad y edad media del profesorado influyen en el rendimiento académico universitario, efecto que no se repite en el resto de periodos.

En los periodos 2, 3 y 4 con $p < 0,01$ la variable área de formación socio – humanística influye en el porcentaje de créditos superados por los estudiantes, y éste resultado no se observa en el periodo 1.

En los periodos 1 ($p < 0,01$) y 3 ($p < 0,1$) la interacción de las variables edad y área de formación biológica y biomédica, en los periodos 2 ($p < 0,01$), 3 y 4 ($p < 0,05$) la interacción de las variables edad y área de formación socio – humanística; y, en el periodo 1 ($p < 0,05$) la interacción de la edad y el área técnica influyen en el rendimiento académico, los efectos no se repiten en los periodos restantes respectivamente.

4.5. ANÁLISIS COMPARATIVO DE LOS MODELOS AJUSTADOS POR NOTA MEDIA PONDERADA Y PORCENTAJE DE CRÉDITOS SUPERADOS

En las Tablas 38 y 39 se resumen los resultados de los modelos nota media ponderada y porcentaje de créditos aprobados, indicándose las variables que influyen en el rendimiento académico universitario. Para los dos modelos y en los cuatro periodos influyen significativamente ($p < 0,01$) las variables edad, hábitat rural, hábitat urbano, beca, sexo y modalidad presencial.

El área socio – humanística incide en los cuatro períodos en el rendimiento académico medido por la nota media ponderada, mientras que en el calculado por el porcentaje de créditos superados influye en los periodos 2, 3 y 4.

La modalidad presencial incide en los cuatro periodos en el rendimiento medido por el porcentaje de créditos superados, mientras que en el modelo nota media ponderada incide en los periodos 2 y 4.

La interacción de las variables edad y área socio – humanística influye en los cuatro periodos en el modelo nota media ponderada mientras que en el modelo porcentaje de créditos superados incide en los periodos 2, 3 y 4.

La influencia de la interacción de las variables modalidad presencial y edad media docente aparece en los periodos 1 y 3 en el rendimiento calculado por la nota media ponderada y no influye en el porcentaje de créditos superados de los estudiantes universitarios.

La interacción de las variables edad y área de formación técnica en el periodo 1, presentan una influencia leve en el rendimiento universitario medido por la nota media ponderada y en el mismo periodo, una influencia mayor en el rendimiento medido por el porcentaje de créditos superados.

La interacción de la modalidad presencial y porcentaje de mujeres, modalidad a distancia y porcentaje de mujeres en el periodo 1, influyen en el rendimiento medido a través del modelo

de nota media ponderada, ésta influencia no se observa en el modelo porcentaje de créditos superados.

La interacción de la variable modalidad a distancia y porcentaje de mujeres docentes influyen en el periodo 2 en el rendimiento calculado por la nota media ponderada y este efecto no se produce en el porcentaje de créditos superados.

Las variables edad media del profesorado e interacción de las variables edad y edad media del profesorado, en el periodo 1, inciden en la nota media ponderada de los estudiantes mientras que en el porcentaje de créditos ponderados no aparece esta influencia.

La interacción de las variables edad y área de formación biológica y biomédica en los periodos 1 y 3 influyen en el rendimiento universitario medido por el porcentaje de créditos superados y no inciden en la nota media ponderada.

Capítulo 5. DISCUSIÓN, LIMITACIONES, FUTURAS INVESTIGACIONES

5.1. DISCUSIÓN

Los objetivos de este estudio son establecer la relación multinivel entre factores inherentes al estudiante (edad, sexo, becas y hábitat de los estudiantes), factores inherentes a la organización académica universitaria (área de formación y modalidad de estudios) y factores inherentes al docente (experiencia docente, retribución media, edad media y sexo) con el rendimiento académico de los estudiantes medido por la nota media ponderada y porcentaje de créditos superados.

Estas variables se agrupan en una estructura multinivel de dos niveles. En el primer nivel se incluyen la edad de los estudiantes, sexo estudiantil, becas de los estudiantes, hábitat de los estudiantes y tipo de estudios de los estudiantes; y, en el segundo nivel el área de formación de los estudiantes, formación docente, experiencia del profesorado, retribución media de los profesores, edad media docente y sexo de los docentes.

5.1.1. DISCUSIÓN MODELO NOTA MEDIA PONDERADA

En relación a los objetivos planteados y considerando como indicador de rendimiento académico la nota media ponderada los resultados reportaron que a medida que los estudiantes universitarios aumentan su edad es mayor el éxito académico en comparación con los estudiantes más jóvenes, resultados que van en congruencia con los estudios de Richardson & Woodley (2003), Sheard (2009), Jansen & Bruinsma (2005); van Rooyen, Dixon, Dixon, & Wells (2006); Wojciechowski & Palmer (2005); Cantwell, Archer & Bourke (2001), Ofori (2000) y en contraposición con el autor Van den Berg & Hofman (2005) que menciona que los estudiantes mayores obtienen menos resultados académicos que los estudiantes más

jóvenes. En este estudio, en concreto los estudiantes del área administrativa y biológica y biomédica, incrementan su nota media ponderada a medida que su edad aumenta en mayor puntuación que los estudiantes del área socio – humanística y técnica.

Las estimaciones obtenidas también sugieren que el rendimiento académico medido por la nota media ponderada de las estudiantes es superior a sus homólogos masculinos de la modalidad presencial, este hallazgo es consistente con investigaciones anteriores como las de Garkaz, Banimahd & Esmaeili (2011); Barrow, Reilly & Woodfield (2009); Castagnetti & Rosti (2009); Smith & Naylor (2001); Gibb, Fergusson & Horwood (2008); Dayioğlu & Türüt-Aşık (2007); Sheard (2009); Mills, Heyworth, Rosenwax, Carr & Rosenberg (2009); Farsides & Woodfield (2007); Woodfield, Jessop & McMillan (2006); Jansen & Bruinsma (2005); Richardson & Woodley (2003); Gracia & Jenkins (2003); Cantwell, Archer & Bourke (2001). Sin embargo, estos estudios no analizan la interrelación del sexo con la modalidad a la que pertenecen los estudiantes universitarios.

La variable beca es un indicador de las condiciones socioeconómicas que en nuestro modelo tiene un incidencia significativa en la nota media de los estudiantes, de forma que los estudiantes becarios tienen una puntuación media superior a los estudiantes que no tienen beca. Este resultado se contrapone con los de Sackett *et al.* (2009); Cantwell, Archer & Bourke (2001); Lumb & Vail (2004) que afirman que el estatus socioeconómico no tiene influencia sobre el rendimiento académico. El hecho de que para obtener beca se exija diferentes requisitos académicos, en función del programa o institución que las concede, puede explicar esta diferencia en los resultados.

El hábitat de los estudiantes es otra de las variables que discrimina el éxito académico pues en países en vías de desarrollo, las condiciones geográficas, tecnológicas, sociales y culturales son diferentes. Este trabajo presenta que los estudiantes de los centros urbanos disminuyen su nota media en menos puntos que los estudiantes de los centros rurales frente a los estudiantes de los centros del extranjero. Estos resultados se alinean con los obtenidos en los trabajos de Halpern (2007), Dutton & Dutton (2005) que manifiestan que mientras es mayor el tiempo que tarda un estudiante para ir y volver de la universidad, peor es su rendimiento académico.

La nota media como indicador de rendimiento académico en este estudio es influida también por la variable modalidad de estudios en interrelación con el sexo, edad media de los docentes y porcentaje de mujeres docentes, a diferencia de lo obtenido en otras investigaciones como las de Buckalew, Daly & Coffield (1986), Woodfield, Jessop & McMillan (2006), Paisey & Paisey (2004), Stanca (2006), que relacionan características específicas de la modalidad presencial como la asistencia a clases con el rendimiento académico. Se encuentra que frente a un incremento en la edad media de docentes y porcentaje de docentes mujeres, los estudiantes de la modalidad presencial de sexo femenino seguidos de los estudiantes de sexo masculino presentan una nota media mayor que los estudiantes hombres y mujeres de la modalidad a distancia que aparecen con un decremento en la nota media. Esta disminución en la nota media de los estudiantes a distancia se puede explicar por la dedicación del tiempo a los estudios que destinan los estudiantes a distancia, quienes por sus compromisos familiares y/o laborales no cuentan con el tiempo necesario para estudiar, por lo que el rendimiento académico de los estudiantes a tiempo parcial se ve afectado negativamente según lo explican Halpern (2007), Hunt, Lincoln & Walker (2004), Callender (2008). Los resultados de esta investigación se contraponen con el estudio de Duvall & Schwartz (2000) que indican que no encontraron diferencias significativas en el rendimiento académico entre los estudiantes a distancia y presenciales.

Además de la interacción del área de formación con la edad, los resultados del modelo muestran que los estudiantes del área socio – humanística tienen una nota media inferior a los estudiantes del área administrativa. Aunque no es del mismo ciclo universitario, este resultado está en concordancia con el estudio de Buela-Casal *et al.* (2011) aplicado a una selección de doctorandos que muestra que el área con una menor tasa de éxito es la de humanidades.

Ante el incremento de un año de edad media de los docentes la puntuación media superior de los estudiantes de modalidad presencial varía en relación a los alumnos de la modalidad a distancia y para un incremento del 1% de docentes mujeres la puntuación media del rendimiento disminuye para los estudiantes de la modalidad presencial y a distancia.

Las variables experiencia y retribución media de los profesores no influyen en el rendimiento académico universitario con los datos utilizados en este estudio.

5.1.2. DISCUSIÓN MODELO PORCENTAJE DE CRÉDITOS APROBADOS

El segundo modelo planteado en esta investigación considera la relación multinivel de las variables inherentes al estudiante, organización académica universitaria y profesorado con el rendimiento académico medido por el porcentaje de créditos superados por los estudiantes universitarios; los resultados sugieren que ante un incremento de edad de los estudiantes existe un incremento del porcentaje de créditos superados en las cuatro áreas analizadas. Los hallazgos muestran correspondencia con las afirmaciones de Richardson & Woodley (2003), Sheard (2009), Jansen & Bruinsma (2005); van Rooyen, Dixon, Dixon, & Wells (2006); Wojciechowski & Palmer (2005); Cantwell, Archer & Bourke (2001), Ofori (2000) y contraposición con Van den Berg & Hofman (2005), a pesar que todos estos estudios no consideran la edad media de docentes como variable interrelacionada. El presente estudio específicamente muestra que el incremento del rendimiento académico es igual en el área administrativa y socio-humanística, mayor en el área biológica y biomédica y con un menor aumento en el área técnica.

Las estudiantes presentan un mayor porcentaje de créditos superados que los estudiantes de sexo masculino de la modalidad presencial, resultados que concuerdan con los reportados por Garkaz, Banimahd & Esmaili (2011); Barrow, Reilly & Woodfield (2009); Castagnetti & Rosti (2009); Smith & Naylor (2001); Gibb, Fergusson & Horwood (2008); Dayioğlu & Türüt-Aşik (2007); Sheard (2009); Mills, Heyworth, Rosenwax, Carr & Rosenberg (2009); Farsides & Woodfield (2007); Woodfield, Jessop & McMillan (2006); Jansen & Bruinsma (2005); Richardson & Woodley (2003); Gracia & Jenkins (2003); Cantwell, Archer & Bourke (2001). Sin embargo estos estudios no consideran el sexo en interrelación con la modalidad de estudios.

Esta investigación también encontró que el factor beca incide significativamente en el rendimiento académico, de tal manera que los estudiantes que poseen una beca reportan un

mayor porcentaje de créditos superados que los estudiantes que no la poseen, resultados que difieren con los de Sackett *et al.* (2009); Cantwell, Archer & Bourke (2001); Lumb & Vail (2004) que señalan que el estatus socioeconómico no tiene influencia sobre el rendimiento académico, ésta diferencia puede estar dada por los requerimientos académicos disímiles entre instituciones.

Otro de los hallazgos de este estudio muestra que el porcentaje de créditos superados de los estudiantes de los centros urbanos disminuyen en menos puntos porcentuales que los estudiantes de los centros rurales en relación a los estudiantes ubicados en los centros del extranjero, resultados que van en la misma línea con los expresados por Halpern (2007), Dutton & Dutton (2005) que indican que mientras más demora un alumno para ir y volver de la universidad, su rendimiento académico disminuye.

Así mismo se encuentra que las estudiantes de la modalidad presencial seguidas de sus homólogos masculinos presentan un mayor porcentaje de créditos superados que los alumnos de sexo femenino y masculino que desarrollan sus actividades académicas en la modalidad a distancia, la diferencia entre las dos modalidades puede dilucidarse por las características propias de los estudiantes a distancia que al desarrollar paralelamente funciones de tipo laboral, familiar y estudiantil, se convierten en estudiantes a tiempo parcial y que de acuerdo a Halpern (2007), Hunt, Lincoln & Walker (2004), Callender (2008) éste tipo de estudiantes presentan un menor rendimiento. Duvall & Schwartz (2000) por su parte señalan que no existen diferencias significativas en el rendimiento académico entre los estudiantes a distancia y presenciales.

El área de formación en interacción con la edad muestra que en relación al área administrativa el menor porcentaje de créditos superados está en el área biológica y biomédica, resultados que al ser comparados con el estudio a nivel de doctorado de Buela-Casal *et al.* (2011) difieren al encontrar que el área de humanidades tiene un menor rendimiento, en el presente estudio el área socio – humanística tiene igual porcentaje de créditos superados que el área administrativa.

Esta investigación también muestra que en la interacción edad media docente y edad, ante el incremento en la edad media del profesorado universitario y el incremento de la edad de los

estudiantes el porcentaje de créditos superados disminuye. Así mismo se encuentra que con los datos utilizados, las variables experiencia y retribución media de los profesores no influyen en el rendimiento académico medido por el porcentaje de créditos superados.

5.2. LIMITACIONES

Las principales limitaciones de esta investigación son:

1. Los modelos multinivel establecidos valoran la influencia de las variables del estudiante, organización académica y docente en el rendimiento académico universitario y al no contar con variables adicionales inherentes al estudiante que también influyen en el rendimiento académico como coeficiente intelectual, aptitud, formación académica previa a la universidad, nota media de acceso a la universidad, situación socioeconómica de los padres, estudios y profesión de los padres; y de variables inherentes al docente como metodología que utiliza en clase y métodos de evaluación no se pueden constituir como modelos predictivos.
2. Al medir el rendimiento académico de los estudiantes universitarios de forma global, resulta complejo medir 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.
3. Existe una gran variedad de asignaturas que se valoran de forma global, provocando ciertas limitaciones para ajustar el rendimiento del alumno, no obstante los resultados del modelo son válidos para todas las materias de forma conjunta.

5.3. FUTURAS INVESTIGACIONES

1. Adaptación de otros modelos estadísticos para medir el rendimiento académico universitario, como modelos multivariantes multinivel que son modelos poco utilizados y que no se encuentran en fases muy desarrolladas desde la perspectiva del software.
2. Aplicación de modelos multinivel considerando las variables inherentes al estudiante y al docente que también influyen en el rendimiento académico universitario como coeficiente intelectual, aptitud, formación académica previa a la universidad, nota media de acceso a la universidad, situación socioeconómica de los padres, estudios y profesión de los padres; y de variables inherentes al docente como metodología que utiliza en clase, métodos de evaluación y otras. Para el efecto será necesario delimitar la población a los estudiantes que toman materias específicas o que están cursando sus estudios en un semestre específico.

BIBLIOGRAFÍA

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6), 716-723.
- Al-Hattami, A. A. (2012). *Differential predictive validity of high school GPA and college entrance test scores for university students in Yemen (Doctoral dissertation, University of Pittsburgh)*.
- Ayán, M. N., & Díaz, M. Á. (2011). Indicadores de rendimiento de estudiantes universitarios: calificaciones versus créditos acumulados. *Revista de Educación*, (355), 467-492.
- Barrow, M., Reilly, B., & Woodfield, R. (2009). The determinants of undergraduate degree performance: how important is gender? *British Educational Research Journal*, 35(4), 575-597.
- Bermúdez, M. P., Guillén-Riquelme, A., Gómez-García, A., Quevedo-Blasco, R., Sierra, J. C., & Buela-Casal, G. (2011). Análisis del rendimiento en el doctorado en función del sexo. *Educación XX1*, 14(1), 17-33.
- Browne, W., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, 1, 103-124.
- Buckalew, L. W., Daly, J. D., & Coffield, K. E. (1986). Relationship of initial class attendance and seating location to academic performance in psychology classes. *Bulletin of the Psychonomic Society*, 24(1), 63-64.
- Buela-Casal, G., Guillén-Riquelme, A., Guglielmi, O., Quevedo-Blasco, R., & Ramiro, M. T. (2011). Rendimiento en el doctorado en función del área de conocimiento. *Revista de Psicodidáctica*, 16(1), 181-192.
- Byrne, M., & Flood, B. (2008). Examining the relationships among background variables and academic performance of first year accounting students at an Irish University. *Journal of Accounting Education*, Volume 26, Issue 4, 202-212.
- Callender, C. (2008). The impact of term-time employment on higher education students' academic attainment and achievement. *Journal of Education Policy*, 23(4), 359-377.
- Cantwell, R., Archer, J., & Bourke, S. (2001). A comparison of the academic experiences and achievement of university students entering by traditional and non-traditional means. *Assessment & Evaluation in Higher Education*, 26(3), 221-234.

- Castagnetti, C., & Rosti, L. (2009). Effort allocation in tournaments: The effect of gender on academic performance in Italian universities. *Economics of Education Review*, 28(3), 357-369.
- Centre for Multilevel Modelling. University of Bristol. (s.f.). Obtenido de <http://www.bristol.ac.uk/cmm/learning/multilevel-models/data-structures.html>
- Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, 37(4), 319-338.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778-820.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673-682.
- Considine, G., & Zappalá, G. (2002). The influence of social and economic disadvantage in the academic performance of school students in Australia. *Journal of Sociology*, 38(2), 129-148.
- Cortés, A., & Palomar, J. (2008). El proceso de admisión como predictor del rendimiento académico en la educación superior. *Universitas psychologica*, 7(1), 199-215.
- Coto, M. S. (2007). *Algunos factores asociados con el rendimiento académico de los estudiantes regulares de grado de la Universidad Nacional*. Costa Rica: IDESPO - Universidad Nacional.
- Crawford, I., & Wang, Z. (2014). The impact of individual factors on the academic attainment of Chinese and UK students in higher education. *Studies in Higher Education*, 1-19.
- Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, 26(3), 312-324.
- Dayioğlu, M., & Türüt-Aşık, S. (2007). Gender differences in academic performance in a large public university in Turkey. *Higher Education*, 53(2), 255-277.
- de Miguel Díaz, M., Urquijo, P. A., Blanco, J. M., Escorza, T. E., Espinar, S. R., & García, J. V. (2002). Evaluación del rendimiento en la enseñanza superior. Comparación de resultados entre alumnos procedentes de la LOGSE y del COU. *Revista de Investigación Educativa*, 20(2), 357-383.

- Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and individual differences*, 36(8), 1907-1920.
- Dutton, J., & Dutton, M. (2005). Characteristics and Performance of Students in an Online Section of Business Statistics. *Journal of Statistics Education*, 13(3), 1-26.
- Duvall, C. K., & Schwartz, R. G. (2000). Distance education: relationship between academic performance and technology-adept adult students. *Education and Information Technologies*, 5(3), 177-187.
- ECTS: Guía de uso. European Commission. (6 de Febrero de 2009). <http://ec.europa.eu>. Recuperado el 18 de Febrero de 2015, de http://ec.europa.eu/education/tools/docs/ects-guide_es.pdf
- Edel, R. (2003). El rendimiento académico: concepto, investigación y desarrollo. *REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2).
- El Ansari, W. (2002). Student nurse satisfaction levels with their courses: Part I—effects of demographic variables. *Nurse education today*, 22(2), 159-170.
- El Ansari, W. (2003). Satisfaction trends in undergraduate physiotherapy education. *Physiotherapy*, 89(3), 171-185.
- Farsides, T., & Woodfield, R. (2007). Individual and gender differences in good and first-class undergraduate degree performance. *British Journal of Psychology*, 98(3), 467-483.
- Federation University Australia. (08 de Octubre de 2014). <http://federation.edu.au>. Recuperado el 19 de Marzo de 2014, de <http://federation.edu.au/current-students/essential-info/administration/examinations-and-results/results/grade-point-average-gpa>
- Fenollar, P., Román, S., & Cuestas, P. (2007). University students' academic performance: An integrative conceptual framework and empirical analysis. *British Journal of Educational Psychology*, 77(4), 873-891.
- Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, 91(5), 901-914.
- ForeignCredits. Translation and Evaluation. (2013). <http://www.foreigncredits.com>. Recuperado el 19 de Marzo de 2015, de <http://www.foreigncredits.com/Credential-Evaluation/>
- Fryer, R. G. (2011). Teacher incentives and student achievement: Evidence from New York City public schools. *National Bureau of Economic Research*.

- García, C., Melchor, M. J., & García, C. (2008). *Análisis de los factores del rendimiento académico: Un estudio aplicado a la Diplomatura en Ciencias Empresariales de la Universidad de Alcalá (2004-2007)*. Alcalá de Henares: Universidad de Alcalá.
- García, L. (1987). *Hacia una definición de Educación a Distancia*. Asociación Iberoamericana de Educación Superior a Distancia.
- García, L. (1988). *Rendimiento académico y abandono en la educación superior a distancia*. Madrid: Universidad Nacional a Distancia.
- García, T., & García, J. C. (2013). Aplicación de cuestionarios online sobre el rendimiento académico. *Revista de Informática Educativa y Medios Audiovisuales*, 10(17), 1-7.
- Garkaz, M., Banimahd, B., & Esmaeili, H. (2011). Factors affecting accounting students' performance: the case of Students at the islamic azad university. *Procedia-Social and Behavioral Sciences*, 29, 122-128.
- Gaviria, J., & Castro, M. (2005). *Modelos Jerárquicos Lineales*. Madrid: La Muralla, S.A.
- Gibb, S. J., Fergusson, D. M., & Horwood, L. J. (2008). Gender differences in educational achievement to age 25. *Australian Journal of Education*, 52(1), 63-80.
- Glewwe, P., Ilias, N., & Kremer, M. (2003). Teacher incentives (No. w9671). *National Bureau of Economic Research*.
- Goldhaber, D. (2006). *Teacher pay reforms: The political implications of recent research*. Center for American Progress. University of Washington.
- Goldhaber, D. (2007). Everyone's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness? *Journal of Human Resources*, 42(4), 765-794.
- Goldhaber, D. D., & Brewer, D. J. (1996). *Evaluating the Effect of Teacher Degree Level on Educational Performance*.
- Goldhaber, D., & Anthony, E. (2003). *Indicators of Teacher Quality*. New York: ERIC Digest.
- Goldstein, H. (2011). *Multilevel Statistical Models (4th Edition)*. Chichester - United Kingdom: J. Wiley & Sons.
- Gómez, M. H. (2009). Gasto en farmacia y médico de atención primara: un enfoque multinivel. *Estadística española*, 51(171), 331-361.
- Gracia, L., & Jenkins, E. (2003). A quantitative exploration of student performance on an undergraduate accounting programme of study. *Accounting Education*, 12(1), 15-32.

- Halpern, N. (2007). The impact of attendance and student characteristics on academic achievement: findings from an undergraduate business management module. *Journal of Further and Higher Education, 31(4)*, 335-349.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review, 30(3)*, 466-479.
- Houston, M., & Rimmer, R. (2005). A comparison of academic outcomes for business and other students. *The international journal of management education, 4(3)*, 11-20.
- Hunt, A., Lincoln, I., & Walker, A. (2004). Term-time employment and academic attainment: evidence from a large-scale survey of undergraduates at Northumbria University. *Journal of Further and Higher Education, 28(1)*, 3-18.
- Jansen, E. P., & Bruinsma, M. (2005). Explaining achievement in higher education. *Educational Research and Evaluation, 11(3)*, 235-252.
- Jepsen, C. (2005). Teacher characteristics and student achievement: evidence from teacher surveys. *Journal of Urban Economics, 57(2)*, 302-319.
- Jiménez, M. V., Izquierdo, J. M., & Blanco, A. J. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema, 12(Suplemento)*, 248-252.
- Junco, R. (2012). Too much face and not enough books: The relationship between multiple indices of Facebook use and academic performance. *Computers in Human Behavior, 28(1)*, 187-198.
- Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of political Economy, 110(6)*, 1286-1317.
- Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The future of children, 17(1)*, 87-109.
- Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. *Studies in Higher education, 27(1)*, 27-52.
- Lumb, A. B., & Vail, A. (2004). Comparison of academic, application form and social factors in predicting early performance on the medical course. *Medical education, 38(9)*, 1002-1005.
- McKenzie, K., & Gow, K. (2004). Exploring the first year academic achievement of school leavers and mature-age students through structural equation modelling. *Learning and Individual Differences, 14(2)*, 107-123.

- McKenzie, K., & Schweitzer, R. (2001). Who Succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher education research and development*, 20(1), 21-33.
- Mills, C., Heyworth, J., Rosenwax, L., Carr, S., & Rosenberg, M. (2009). Factors associated with the academic success of first year Health Science students. *Advances in health sciences education*, 14(2), 205-217.
- Ministerio de Educación, Cultura y Deporte. (2014). <https://sede.educacion.gob.es>. Recuperado el 19 de Marzo de 2015, de <https://sede.educacion.gob.es/catalogo-tramites/profesores/formacion/universitarios/fpu.html>
- Moncada, L. F., & Rubio, M. J. (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(2), 77-95.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of education review*, 13(2), 125-145.
- Montero, E., Villalobos, J., & Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. *RELIEVE*, 13(2), 215-234.
- Muralidharan, K., & Sundararaman, V. (2009). Teacher performance pay: Experimental evidence from India (No. w15323). *National Bureau of Economic Research*.
- Ofori, R. (2000). Age and 'type' of domain specific entry qualifications as predictors of student nurses' performance in biological, social and behavioural sciences in nursing assessments. *Nurse education today*, 20(4), 298-310.
- Olmos, E. (1997). El rendimiento estudiantil: Una metodología para su medición. *Revista Economía*, (13), 7-26.
- Paisey, C., & Paisey, N. J. (2004). Student attendance in an accounting module—reasons for non-attendance and the effect on academic performance at a Scottish University. *Accounting education*, 13(sup1), 39-53.
- Paradis, E. (2003). *R para Principiantes*. France: Institut des Sciences de l'Évolution, Université Montpellier II.
- Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-Plus*. New York: Springer Verlag.

- Porto, A., & Di Gresia, L. (2004). Rendimiento de estudiantes universitarios y sus determinantes. *Revista de Economía y Estadística*, 42(1), 93-113.
- Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross classified random structures using a multilevel model. *Journal of Educational and Behavioural Statistics*, 19, 337-350.
- Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2009). *A User's Guide to MLwiN. Version 2.10*. Bristol, United Kingdom: Centre for Multilevel Modelling, University of Bristol.
- Richardson, J. T. (1995). Mature students in higher education: II. An investigation of approaches to studying and academic performance. *Studies in Higher Education*, 20(1), 5-17.
- Richardson, J. T., & Woodley, A. (2003). Another look at the role of age, gender and subject as predictors of academic attainment in higher education. *Studies in Higher Education*, 28(4), 475-493.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 247-252.
- Rodríguez, S., Fita, E., & Torrado, M. (2004). El rendimiento académico en la transición secundaria - universidad. *Revista de Educación*, 334, 391-414.
- Rubio, M. (2009). *Nuevas Orientaciones y Metodología para la Educación a Distancia*. Loja: Universidad Técnica Particular de Loja.
- Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological bulletin*, 135(1), 1-22.
- Sánchez, D. G., Marín, R., & López, Y. E. (2011). Factores que influyen en el rendimiento académico del estudiante universitario. *TECNOCIENCIA Chihuahua*, 5(2).
- Sarramona, J. (1993). El rendimiento en la enseñanza a distancia. *Teoría de la educación*, Vol. V, 127-137.
- Sheard, M. (2009). Hardiness commitment, gender, and age differentiate university academic performance. *British Journal of Educational Psychology*, 79(1), 189-204.

- Smith, J., & Naylor, R. (2001). Determinants of degree performance in UK universities: a statistical analysis of the 1993 student cohort. *Oxford Bulletin of Economics and Statistics*, 63(1), 29-60.
- Smith, R. (2011). *Multilevel models os social problems. A Causal Perspective*. Cambridge, Massachusetts. USA: Springer.
- Springer, M. G., Ballou, D., Hamilton, L., Le, V. N., Lockwood, J. R., McCaffrey, D. F., y otros. (2011). *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)*. Society for Research on Educational Effectiveness.
- Stanca, L. (2006). The effects of attendance on academic performance: Panel data evidence for introductory microeconomics. *The Journal of Economic Education*, 37(3), 251-266.
- Stanley, L. R., Comello, M. L., Edwards, R. W., & Marquart, B. S. (2008). School adjustment in rural and urban communities: Do students from “Timbuktu” differ from their “City Slicker” peers? *Journal of Youth and Adolescence*, 37(2), 225-238.
- Tejedor, F. J., & Muñoz-Repiso, A. G. (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.
- Universidad de Granada. (13 de Mayo de 2010). <http://internacional.ugr.es>. Recuperado el 18 de Mayo de 2015, de [http://internacional.ugr.es/pages/conversion-calificaciones/tablaconversioncalificaciones/!](http://internacional.ugr.es/pages/conversion-calificaciones/tablaconversioncalificaciones/)
- Universidad de Sevilla. (18 de Marzo de 2010). <http://www.us.es>. Recuperado el 17 de Marzo de 2015, de <http://www.us.es/downloads/acerca/normativa/normativa-examenes.pdf>
- Universidad Técnica Particular de Loja. (Febrero de 2015). <http://www.utpl.edu.ec>. Recuperado el 16 de Marzo de 2015, de <http://www.utpl.edu.ec/sites/default/files/pregrado/guia-general-MAD.pdf>
- University of Toronto Scarborough. (2015). <http://www.utsc.utoronto.ca>. Recuperado el 18 de Marzo de 2015, de http://www.utsc.utoronto.ca/webapps/aacc-tools/GPA_calculator/index.cgi
- Van den Berg, M. N., & Hofman, W. H. (2005). Student success in university education: A multi-measurement study of the impact of student and faculty factors on study progress. *Higher Education*, 50(3), 413-446.
- van Rooyen, P., Dixon, A., Dixon, G., & Wells, C. (2006). Entry criteria as predictor of performance in an undergraduate nursing degree programme. *Nurse education today*, 26(7), 593-600.

- 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.
- Vargas, G. M. (2013). Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17(3), 57-87.
- Vegas, E. (2006). Incentivos docentes y sus efectos en el aprendizaje del alumnado en Latinoamérica. *Revista de Educación*, 340, 213-241.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89-122.
- West, B. T., Welch, K., & Galecki, A. (2007). *Linear Mixed Models. A Practical Guide Using Statistical Software*. Boca Raton, USA: Chapman & Hall/CRC.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461-481.
- Wimshurst, K., Wortley, R., Bates, M., & Allard, T. (2006). The impact of institutional factors on student academic results: implications for 'quality' in universities. *Higher Education Research & Development*, 25(02), 131-145.
- Wimshurst, K., Wortley, R., Bates, M., & Allard, T. (2006). The impact of institutional factors on student academic results: implications for 'quality' in universities. *Higher Education Research & Development*, 25(02), 131-145.
- Win, R., & Miller, P. W. (2005). The Effects of Individual and School Factors on University Students' Academic Performance. *Australian Economic Review*, 38(1), 1-18.
- Wojciechowski, A., & Palmer, L. B. (2005). Individual student characteristics: Can any be predictors of success in online classes? *Online Journal of Distance Learning Administration*, 8(2).
- Woodfield, R., Jessop, D., & McMillan, L. (2006). Gender differences in undergraduate attendance rates. *Studies in Higher Education*, 31(1), 1-22.
- Woodfield, R., Jessop, D., & McMillan, L. (2006). Gender differences in undergraduate attendance rates. *Studies in Higher Education*, 31(1), 1-22.
- Zhang, D. (2008). The Effects of Teacher Education Level, Teaching Experience, And Teaching Behaviors On Student Science Achievement. *Graduate Theses and Dissertations*, 155.

Zuzovsky, R. (2003). Teachers' qualifications and their impact on student achievement: Findings from TIMSS 2003 data for Israel. *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments*, 2, 37-62.

Anexos. TRATAMIENTO DE INFORMACIÓN

Anexo 1. UNIÓN DE ARCHIVOS ALUMNOS

```
require("openxlsx")

xlsxFile1 = "datosacademicos/NotasAbr2012Ago2012.xlsx"
df01_hoja1 <- read.xlsx(xlsxFile = xlsxFile1, sheet = 1)
df01_hoja2 <- read.xlsx(xlsxFile = xlsxFile1, sheet = 2)
df01_hoja3 <- read.xlsx(xlsxFile = xlsxFile1, sheet = 3)

datos_abr2012ago2012 = rbind(df01_hoja1,df01_hoja2,df01_hoja3)
save(datos_abr2012ago2012, file="datos_abr2012ago2012.RData")

xlsxFile2 = "datosacademicos/Octubre2011Febrero2012.xlsx"
df02_hoja1 <- read.xlsx(xlsxFile = xlsxFile2, sheet = 1)
df02_hoja2 <- read.xlsx(xlsxFile = xlsxFile2, sheet = 2)
df02_hoja3 <- read.xlsx(xlsxFile = xlsxFile2, sheet = 3)

df02_hoja2 = df02_hoja2[,-(1:2)]
df02_hoja3 = df02_hoja3[,-(1:2)]

rm(df02_hoja2g)

datos_oct2011feb2012 = rbind(df02_hoja1,df02_hoja2,df02_hoja3)
save(datos_oct2011feb2012, file="datos_oct2011feb2012.RData")

xlsxFile3 = "datosacademicos/Abril2011Agosto2011.xlsx"
df03_hoja1 <- read.xlsx(xlsxFile = xlsxFile3, sheet = 1)
df03_hoja2 <- read.xlsx(xlsxFile = xlsxFile3, sheet = 2)
df03_hoja3 <- read.xlsx(xlsxFile = xlsxFile3, sheet = 3)

datos_Abril2011Agosto2011 = rbind(df03_hoja1,df03_hoja2,df03_hoja3)
save(datos_Abril2011Agosto2011, file="datos_Abril2011Agosto2011.RData")

xlsxFile4 = "datosacademicos/Octubre2010Febrero2011.xlsx"
df04_hoja1 <- read.xlsx(xlsxFile = xlsxFile4, sheet = 1)
df04_hoja2 <- read.xlsx(xlsxFile = xlsxFile4, sheet = 2)
df04_hoja3 <- read.xlsx(xlsxFile = xlsxFile4, sheet = 3)

datos_Octubre2010Febrero2011 = rbind(df04_hoja1,df04_hoja2,df04_hoja3)
save(datos_Octubre2010Febrero2011, file="datos_Octubre2010Febrero2011.RData")

rm(df04_hoja3)
save(datos_abr2012ago2012, datos_Abril2011Agosto2011, datos_oct2011feb2012,
      datos_Octubre2010Febrero2011, file='todos.RData')

## paso 2

##load("todos.RData")

## Añadir columna año

col_anolectivo = rep(2012, nrow(datos_abr2012ago2012))
```

```

names(col_anolectivo) = "Año lectivo"
datos_abr2012ago2012mod = cbind(datos_abr2012ago2012, col_anolectivo)

col_anolectivo = rep(2012, nrow(datos_oct2011feb2012))
names(col_anolectivo) = "Año lectivo"
datos_oct2011feb2012mod = cbind(datos_oct2011feb2012, col_anolectivo)

col_anolectivo = rep(2011, nrow(datos_Abril2011Agosto2011))
names(col_anolectivo) = "Año lectivo"
datos_Abril2011Agosto2011mod = cbind(datos_Abril2011Agosto2011, col_anolectivo)

col_anolectivo = rep(2011, nrow(datos_Octubre2010Febrero2011))
names(col_anolectivo) = "Año lectivo"
datos_Octubre2010Febrero2011mod = cbind(datos_Octubre2010Febrero2011, col_anolectivo)

## Unir por años

### 2011

datos_2011 = rbind(datos_Octubre2010Febrero2011mod, datos_Abril2011Agosto2011mod)

datos_2011m = datos_2011[, -c(7, 8, 20)]
names(datos_2011m)
AREAFORMACION = rep('', nrow(datos_2011m))
datos_2011m2 = cbind(datos_2011m[, 1:7], AREAFORMACION, datos_2011m[, 8:20])
names(datos_2011m2)
rm(datos_2011m2)
datos_2011final = datos_2011m2

### 2012

names(datos_abr2012ago2012mod)
datos_abr2012ago2012mod2 = datos_abr2012ago2012mod[, -c(1, 20, 21, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36)]
names(datos_abr2012ago2012mod2)

NOMBRES = paste(datos_abr2012ago2012mod2[, 4], datos_abr2012ago2012mod2[, 2],
                datos_abr2012ago2012mod2[, 3])

datos_abr2012ago2012mod2 = cbind(datos_abr2012ago2012mod2, NOMBRES)
names(datos_abr2012ago2012mod2)

datos_abr2012ago2012mod3 = datos_abr2012ago2012mod2[, c(1, 27, 9, 7, 5, 6, 10, 11, 13, 12, 16, 18, 23, 22, 24, 25, 20,
21, 8, 19, 26)]

names(datos_abr2012ago2012mod3)
names(datos_oct2011feb2012mod)

datos_oct2011feb2012mod2 = datos_oct2011feb2012mod[, -c(7, 8, 20)]
names(datos_oct2011feb2012mod2)

AREAFORMACION = rep('', nrow(datos_oct2011feb2012mod2))
datos_oct2011feb2012mod3 = cbind(datos_oct2011feb2012mod2[, 1:7],
                                AREAFORMACION, datos_oct2011feb2012mod2[, 8:20])

names(datos_oct2011feb2012mod3)
names(datos_abr2012ago2012mod3) = names(datos_oct2011feb2012mod3)
names(datos_abr2012ago2012mod3)

datos_2012final = rbind(datos_oct2011feb2012mod3, datos_abr2012ago2012mod3)

names(datos_2012final)

names(datos_2011final)[21] = "ANO"

```

```

names(datos_2012final)[21] = "ANO"

names(datos_2011final)[10] = "CODIGOCARRERA"
names(datos_2012final)[10] = "CODIGOCARRERA"

names(datos_2011final)[16] = "ESTADONOTA"
names(datos_2012final)[16] = "ESTADONOTA"

## unir 2011 con 2012

datos_20112012 = rbind(datos_2011final,datos_2012final)
names(datos_20112012)

save(datos_20112012,datos_2011final,datos_2012final, file='datosacademicos.RData')

rm(list = ls())

###

load("datosacademicos.RData")

which(datos_20112012$MATRICULA=="Legalizada")[1:10]
datos_20112012$MATRICULA[460534:460536]

primer_registro_abril2012 = 460535

table(datos_20112012$MATRICULA)

v.matri = datos_20112012$MATRICULA

v.matri[v.matri=='ACEPTADA'] = 'ANULADA'
table(v.matri)

datos_20112012$MATRICULA = v.matri

##
table(datos_20112012$ESTADONOTA)

## A = Aprobada
## R = Reprobado

datos_20112012$NOTA[datos_20112012$ESTADONOTA=='AN']
head(datos_20112012[datos_20112012$ESTADONOTA=='AN',])

v.nota = datos_20112012$ESTADONOTA
v.nota[v.nota=='AN'] = 'Anulada'
v.nota[v.nota=='A'] = 'Aprobada'
v.nota[v.nota=='R'] = 'Reprobado'
table(v.nota)
datos_20112012$ESTADONOTA = v.nota

### trabajo con el AREAFORMACION

temp = table(datos_20112012$CARRERA)
temp

table(datos_2012final$AREAFORMACION)

ult_registro = 601560
which(table(datos_20112012[primer_registro_abril2012:ult_registro,9],
           datos_20112012[primer_registro_abril2012:ult_registro,8])>0)

names(datos_20112012)

```

```
write.xlsx(table(datos_20112012[primer_registro_abril2012:ult.registro,9],
                datos_20112012[primer_registro_abril2012:ult.registro,8]),
           file='carrerayareadformacion.xlsx')

save(datos_20112012,datos_2011final,datos_2012final,file='datosacademicos_final.RData')
```


Anexo 2. ARÉA DE FORMACIÓN - TITULACIONES - CÓDIGOS

```

## columnas: CARRERA, CODIGOCARRERA

require("openxlsx")
xlsxFile = "NOMENCLATURA DE CARRERAS//carreras2011.xlsx"
df1 <- read.xlsx(xlsxFile = xlsxFile, sheet = 1,)

primer_registro_abril2012
names(datos_20112012)

cols_carrera2011 = datos_20112012[1:(primer_registro_abril2012-1),c(11,12,10)]

cols_carrera2011 = cbind(cols_carrera2011,rep("",primer_registro_abril2012-1),rep(NA,primer_registro_abril2012-1))

names(cols_carrera2011)[4] = 'nuevacarrera'
names(cols_carrera2011)[5] = 'nuevcodigo'

str(cols_carrera2011)

cols_carrera2011[,3] = as.character(cols_carrera2011[,3])
cols_carrera2011[,4] = as.character(cols_carrera2011[,4])
cols_carrera2011[,5] = as.numeric(cols_carrera2011[,5])

trans_areaformacion = function(codigo) {
  valores = c('SOCIO-HUMAN?STICA', 'ADMINISTRATIVA', 'BIOL?GICA Y BIOM?DICA', 'T?CNICA')
  if (codigo==0) {
    resul = "ELIMINAR"
  } else {
    resul = valores[codigo]
  }
  return(resul)
}

nregistros = nrow(cols_carrera2011)
col01 = names(table(cols_carrera2011[,1]))

for (i in 1:nregistros) {
  ##browser()
  ##i = 26
  carrera = cols_carrera2011[i,1]
  ##carrera = iconv(carrera, "LATIN1", "UTF-8")
  ##ind = which(df1[,1]==carrera)
  if ((i %% 10000)==0) {
    cat(i,file='seguimiento.txt',append = TRUE)
    cat("\n",file='seguimiento.txt',append = TRUE)
    ##cat(Sys.time(),file='seguimiento.txt',append = TRUE)
    cat(as.character(Sys.time()),file="seguimiento.txt",append = TRUE)
    cat("\n",file='seguimiento.txt',append = TRUE)
  }
  if ((i %% 50000)==0) {
    save(cols_carrera2011, file='cols_carreras2011temp.RData')
  }

  ind = which(col01==carrera)
  cols_carrera2011[i,4] = df1[ind,5]
  cols_carrera2011[i,5] = df1[ind,3]
  cols_carrera2011[i,3] = trans_areaformacion(df1[ind,6])
}

save(cols_carrera2011, file='cols_carreras2011.RData')
## paso2_2012

```

```

### (tratamiento de las carreras 2012)

## columnas: CARRERA, CODIGOCARRERA

require("openxlsx")
xlsxFile = "NOMENCLATURA DE CARRERAS//carrerayareadformacion CON CAMBIOS 2012.xlsx"
df1 <- read.xlsx(xlsxFile = xlsxFile, sheet = 1,)

primer_registro_abril2012
names(datos_20112012)

cols_carrera2012 = datos_20112012[primer_registro_abril2012:nrow(datos_20112012),c(11,12,10)]
cuantos2012 = 141026

cols_carrera2012 = cbind(cols_carrera2012,rep("", cuantos2012),rep(NA,cuantos2012))

names(cols_carrera2012)[4] = 'nuevacarrera'
names(cols_carrera2012)[5] = 'nuevocodigo'

str(cols_carrera2012)

cols_carrera2012[,3] = as.character(cols_carrera2012[,3])
cols_carrera2012[,4] = as.character(cols_carrera2012[,4])
cols_carrera2012[,5] = as.numeric(cols_carrera2012[,5])

trans_areaformacion = function(codigo) {
  valores = c('SOCIO-HUMANÍSTICA', 'ADMINISTRATIVA', 'BIOLÓGICA Y BIOMÉDICA', 'TÉCNICA')
  if (codigo==0) {
    resul = "ELIMINAR"
  } else {
    resul = valores[codigo]
  }
  return(resul)
}

nregistros = nrow(cols_carrera2012)
col01 = names(table(cols_carrera2012[,1]))

for (i in 1:nregistros) {
  carrera = cols_carrera2012[i,1]
  if ((i %% 10000)==0) {
    cat(i,file='seguimiento2012.txt',append = TRUE)
    cat("\n",file='seguimiento2012.txt',append = TRUE)
    cat(as.character(Sys.time()),file="seguimiento2012.txt",append = TRUE)
    cat("\n",file='seguimiento2012.txt',append = TRUE)
  }

  ind = which(col01==carrera)
  cols_carrera2012[i,4] = df1[ind,4]
  cols_carrera2012[i,5] = df1[ind,2]
  cols_carrera2012[i,3] = trans_areaformacion(df1[ind,5])
}

save(cols_carrera2012,file='cols_carreras2012.RData')

rownames(cols_carrera2012) = NULL

## Unimos los objetos: cols_carrera2011 y 2012
names(cols_carrera2011) = names(cols_carrera2012)

```

```

unidos_carrera20112012 = rbind(cols_carrera2011,cols_carrera2012)

## lo pegamos al objeto "datos_20112012"

datos_20112012temp = cbind(datos_20112012,unidos_carrera20112012)

## eliminar columnas sobrantes

names(datos_20112012temp)

datos_20112012temp2 = datos_20112012temp[,-c(10,11,12,25,26)]
names(datos_20112012temp2)
datos_20112012temp2 = datos_20112012temp2[,c(1,2,3,4,5,6,7,8,9,22,23,24,10,11,12,13,14,15,16,17,18,19,20,21)]
names(datos_20112012temp2)
names(datos_20112012temp2)[11] = 'CARRERA'
names(datos_20112012temp2)[12] = 'CODIGOCARRERA'
names(datos_20112012temp2)

## ELIMINAMOS NA DE CODIGOCARRERA

datos_20112012temp3 = datos_20112012temp2[!is.na(datos_20112012temp2$CODIGOCARRERA),]

summary(datos_20112012temp3$CODIGOCARRERA)
tablita = table(datos_20112012temp3$CARRERA)

rownames(tablita)

valor = rownames(tablita)[24]
datos_20112012temp3[datos_20112012temp3$CARRERA==valor,11] = "INGENIERÍA QUÍMICA"
valor = rownames(tablita)[26]
datos_20112012temp3[datos_20112012temp3$CARRERA==valor,11] = "INGENIERÍA EN ADMINISTRACIÓN DE EMPRESAS TURÍSTICAS Y HOTELERAS"
valor = rownames(tablita)[27]
datos_20112012temp3[datos_20112012temp3$CARRERA==valor,11] = "INGENIERÍA EN ADMINISTRACIÓN DE EMPRESAS TURÍSTICAS Y HOTELERAS/INGENIERÍA EN HOTELERÍA Y TURISMO"

valor = rownames(tablita)[36]
datos_20112012temp3[datos_20112012temp3$CARRERA==valor,11] = "SISTEMAS INFORMÁTICOS Y COMPUTACIÓN"

## CORREGIR PROBLEMA: DISTANCIA / PRESENCIAL EN NOMBRE CARRERA Y EN COLUMNA CODIGOCARRERA
tablita2 = table(datos_20112012temp3$CODIGOCARRERA)
rownames(tablita2)

valor = rownames(tablita2)[24]
datos_20112012temp3[datos_20112012temp3$CODIGOCARRERA==valor,c(11,12,13)]

valor = rownames(tablita2)[24]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia')
datos_20112012temp3[condicion,11] = "CIENCIAS DE LA EDUCACIÓN MENCIÓN INGLÉS"
datos_20112012temp3[condicion,12] = '03106'
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial')
datos_20112012temp3[condicion,11] = "INGLÉS"
datos_20112012temp3[condicion,12] = '12533'

valor = rownames(tablita2)[46]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia')
datos_20112012temp3[condicion,11] = "INGENIERÍA EN ADMINISTRACIÓN DE EMPRESAS TURÍSTICAS Y HOTELERAS"
datos_20112012temp3[condicion,12] = '08902'
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial')

```

```

')
datos_20112012temp3[condicion,11] = "INGENIERÍA EN HOTELERÍA Y TURISMO"
datos_20112012temp3[condicion,12] = '01414'

## RECALCULO DE TABLITA2

valor = rownames(tablita2)[26]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "ADMINISTRACIÓN EN BANCA Y FINANZAS"
datos_20112012temp3[condicion,12] = '04371'
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "ADMINISTRACIÓN EN BANCA Y FINANZAS"
datos_20112012temp3[condicion,12] = '01408'

tablita2 = table(datos_20112012temp3$CODIGOCARRERA)
valor = rownames(tablita2)[29]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "COMUNICACIÓN SOCIAL"
datos_20112012temp3[condicion,12] = '04376'
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "COMUNICACIÓN SOCIAL"
datos_20112012temp3[condicion,12] = '01413'

tablita2 = table(datos_20112012temp3$CODIGOCARRERA)
valor = rownames(tablita2)[30]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "ECONOMÍA"
datos_20112012temp3[condicion,12] = '04377'
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "ECONOMÍA"
datos_20112012temp3[condicion,12] = '01401'

tablita2 = table(datos_20112012temp3$CODIGOCARRERA)
valor = rownames(tablita2)[32]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "PSICOLOGÍA"
datos_20112012temp3[condicion,12] = "04379"
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "PSICOLOGÍA"
datos_20112012temp3[condicion,12] = '01854'

tablita2 = table(datos_20112012temp3$CODIGOCARRERA)
valor = rownames(tablita2)[33]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "GESTIÓN AMBIENTAL"
datos_20112012temp3[condicion,12] = "04524"
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "GESTIÓN AMBIENTAL"
datos_20112012temp3[condicion,12] = '01404'

tablita2 = table(datos_20112012temp3$CODIGOCARRERA)

```

```

valor = rownames(tablita2)[34]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "ADMINISTRACIÓN DE EMPRESAS"
datos_20112012temp3[condicion,12] = "06090"
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "ADMINISTRACIÓN DE EMPRESAS"
datos_20112012temp3[condicion,12] = '01407'

tablita2 = table(datos_20112012temp3$CODIGOCARRERA)
valor = rownames(tablita2)[36]
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Distancia'
)
datos_20112012temp3[condicion,11] = "INGENIERÍA EN CONTABILIDAD Y AUDITORÍA"
datos_20112012temp3[condicion,12] = "08898"
condicion = (datos_20112012temp3$CODIGOCARRERA==valor) & (datos_20112012temp3$MODALIDAD=='Presencial'
')
datos_20112012temp3[condicion,11] = "INGENIERÍA EN CONTABILIDAD Y AUDITORÍA"
datos_20112012temp3[condicion,12] = '01405'

alumnos_20112012 = datos_20112012temp3
rownames(alumnos_20112012) = NULL
save(alumnos_20112012, file="alumnos_20112012.RData")

names(alumnos_20112012)
table(alumnos_20112012[,c(15)])

alumnos_20112012[alumnos_20112012$COMPONENTE==rownames(tt1)[2418],c(15,16)]
alumnos_20112012[alumnos_20112012$COMPONENTE==rownames(tt1)[2417],c(15,16)]

tt1=table(alumnos_20112012[,c(15)])
tt2=table(alumnos_20112012[,c(16)])
length(rownames(tt2))
rownames(tt2)[4505]

table(alumnos_20112012$AREAFORMACION)

## problemas con componente y codigo de componente

require(sqldf)
codycomp_ord_codigo = sqldf("select distinct ASG_CODIGO,COMPONENTE from alumnos_20112012")
codycomp_ord_compon = sqldf("select distinct ASG_CODIGO,COMPONENTE from alumnos_20112012 order by CO
MPONENTE")

write.xlsx(codycomp_ord_codigo, file='codycomp_ord_codigo.xlsx')
write.xlsx(codycomp_ord_compon, file='codycomp_ord_compon.xlsx')

```

Anexo 3. CENTROS UNIVERSITARIOS - MODALIDAD

```

## problemas con centro: mayúsculas,...
alumnos_20112012$CENTRO = toupper(alumnos_20112012$CENTRO)

(tablita4 = table(alumnos_20112012$CENTRO))

length(which(alumnos_20112012$CENTRO=='GUAYAQUIL-CENTENARIO'))

alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='GUAYAQUIL-CENTENARIO'] = 'GUAYAQUIL - CENTENARIO'

## ELIMINAR ALUMNOS COMPLETOS DE LOS CENTROS SIGUIENTES:

length(which(alumnos_20112012$CENTRO=='CARIAMANGA EXTENSIÓN'))
alumnos_20112012 = alumnos_20112012[alumnos_20112012$CENTRO!='CARIAMANGA EXTENSIÓN',]
alumnos_20112012 = alumnos_20112012[alumnos_20112012$CENTRO!='ZAMORA EXTENSIÓN',]

#Load("alumnos_20112012final.RData")
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='BAHÍA CARÁQUEZ'] = 'BAHÍA DE CARÁQUEZ'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='EL ANGEL'] = 'EL ÁNGEL'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='ARMADA NACIONAL'] = 'GUAYAQUIL'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='CENTRO DE REHABILITACION - GUAYAQUIL'] = 'GUAYAQUI
L'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='CENTRO DE REHABILITACION - QUITO'] = 'QUITO'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='ESMA'] = 'SALINAS'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='LA ISABELA'] = 'ISABELA'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (BARCELONA)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (BILBAO)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (LONDRES)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (MURCIA)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (PALMA DE MALLORCA)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (PAMPLONA)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (SUIZA)'] = 'MADRID'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='MADRID (VALENCIA)'] = 'MADRID'

alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='ROMA (GÉNOVA)'] = 'ROMA'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='ROMA (MILÁN)'] = 'ROMA'
alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='ROMA (PARIS)'] = 'ROMA'

length(alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='SANTA ISABEL'])
length(alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='SANTA ISABEL '])

alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='SANTA ISABEL '] = 'SANTA ISABEL'

alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='SHUSHUFINDI (LIMONCOCHA)'] = 'SHUSHUFINDI'

alumnos_20112012$CENTRO[alumnos_20112012$CENTRO=='PUCARA'] = 'PUCARÁ'

## ELIMINAR REGISTROS DE: BOLIVIA (CENTRO) Y BOLIVIA (SEDE)
alumnos_20112012T = alumnos_20112012[alumnos_20112012$CENTRO!='BOLIVIA (CENTRO)',]
alumnos_20112012T = alumnos_20112012T[alumnos_20112012T$CENTRO!='BOLIVIA (SEDE)',]
rownames(alumnos_20112012T) = NULL

alumnos_20112012 = alumnos_20112012T

## PROBLEMAS CON MODALIDAD

alumnos_20112012$MODALIDAD[alumnos_20112012$MODALIDAD=='ABIERTA'] = 'DISTANCIA'
alumnos_20112012$MODALIDAD[alumnos_20112012$MODALIDAD=='CLASICA'] = 'PRESENCIAL'
alumnos_20112012$MODALIDAD = toupper(alumnos_20112012$MODALIDAD)

```

Anexo 4. CIUDAD - PROVINCIA - CANTÓN

```

## ASOCIAR CIUDAD A ALUMNOS QUE NO TIENEN EN ALGUNOS REGISTROS Y SI EN OTROS REGISTROS

#CIUDAD
condicion = (alumnos_20112012$CIUDAD=='')

temp2 = table(alumnos_20112012$NOMBRES[condicion])
length(rownames(temp2))

numero.al = length(rownames(temp2))
for (i in 1:numero.al) {
  tablaciudades = table(alumnos_20112012$CIUDAD[alumnos_20112012$NOMBRES==rownames(temp2)[i]])
  num.filas = length(tablaciudades)
  if (num.filas>1) {
    alumnos_20112012$CIUDAD[alumnos_20112012$NOMBRES==rownames(temp2)[i]] = rownames(tablaciudades)[
2]
  }
}

## canton

condicion = (alumnos_20112012$CANTON=='')

temp2 = table(alumnos_20112012$NOMBRES[condicion])
length(rownames(temp2))

numero.al = length(rownames(temp2))
for (i in 1:numero.al) {
  tablaciudades = table(alumnos_20112012$CANTON[alumnos_20112012$NOMBRES==rownames(temp2)[i]])
  num.filas = length(tablaciudades)
  if (num.filas>1) {
    alumnos_20112012$CANTON[alumnos_20112012$NOMBRES==rownames(temp2)[i]] = rownames(tablaciudades)[
2]
  }
}

## provincia

condicion = (alumnos_20112012$PROVINCIA=='')

temp2 = table(alumnos_20112012$NOMBRES[condicion])
length(rownames(temp2))

numero.al = length(rownames(temp2))
for (i in 1:numero.al) {
  tablaciudades = table(alumnos_20112012$PROVINCIA[alumnos_20112012$NOMBRES==rownames(temp2)[i]])
  num.filas = length(tablaciudades)
  if (num.filas>1) {
    alumnos_20112012$PROVINCIA[alumnos_20112012$NOMBRES==rownames(temp2)[i]] = rownames(tablaciudades)[
2]
  }
}

## DIRECCION

condicion = (alumnos_20112012$DIRECCION=='')

temp2 = table(alumnos_20112012$NOMBRES[condicion])
length(rownames(temp2))

numero.al = length(rownames(temp2))
for (i in 1:numero.al) {
  tablaciudades = table(alumnos_20112012$DIRECCION[alumnos_20112012$NOMBRES==rownames(temp2)[i]])

```

```

num.filas = length(tablaciudades)
if (num.filas>1) {
  alumnos_20112012$DIRECCION[alumnos_20112012$NOMBRES==rownames(temp2)[i]] = rownames(tablaciudades)[2]
}
}

save(alumnos_20112012, file="alumnos_20112012lunes.RData")

## SOBRE PROVINCIA Y CANTON RECODIFICAMOS CON NA

(tablita4 = table(alumnos_20112012$PROVINCIA))

length(which(alumnos_20112012$PROVINCIA=='DESCONOCIDA'))
alumnos_20112012$PROVINCIA[alumnos_20112012$PROVINCIA=='DESCONOCIDA'] = NA
alumnos_20112012$PROVINCIA[alumnos_20112012$PROVINCIA=='(OTRA)'] = NA
length(which(alumnos_20112012$PROVINCIA==' '))

alumnos_20112012$PROVINCIA[alumnos_20112012$PROVINCIA==' '] = NA

(tablita5 = table(alumnos_20112012$CANTON))

alumnos_20112012$CANTON[alumnos_20112012$CANTON=='DESCONOCIDO'] = NA
alumnos_20112012$CANTON[alumnos_20112012$CANTON==' '] = NA
alumnos_20112012$CANTON[alumnos_20112012$CANTON=='(OTRA)'] = NA
alumnos_20112012$CANTON[alumnos_20112012$CANTON=='NO APLICA'] = NA
alumnos_20112012$CANTON[alumnos_20112012$CANTON=='NO DEFINIDO'] = NA

## SOBRE CIUDAD Y DIRECCION RECODIFICAMOS ESPACIOS EN BLANCO COMO NA

(tablita6 = table(alumnos_20112012$CIUDAD))
length(which(alumnos_20112012$CIUDAD==' '))
alumnos_20112012$CIUDAD[alumnos_20112012$CIUDAD==' '] = NA
length(which(alumnos_20112012$CIUDAD==''))
alumnos_20112012$CIUDAD[alumnos_20112012$CIUDAD==''] = NA

(tablita7 = table(alumnos_20112012$DIRECCION))
length(which(alumnos_20112012$DIRECCION==' '))
length(which(alumnos_20112012$DIRECCION==''))
alumnos_20112012$DIRECCION[alumnos_20112012$DIRECCION==''] = NA

save(alumnos_20112012, file="alumnos_20112012final.RData")

```


Anexo 5. HÁBITAT

```

## añadir columna HABITAT

col_habitat = c(rep('RURAL',2),rep('URBANA',1),rep('RURAL',1),rep('URBANA',1),rep('RURAL',9),rep('UR
BANA',2),
               rep('RURAL',4),rep('URBANA',1),rep('RURAL',4),rep('URBANA',3),rep('RURAL',2),rep('UR
BANA',1),
               rep('RURAL',4),rep('URBANA',1),rep('RURAL',2),rep('URBANA',1),rep('RURAL',1),rep('UR
BANA',1),
               rep('RURAL',1),rep('URBANA',1),rep('EXTRANJERO',1),rep('RURAL',4),rep('EXTRANJERO',1
),rep("URBANA",1),
               rep('RURAL',8),rep('URBANA',1),rep('RURAL',1),rep('URBANA',1),rep('RURAL',2),rep('UR
BANA',2),
               rep('RURAL',1),rep('URBANA',4),rep("EXTRANJERO",1),rep('RURAL',2),rep('URBANA',1),re
p('RURAL',6),rep('URBANA',1),
               rep('RURAL',6),rep('URBANA',2),rep('RURAL',3),rep('URBANA',1),rep('RURAL',2))
col_habitat = c(col_habitat[1:35], "RURAL", col_habitat[36:95])
col_CENTRO = rownames(tablita4)

df = data.frame(col_CENTRO,col_habitat)

alumnos_20112012T = cbind(alumnos_20112012, "HABITAT"=rep(NA, 589995))

for (i in 1:96) {
  alumnos_20112012T$HABITAT[alumnos_20112012T$CENTRO==col_CENTRO[i]] = col_habitat[i]
}

names(alumnos_20112012T)
alumnos_20112012T2 = alumnos_20112012T[,c(1:8,25,9:24)]
names(alumnos_20112012T2)

alumnos_20112012 = alumnos_20112012T2

save(alumnos_20112012, file="alumnos_20112012final_02072014.RData")

```

Anexo 6. FECHA DE NACIMIENTO

```
##problemas con fecha_nacimiento

TTEMP2 = convertToDate(as.numeric(alumnos_20112012$FECHA_NACIMIENTO))
summary(TTEMP2)
tablita33 = table(TTEMP2)
tablita33[12160:12218]
str(TTEMP2)
alumnos_20112012$FECHA_NACIMIENTO = TTEMP2

## convertir a na, fechas en límites extraños
which(alumnos_20112012$FECHA_NACIMIENTO <= '1944-12-31')
which(is.na(alumnos_20112012$FECHA_NACIMIENTO))
which(alumnos_20112012$FECHA_NACIMIENTO <= '1946-12-31')
which(alumnos_20112012$FECHA_NACIMIENTO >= '1996-1-1')

alumnos_20112012$FECHA_NACIMIENTO[alumnos_20112012$FECHA_NACIMIENTO <= '1944-12-31'] = NA
alumnos_20112012$FECHA_NACIMIENTO[alumnos_20112012$FECHA_NACIMIENTO >= '1996-01-01'] = NA
```

Anexo 7. EDAD

```
## obtener la edad: 2012 - año de nacimiento

temp = 2012-as.integer(strftime(alumnos_20112012conbecas$FECHA_NACIMIENTO,format="%Y"))
table(temp)

alumnos_20112012conbecast = cbind(alumnos_20112012conbecas,EDAD=temp)
names(alumnos_20112012conbecast)
View(alumnos_20112012conbecast[,c(5,31)])
alumnos_20112012conbecas = alumnos_20112012conbecast[,c(1:5,31,6:30)]

## guardamos en fichero R

save(alumnos_20112012conbecas,file='alumnos_20112012finalconbecas_03072014.RData')
##load("alumnos_20112012finalconbecas_03072014.RData")
```

Anexo 8. CRÉDITOS

```

library(dplyr)
library(openxlsx)
codigoscreditos_utpl = read.xlsx("Oferta4PeriodosV2.xlsx")
pre_codigoscreditos_utpl = codigoscreditos_utpl[codigoscreditos_utpl$NIVEL_ACADEMICO=="PREGRADO",]

colnames(pre_codigoscreditos_utpl)[11] = "ASG_CODIGO"

tbl_pre_codigoscreditos_utpl = tbl_df(pre_codigoscreditos_utpl)

### obtener por código asignatura el mínimo de créditos y el máximo
pru01 =
  tbl_pre_codigoscreditos_utpl %>%
  group_by(ASG_CODIGO) %>%
  summarise(MINCRE=min(`Número.de.Créditos`),MAXCRE=max(`Número.de.Créditos`),
            DIFCRE=MAXCRE-MINCRE)
## hay 1590 códigos de asignaturas distintos

## OBTENER TODOS LOS CÓDIGOS DE ASIGNATURAS EXISTENTES EN TABLA DE DATOS DE "ALUMNOS:"

##### CÁLCULOS PREVIOS PARA INVESTIGAR CRÉDITOS SOSPECHOSOS

temp3 =
  tbl_df(alumnos3) %>%
  select(ASG_CODIGO,COMPONENTE,CODIGOCARRERA,CARRERA,CREDITOS,PERIODO,MODALIDAD,AREAFORMACION) %>%
  group_by(ASG_CODIGO,CODIGOCARRERA) %>%
  summarise(FCODASIG= first(ASG_CODIGO),FCOMP=first(COMPONENTE),FPER=first(PERIODO),FCCAR=first(CODI
GOCARRERA),FCAR=first(CARRERA),FCRE=func_mejora_creditos(CREDITOS),
            MODALI=first(MODALIDAD),AREAFORMA=first(AREAFORMACION)) %>%
  select(FCODASIG,FCOMP,FCRE,FPER,FCCAR,FCAR,MODALI,AREAFORMA) %>%
  ## filter(FCRE==0 | is.na(FCRE) | FCRE>=8) %>%
  arrange(FCCAR)

names(temp3)
temp3reducido = temp3[,c(8,9,6,7,1,3,4)]

t_temp3reducido = tbl_df(temp3reducido)
temp3reducido2 = arrange(t_temp3reducido,MODALI,AREAFORMA,FCCAR,ASG_CODIGO,FCOMP)
names(temp3reducido2)
temp3reducido2$FCRE[2]

##library(openxlsx)
##write.xlsx(data.frame(temp3reducido2),file="creditos_carreras_provisional.xlsx")

##### FIN DE CÁLCULOS PREVIOS

### ¿Cuántos códigos de asignaturas distintos hay en temp3reducido2?

temp = table(temp3reducido2$ASG_CODIGO)
length(temp)
## 4505 (código-asignatura y carrera)

codigosasignaturas_en_alumnos =
  temp3reducido2 %>%
  group_by(ASG_CODIGO) %>%
  summarise(ASIGNATURA_CODIGO=first(ASG_CODIGO),ASIGNATURA_NOMBRE=first(FCOMP),CREDITOS_MAL=first(FC
RE)) %>%
  select(ASIGNATURA_CODIGO,ASIGNATURA_NOMBRE,CREDITOS_MAL)

```

```

##
pru02 = left_join(codigosasignaturas_en_alumnos,pru01,by=c("ASIGNATURA_CODIGO"="ASG_CODIGO"))

CREDITOSBUENOS = rep(NA,nrow(pru02))
for (i in 1:nrow(pru02)) {
  if (!is.na(pru02$DIFCRE[i])) {
    CREDITOSBUENOS[i] = pru02$MAXCRE[i]
  } else {
    if (!is.na(pru02$CREDITOS_MAL[i])) {
      if ( (pru02$CREDITOS_MAL[i]>=1) & (pru02$CREDITOS_MAL[i]<=7) ) {
        CREDITOSBUENOS[i] = pru02$CREDITOS_MAL[i]
      }
    }
  }
}

CREDITOSBUENOS[CREDITOSBUENOS==0] = NA

tabla_codigosasignaturas_creditos_final = cbind(pru02,CREDITOSBUENOS)

length(tabla_codigosasignaturas_creditos_final$CREDITOSBUENOS)
table(tabla_codigosasignaturas_creditos_final$CREDITOSBUENOS)
summary(tabla_codigosasignaturas_creditos_final$CREDITOSBUENOS)

cod_asignatura_concreditos_na = tabla_codigosasignaturas_creditos_final$ASIGNATURA_CODIGO[is.na(tabla_codigosasignaturas_creditos_final$CREDITOSBUENOS)]

## ESTUDIAMOS SI LOS CÓDIGOS DE ASIGNATURAS CON NA, SUS NOMBRES APARECEN EN EL LISTADO QUE NOS HAN ENVIADO

tbl_cod_asignatura_concreditos_na = tbl_df(data.frame(ASIGNATURA_CODIGO=cod_asignatura_concreditos_na))
temp = inner_join(codigosasignaturas_en_alumnos,tbl_cod_asignatura_concreditos_na,by=c("ASIGNATURA_CODIGO"="ODIGO"))

names(tbl_pre_codigoscreditos_utpl)
colnames(tbl_pre_codigoscreditos_utpl)[12] = "ASIGNATURA"
names(temp)
temp_nombasig = inner_join(tbl_pre_codigoscreditos_utpl,temp,by=c("ASIGNATURA"="ASIGNATURA_NOMBRE"))

names(temp_nombasig)
temp_nombasig_red = temp_nombasig[,c(21,22,13,12)]
colnames(temp_nombasig_red)[3] = "NUMCREDITOS"

temp40 = temp_nombasig_red %>%
  group_by(ASIGNATURA_CODIGO) %>%
  summarise(AS_CODIGO=first(ASIGNATURA_CODIGO),
            CREMAL=first(CREDITOS_MAL),
            MINCRE=min(NUMCREDITOS),
            MAXCRE=max(NUMCREDITOS),DIFCRE=MAXCRE-MINCRE)

CREDITOSBUENOS2 = rep(NA,nrow(temp40))
for (i in 1:nrow(temp40)) {
  if (temp40$DIFCRE[i]==0) {
    CREDITOSBUENOS2[i] = temp40$MAXCRE[i]
  } else {
    if (is.na(temp40$CREMAL[i])) {
      CREDITOSBUENOS2[i] = temp40$MAXCRE[i]
    } else {
      if ((temp40$CREMAL[i]>=1) & (temp40$CREMAL[i]<=15)) {
        CREDITOSBUENOS2[i] = temp40$CREMAL[i]
      } else {
        CREDITOSBUENOS2[i] = temp40$MAXCRE[i]
      }
    }
  }
}

```

```

    }
  }
}

CREDITOSBUENOS2[CREDITOSBUENOS2==0] = NA

table(CREDITOSBUENOS2)
summary(CREDITOSBUENOS2)

# CREDITOSBUENOS2
#  2  3  4  5  6  7  8  9 10 12 13 14 15 16 20 28
#  4  5 212 192 40  4 17  2  6  2  2  5  4  2  1  1
# > summary(CREDITOSBUENOS2)
#   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
#  2.000  4.000  5.000  5.156  5.000 28.000   145
# >

temp40bueno = cbind(temp40,CREDITOSBUENOS2)

tabla_codigosasignaturas_creditos_final2 = tabla_codigosasignaturas_creditos_final
for (i in 1:nrow(temp40bueno)) {
  cod = temp40bueno$ASIGNATURA_CODIGO[i]
  indices = tabla_codigosasignaturas_creditos_final2$ASIGNATURA_CODIGO==cod
  creditosbien = temp40bueno$CREDITOSBUENOS2[i]
  tabla_codigosasignaturas_creditos_final2$CREDITOSBUENOS[indices] = creditosbien
}

cod_asignatura_concreditos_na2 = tabla_codigosasignaturas_creditos_final2$ASIGNATURA_CODIGO[is.na(tabla_codigosasignaturas_creditos_final2$CREDITOSBUENOS)]

temp41 = tabla_codigosasignaturas_creditos_final2[is.na(tabla_codigosasignaturas_creditos_final2$CREDITOSBUENOS),c(1,2,3,7)]
#names(tabla_codigosasignaturas_creditos_final2)

head(tabla_codigosasignaturas_creditos_final2)
dim(tabla_codigosasignaturas_creditos_final2)

colnames(temp41)
colnames(temp3reducido2)
tbl_temp41 = tbl_df(temp41)
colnames(tbl_temp41)
class(tbl_temp41$ASIGNATURA_CODIGO)
class(temp3reducido2$ASG_CODIGO)
class(temp3reducido2)
temp42 = inner_join(tbl_temp41,temp3reducido2,by=c("ASIGNATURA_CODIGO"="ASG_CODIGO"))
head(temp42)
dim(temp42)

write.xlsx(data.frame(temp42),file="codigosasignaturas_creditos_na.xlsx")

## SE LEE FICHERO REVISADO MANUALMENTE

creditosfinales = read.xlsx("creditos_finales.xlsx")

listado_a_borrar = creditosfinales$ASIGNATURA_CODIGO[creditosfinales$CREDCORREGUIDOS==0]
listado_a_borrar
listado_a_borrar_distintos = names(table(listado_a_borrar))

# [1] "10710001"      "5320001"      "5720001"      "5720002"      "DIST-TNEB001" "EXACOMP"      "EXA
UBI"           "EXUBICA"

```

```

# [9] "JAS"

##

tb_creditosfinales = tbl_df(creditosfinales)

tb_solo_codasig_y_creditos =
tb_creditosfinales %>%
  distinct(ASIGNATURA_CODIGO) %>%
  select(ASIGNATURA_CODIGO,CREDCORREGUIDOS)

summary(tb_solo_codasig_y_creditos$CREDCORREGUIDOS)
summary(creditosfinales$CREDCORREGUIDOS)

creditosfinales$ASIGNATURA_CODIGO[is.na(creditosfinales$CREDCORREGUIDOS)]

# [1] "BF1015A"      "CH-PD01"      "DIST-TNCJ038-A" "EC-TE-1002"    "EI0220"      "EI0225"
"EI0305"
# [8] "EI0620"      "FM0410A"      "FM0410A"      "FM0420A"      "GAA200"      "GAA500"
"GAA925"
# [15] "II06306"      "II09603"      "LI0810"      "LL0525"

## todos los que tienen NA son realmente un 4 créditos

tb_solo_codasig_y_creditos$CREDCORREGUIDOS[is.na(tb_solo_codasig_y_creditos$CREDCORREGUIDOS)] = 4
summary(tb_solo_codasig_y_creditos$CREDCORREGUIDOS)
table(tb_solo_codasig_y_creditos$CREDCORREGUIDOS)

save(alumnos3, file="alumnos3.RData")
save(tb_solo_codasig_y_creditos, file="creditosbuenos.RData")

names(alumnos3)
codigos_temp = alumnos3$ASG_CODIGO
prueba01 = ( codigos_temp %in% listado_a_borrar_distintos)
prueba01
table(prueba01)

prueba02 = alumnos3$COMPONENTE[prueba01]
table(prueba02)
# prueba02
#
#                CURSO DE VERANO
#                1178
#                Examen de Ubicacion de Computación
#                3
#                EXAMEN DE UBICACION DE INGLES
#                46
#                EXAMEN DE UBICACION PARA LOS CURSOS DE INGLES
#                88
#                EXAMEN UBICACION LICENCIATURA INGLES
#                34
#                EXAMEN UBICACION NIVELES INGLES
#                188
#                JORNADA DE ASESORIA DE SISTEMA
#                22756
# JORNADA DE ASESORÍA DE SISTEMA Y TÉCNICAS PARA EL AUTOESTUDIO
#                7134

alumnos4 = alumnos3[prueba01==FALSE,]
save(alumnos4, file="alumnos04.RData")

table(tb_solo_codasig_y_creditos$CREDCORREGUIDOS)
tb_solo_codasig_y_creditos4 = tb_solo_codasig_y_creditos[tb_solo_codasig_y_creditos$CREDCORREGUIDOS!
=0,]
table(tb_solo_codasig_y_creditos4$CREDCORREGUIDOS)

```

```
tempx1 = tabla_codigosasignaturas_creditos_final2[!is.na(tabla_codigosasignaturas_creditos_final2$CR
EDITOSBUENOS),c(1,7)]
table(tempx1$CREDITOSBUENOS)

names(tempx1)
names(tb_solo_codasig_y_creditos4)[2] = "CREDITOSBUENOS"
creditosfinales01 = rbind(tempx1,tb_solo_codasig_y_creditos4)
summary(creditosfinales01)
## hemos conseguido codigos asignaturas con número de créditos correcto
save(creditosfinales01,file="creditosfinales01.RData")
names(alumnos4)

##
class(alumnos4)
class(creditosfinales01)
tb_creditosfinales01 = tbl_df(creditosfinales01)

names(alumnos4)
names(tb_creditosfinales01)
alumnos5 = inner_join(alumnos4,tb_creditosfinales01,by=c("ASG_CODIGO"="ASIGNATURA_CODIGO"))

names(alumnos5)
alumnos6 = alumnos5[,-22]
alumnos6 = alumnos6[,c(1:21,42,22:41)]
names(alumnos6)[22] = "CREDITOS"
names(alumnos6)

save(alumnos6,file="alumnos6.RData")
```


Anexo 9. NOTA MEDIA PONDERADA – PORCENTAJE DE CRÉDITOS APROBADOS – PORCENTAJE DE ASIGNATURAS

```

notas_medias_por_periodo_cedula =
alumnos6 %>%
  group_by(PERIODO, CEDULA) %>%
  summarise(NOTAMEDIA = mean(NOTA))

func_media_ponderada_creditos = function(notas, creditos) {
  sumaproducto = sum(notas*creditos)
  sumacreditos = sum(creditos)
  media = sumaproducto/sumacreditos
  return(media)
}

notas_medias_creditos_por_periodo_cedula =
alumnos6 %>%
  group_by(PERIODO, CEDULA) %>%
  summarise(NOTAMEDIA = func_media_ponderada_creditos(NOTA, CREDITOS))

alumnos6$NOTA[1:20]
alumnos6$ESTADONOTA[1:20]
table(alumnos6$ESTADONOTA)

# x = c(T, F, T)
# y = c(10, 5, 8)
# x*y
# sum(x*y)

func_porcreditosaprobados = function(notaestado, creditos) {
  notasaprobadas = (notaestado!="Reprobado")
  sumaproducto = sum(notasaprobadas*creditos)
  sumacreditos = sum(creditos)
  porcreditosaprobados = (sumaproducto/sumacreditos)*100
  return(porcreditosaprobados)
}

porc_creditos_aprobados_por_periodo_cedula =
alumnos6 %>%
  group_by(PERIODO, CEDULA) %>%
  summarise(PORCREDITOSAPROBADOS = func_porcreditosaprobados(ESTADONOTA, CREDITOS))

func_porasigaprobadas = function(notaestado) {
  notasaprobadas = (notaestado!="Reprobado")
  numaprobadas = sum(notasaprobadas)
  numasignaturas = length(notaestado)
  porasigaprobadas = (numaprobadas/numasignaturas)*100
  return(porasigaprobadas)
}

porc_asignaturas_aprobadas_por_periodo_cedula =
alumnos6 %>%
  group_by(PERIODO, CEDULA) %>%
  summarise(PORCREDITOSAPROBADOS = func_porasigaprobadas(ESTADONOTA))

save(notas_medias_por_periodo_cedula,
      notas_medias_creditos_por_periodo_cedula,
      porc_creditos_aprobados_por_periodo_cedula,
      porc_asignaturas_aprobadas_por_periodo_cedula,

```

```
file="calculos_periodo_cedula.RData")

## Añadimos a alumnos6 las columnas: nota media, nota media ponderada, porcentaje créditos aprobados
y
## porcentaje asignaturas aprobadas, por periodo y cédula

# notas_medias_por_periodo_cedula
# notas_medias_creditos_por_periodo_cedula
# porc_creditos_aprobados_por_periodo_cedula
# porc_asignaturas_aprobadas_por_periodo_cedula

names(notas_medias_por_periodo_cedula)
names(notas_medias_creditos_por_periodo_cedula)[3] = "NOTAMEDIAPONDERADA"
names(porc_creditos_aprobados_por_periodo_cedula)
names(porc_asignaturas_aprobadas_por_periodo_cedula)[3] = "PORASIGNATURASAPROBADAS"

tempa1 = inner_join(notas_medias_por_periodo_cedula, notas_medias_creditos_por_periodo_cedula)
tempa1[1:20, ]
tempa2 = inner_join(tempa1, porc_creditos_aprobados_por_periodo_cedula)
tempa2[1:20, ]
tempa3 = inner_join(tempa2, porc_asignaturas_aprobadas_por_periodo_cedula)
tempa3[1:20, ]

alumnos7 = inner_join(alumnos6, tempa3)
names(alumnos7)[46] = "PORCOMPONENTEAPROBADOS"

alumnosbecasdocentes_completo = data.frame(alumnos7)
class(alumnosbecasdocentes_completo)

save(alumnosbecasdocentes_completo, file="alumnosbecasdocentes_completo.RData")
```

Anexo 10. BECAS

```
## manipulación de Becas

require("openxlsx")
xlsxFile = "Becas finales 2011-2012//Becas 2011 final.xlsx"
dbecas2011 <- read.xlsx(xlsxFile = xlsxFile, sheet = 1)
xlsxFile = "Becas finales 2011-2012//Becas 2012 final.xlsx"
dbecas2012 <- read.xlsx(xlsxFile = xlsxFile, sheet = 1)

## seleccionamos columnas que nos interesan
names(dbecas2011)

dbecas2011final = dbecas2011[,c(9,4,5,6,7,8,3)]
names(dbecas2011final)

names(dbecas2012)

dbecas2012final = dbecas2012[,c(9,4,5,6,7,8,3)]
names(dbecas2012final)

save(dbecas2011final,dbecas2012final,file="Becas20112012.RData")
```

Anexo 11. UNIÓN DE ARCHIVOS ALUMNOS - BECAS

```

load("alumnos_20112012final_02072014.RData")
names(alumnos_20112012)
table(alumnos_20112012$ANO)

## sobre 2011
alum2011 = alumnos_20112012[alumnos_20112012$ANO==2011,]

union2011t = merge(alum2011,dbecas2011final,by=c("CEDULA","MODALIDAD"),all.x=TRUE)
#rm(union2011t)

#pruebas 2011
alum2011[1:10,1]

(cedula = alum2011[20,1])
(cedula="1105601304")
(cedula="0104649355")
(cedula="0102078771")
dbecas2011final[dbecas2011final$CEDULA==cedula,]

union2011t[union2011t$CEDULA==cedula,c(1,2,26:30)]

dbecas2011final$CEDULA[dbecas2011final$BECA_SOCIO_ECONOMICA=='SI']

alum2011[11121,]
alum2011[alum2011$CEDULA==cedula,]

temp = table(dbecas2011final$CEDULA)
which(temp > 1)

# sobre 2012

alum2012 = alumnos_20112012[alumnos_20112012$ANO==2012,]

union2012t = merge(alum2012,dbecas2012final,by=c("CEDULA","MODALIDAD"),all.x=TRUE)

## pegamos los dos años

alumnos_20112012conbecas = rbind(union2011t,union2012t)
head(alumnos_20112012conbecas)

## guardamos en fichero R

save(alumnos_20112012conbecas,file='alumnos_20112012finalconbecas_03072014.RData')
##load("alumnos_20112012finalconbecas_03072014.RData")

```

Anexo 12. SELECCIÓN VARIABLES PROFESORES

```

## Se añade información con profesores

names(dpro2011)

table(dpro2011$CEDULA)

RESUMEN2 = aggregate( SEXO ~ CEDULA,dpro2011,length)

TEMP2 = merge(dpro2011,RESUMEN2,by=c("CEDULA"),all.x=TRUE)
TEMP3 = TEMP2[(TEMP2$SEXO.y==1) | ((TEMP2$SEXO.y==2) & (TEMP2$TIPO_PERSONAL_IES=='PROFESOR') ),]
table(TEMP3$CEDULA)
names(TEMP3)
dprofasig2011t = TEMP3[,c(1,6,7,8,9,13,15,17,19,21,23,24,25)]
names(dprofasig2011t)

dpro2011[dpro2011$CEDULA=='1718130972',]

dpro2011[dpro2011$CEDULA=='1103777759',]

names(alumnos_20112012conbecas_2011)
alumnos_20112012conbecas_2011[alumnos_20112012conbecas_2011$ASG_CODIGO=='1010004',c("PERIODO","CEDULA")]

tablita = table(dprofasig2011final$CEDULA,dprofasig2011final$CODIGO)

tablita[375,3]
rownames(tablita)[375]
hola = tablita[,3]
tablita[2,350:370]>1
dim(tablita)

guarda = c()
for (i in 1:1500) {
  hola = tablita[,i]
  names(hola) = NULL
  inds = which(hola>5)
  if (length(inds)>0) {
    guarda = c(guarda,i)
  }
}

length(guarda)
i = guarda[1]
tablita[inds,i]
inds
rownames(tablita)[inds]
colnames(tablita)[i]
margin.table(tablita,2)

which(colnames(tablita)=='1010004')

which(tablita>1)
tablita[1089047]
names(dprofasig2011final)

a1 = dprofasig2011t[,c(1,6)]
a2 = dprofasig2011t[,c(1,7)]
a3 = dprofasig2011t[,c(1,8)]
a4 = dprofasig2011t[,c(1,9)]
a5 = dprofasig2011t[,c(1,10)]
names(a1) = 1:2
names(a2) = 1:2

```

```

names(a3) = 1:2
names(a4) = 1:2
names(a5) = 1:2
atodos = rbind(a1,a2,a3,a4,a5)
names(atodos) = c('CEDULA','CODIGO')

atodost = atodos[atodos$CODIGO!='NO APLICA',]
rownames(atodost) = NULL

which(table(atodost$CODIGO)>20)
temp = table(atodost$CODIGO)
which(temp>20)
temp[711]
temp[1436]
names(atodost)
names(dprofasig2011t)
datosclaveprof2011 = dprofasig2011t[,c(1,2,3,4,5,11,12,13)]
dprofasig2011final = merge(atodost,datosclaveprof2011,by=c("CEDULA"),all.x=TRUE)
names(datosclaveprof2011)
names(dprofasig2011final)
##temp = table(datosclaveprof2011$CEDULA)

##length(temp>1)

## 2012

RESUMEN2 = aggregate( SEXO ~ CEDULA,dpro2012,length)

TEMP2 = merge(dpro2012,RESUMEN2,by=c("CEDULA"),all.x=TRUE)
TEMP3 = TEMP2[(TEMP2$SEXO.y==1) | ((TEMP2$SEXO.y==2) & (TEMP2$TIPO_PERSONAL_IES=='PROFESOR') ),]
dprofasig2012t = TEMP3[,c(1,6,7,8,9,13,15,17,19,21,23,24,25)]

a1 = dprofasig2012t[,c(1,6)]
a2 = dprofasig2012t[,c(1,7)]
a3 = dprofasig2012t[,c(1,8)]
a4 = dprofasig2012t[,c(1,9)]
a5 = dprofasig2012t[,c(1,10)]
names(a1) = 1:2
names(a2) = 1:2
names(a3) = 1:2
names(a4) = 1:2
names(a5) = 1:2
atodos = rbind(a1,a2,a3,a4,a5)
names(atodos) = c('CEDULA','CODIGO')

atodost = atodos[atodos$CODIGO!='NO APLICA',]
rownames(atodost) = NULL

datosclaveprof2012 = dprofasig2012t[,c(1,2,3,4,5,11,12,13)]
dprofasig2012final = merge(atodost,datosclaveprof2012,by=c("CEDULA"),all.x=TRUE)
names(datosclaveprof2012)
names(dprofasig2012final)

save(dprofasig2011final,dprofasig2012final,file="datosprofesores_final_anexo10.RData")

```

Anexo 13. CREACIÓN VARIABLES DOCENTES

```

load("datosprofesores_final_anexo10.RData")

require(openxlsx)
FECHAT1 = convertToDate(as.numeric(dprofasig2011final$FECHA_INGRESO_IES))
FECHAT2 = convertToDate(as.numeric(dprofasig2011final$FECHA_NACIMIENTO))

ANOANTIGUEDAD = 2012 - as.integer(strftime(FECHAT1,format="%Y"))
EDAD = 2012 - as.integer(strftime(FECHAT2,format="%Y"))

CATEGORIA = factor(dprofasig2011final$CATEGORIA_PERSONAL_ACADEMICO,ordered = T)
NIVELFORMACION = factor(dprofasig2011final$NIVEL_FORMACION,levels=c('NINGUNO','DIPLOMA SUPERIOR','E
SPECIALISTA','ESPECIALISTA MÉDICO','MAGISTER','DOCTOR Ph.D'),ordered = T)

TIEMPODEDICACION = factor(dprofasig2011final$TIEMPO_DEDICACION,levels=c('TIEMPO PARCIAL','SEMI EXCLU
SIVA O MEDIO TIEMPO','EXCLUSIVA O TIEMPO COMPLETO'),ordered = T)

dprofasig2011final = cbind(dprofasig2011final,ANOANTIGUEDAD,EDAD,CATEGORIA,NIVELFORMACION, TIEMPODEDI
CACION)

RESUMEN_NUMEROPROFESORES = aggregate( ANOANTIGUEDAD ~ CODIGO,dprofasig2011final,length)

RESUMEN_ANTIGUEDAD = aggregate( ANOANTIGUEDAD ~ CODIGO,dprofasig2011final,mean)
RESUMEN_EDAD = aggregate( EDAD ~ CODIGO,dprofasig2011final,mean)
RESUMEN_RETRIBU = aggregate( REMUNERACIONFINAL ~ CODIGO,dprofasig2011final,mean)

RESUMEN_CATEGORIA= aggregate( CATEGORIA ~ CODIGO,dprofasig2011final,max)
RESUMEN_FORMACION= aggregate( NIVELFORMACION ~ CODIGO,dprofasig2011final,max)
RESUMEN_DEDICACION= aggregate( TIEMPO_DEDICACION ~ CODIGO,dprofasig2011final,max)

as.integer(dprofasig2011final$SEXO.x[1:10])
temp2 = rep(NA,length(dprofasig2011final$SEXO.x))
for (i in 1:length(dprofasig2011final$SEXO.x)) {
  if (dprofasig2011final$SEXO.x[i]=="HOMBRE") {
    temp2[i] = 0
  } else {
    temp2[i] = 1
  }
}

table(temp2)
table(dprofasig2011final$SEXO.x)
dprofasig2011final = cbind(dprofasig2011final,TEMP2 = temp2)

RESUMEN_PORCMUJERES= aggregate( TEMP2 ~ CODIGO,dprofasig2011final,sum)
RESUMEN_PORCMUJERES2 = (RESUMEN_PORCMUJERES$TEMP2 / RESUMEN_NUMEROPROFESORES$ANOANTIGUEDAD)*100
str(RESUMEN_PORCMUJERES2)
table(RESUMEN_PORCMUJERES2)

names(dprofasig2011final)

dprofcodigoasig2011final = cbind(RESUMEN_ANTIGUEDAD,RESUMEN_EDAD$EDAD,RESUMEN_RETRIBU$REMUNERACIONFI
NAL,RESUMEN_CATEGORIA$CATEGORIA,
                                RESUMEN_PORCMUJERES2,RESUMEN_FORMACION$NIVELFORMACION,RESUMEN_DEDIC
ACION$TIEMPO_DEDICACION,
                                RESUMEN_NUMEROPROFESORES$ANOANTIGUEDAD)

names(dprofcodigoasig2011final)

```

```

which(dprofasig2011final$CODIGO=='1010014')
dprofasig2011final[3045,]
which(dpro2011$CEDULA=="1104602865")
dpro2011[c(410,816),]

names(dprofcodigoasig2011final) = c("ASG_CODIGO", "EXPERIENCIAMEDIA", "EDADMEDIA", "RETRIBUCIONMEDIA", "
CATEGORIAMAXIMA", "PORCENTAJEMUJERES", "FORMACIONMAXIMA", "DEDICACIONMAXIMA", "NUMEROPROFESORESOPRASG")

levels(dprofcodigoasig2011final$FORMACIONMAXIMA)[1] = 'TERCER NIVEL'

## 2012

datosclaveprof2012 = dprofasig2012t[,c(1,2,3,4,5,11,12,13)]
dprofasig2012final = merge(atodost,datosclaveprof2012,by=c("CEDULA"),all.x=TRUE)
names(datosclaveprof2012)
names(dprofasig2012final)

require(openxlsx)
FECHAT1 = convertToDate(as.numeric(dprofasig2012final$FECHA_INGRESO_IES))
FECHAT2 = convertToDate(as.numeric(dprofasig2012final$FECHA_NACIMIENTO))

ANOANTIGUEDAD = 2012 - as.integer(strftime(FECHAT1,format="%Y"))
EDAD = 2012 - as.integer(strftime(FECHAT2,format="%Y"))

CATEGORIA = factor(dprofasig2012final$CATEGORIA_PERSONAL_ACADEMICO,ordered = T)
NIVELFORMACION = factor(dprofasig2012final$NIVEL_.FORMACION,levels=c('NINGUNO','DIPLOMA SUPERIOR','E
SPECIALISTA','ESPECIALISTA MÉDICO','MAGISTER','DOCTOR Ph.D'),ordered = T)

TIEMPODEDICACION = factor(dprofasig2012final$TIEMPO_DEDICACION,levels=c('TIEMPO PARCIAL','SEMI EXCLU
SIVA O MEDIO TIEMPO','EXCLUSIVA O TIEMPO COMPLETO'),ordered = T)

dprofasig2012final = cbind(dprofasig2012final,ANOANTIGUEDAD,EDAD,CATEGORIA,NIVELFORMACION, TIEMPODEDI
CACION)

save(dprofasig2011final,dprofasig2012final,file="datosprofesores_medio_anexo11.RData")

RESUMEN_NUMEROPROFESORES = aggregate( ANOANTIGUEDAD ~ CODIGO,dprofasig2012final,length)

RESUMEN_ANTIGUEDAD = aggregate( ANOANTIGUEDAD ~ CODIGO,dprofasig2012final,mean)
RESUMEN_EDAD = aggregate( EDAD ~ CODIGO,dprofasig2012final,mean)
RESUMEN_RETRIBU = aggregate( REMUNERACIONFINAL ~ CODIGO,dprofasig2012final,mean)

RESUMEN_CATEGORIA= aggregate( CATEGORIA ~ CODIGO,dprofasig2012final,max)
RESUMEN_FORMACION= aggregate( NIVELFORMACION ~ CODIGO,dprofasig2012final,max)
RESUMEN_DEDICACION= aggregate( TIEMPO_DEDICACION ~ CODIGO,dprofasig2012final,max)

temp2 = rep(NA,length(dprofasig2012final$SEXO.x))
for (i in 1:length(dprofasig2012final$SEXO.x)) {
  if (dprofasig2012final$SEXO.x[i]=="HOMBRE") {
    temp2[i] = 0
  } else {
    temp2[i] = 1
  }
}

table(temp2)
table(dprofasig2012final$SEXO.x)
dprofasig2012final = cbind(dprofasig2012final,TEMP2 = temp2)

RESUMEN_PORCMUJERES= aggregate( TEMP2 ~ CODIGO,dprofasig2012final,sum)
RESUMEN_PORCMUJERES2 = (RESUMEN_PORCMUJERES$TEMP2 / RESUMEN_NUMEROPROFESORES$ANOANTIGUEDAD)*100
table(RESUMEN_PORCMUJERES2)

```



```
dprofcodigoasig2012final = cbind(RESUMEN_ANTIGUEDAD, RESUMEN_EDAD$EDAD, RESUMEN_RETRIBU$REMUNERACIONFI  
NAL, RESUMEN_CATEGORIA$CATEGORIA,  
                                RESUMEN_PORCMUJERES2, RESUMEN_FORMACION$NIVELFORMACION, RESUMEN_DEDIC  
ACION$TIEMPO_DEDICACION,  
                                RESUMEN_NUMEROPROFESORES$ANOANTIGUEDAD)  
  
names(dprofcodigoasig2012final)  
  
names(dprofcodigoasig2012final) = c("ASG_CODIGO", "EXPERIENCIAMEDIA", "EDADMEDIA", "RETRIBUCIONMEDIA", "  
CATEGORIAMAXIMA", "PORCENTAJEMUJERES", "FORMACIONMAXIMA", "DEDICACIONMAXIMA", "NUMEROPROFESORES  
PORASG")  
levels(dprofcodigoasig2012final$FORMACIONMAXIMA)[1] = 'TERCER NIVEL'
```

Anexo 14. UNIÓN BASE PROFESORES – ESTUDIANTES - BECAS

```
## 2011

alumnos_20112012conbecas_2011 = alumnos_20112012conbecas[alumnos_20112012conbecas$ANO=='2011',]

union2011temp = merge(alumnos_20112012conbecas_2011,dprofcodigoasig2011final,by=c("ASG_CODIGO"),all.
x=TRUE)

dprofasig2011final[dprofasig2011final$CODIGO=='1010004', ]

dpro2011[dpro2011$CEDULA=='1103346035',]
TEMP3[TEMP3$CEDULA=='1103346035',]

names(dprofasig2011final)
names(alumnos_20112012conbecas)

## 2012

alumnos_20112012conbecas_2012 = alumnos_20112012conbecas[alumnos_20112012conbecas$ANO=='2012',]

union2012temp = merge(alumnos_20112012conbecas_2012,dprofcodigoasig2012final,by=c("ASG_CODIGO"),all.
x=TRUE)

alumnos_20112012_becasyprofesores = rbind(union2011temp,union2012temp)

##pie(table(union2011temp$AREAFORMACION))
##barplot(prop.table(table(union2011temp$SEXO))*100,main="Diagrama de Barras sobre el Sexo")

names(alumnos_20112012_becasyprofesores)
head(alumnos_20112012_becasyprofesores)
table(alumnos_20112012_becasyprofesores$PERIODO)

alumnos_20112012_becasyprofesores[1:5,]

save(alumnos_20112012_becasyprofesores, file="estudiantesbecasdocentes_ultimo.RData")
```

