



Facultad de Ciencias Económicas y Empresariales

EL ESTUDIO DE LA ESTADÍSTICA EN GRADOS NO STEM

Gonzalo Villarías Vozmediano

5º E3 + Analytics

Director: José Luis Arroyo-Barrigüete.

Madrid, 2024-25

Resumen

Las asignaturas de estadística están adquiriendo una importancia creciente en un mundo marcado por el *Big Data* y la inteligencia artificial, incluso para aquellos profesionales de formación no técnica. Es por ello que la enseñanza de las distintas asignaturas de estadística en la universidad ha adquirido una relevancia en estos grados de la que probablemente carecía anteriormente. Por este motivo, se ha llevado a cabo un estudio con el propósito de analizar qué perfiles de alumnos existen en relación con el estudio de la estadística en grados no-STEM, y cuáles son los principales factores que influyen en su rendimiento. Con tal propósito, se ha seleccionado una muestra de 151 alumnos de grados en Analytics de ICADE, a quienes se ha hecho una encuesta. A los resultados de la misma se les ha aplicado un algoritmo de *clustering* jerárquico para identificar los distintos perfiles existentes, así como una red neuronal, que posteriormente ha sido objeto de un análisis *post-hoc* para estudiar los principales *drivers* del rendimiento de los alumnos. De este estudio se ha concluido que existen tres perfiles de alumnos a la hora de enfocar estas asignaturas. En cuanto a las variables que impulsan el rendimiento en las asignaturas de estadística, no se ha logrado determinar con carácter estadísticamente significativo cuál es el impacto de elementos como el género, el bachillerato cursado, la carrera universitaria, el tiempo de estudio o el soporte de estudio utilizado, habiéndose demostrado únicamente que una menor dificultad percibida se asocia a un mayor rendimiento relativo declarado.

Palabras Clave

Grados no-STEM, Estadística, Género, Bachillerato, Método de Estudio

Abstract

Statistics courses are gaining increasing importance in a world shaped by Big Data and artificial intelligence, even for professionals with non-technical backgrounds. For this reason, the teaching of various statistics courses at the university level has become more relevant in these degree programs than it likely was in the past. In this context, a study was conducted to analyse the different student profiles related to the study of statistics in non-STEM degrees and to identify the main factors influencing their performance. To this end, a sample of 151 students from Analytics programs at ICADE was selected, and a survey was administered. The survey results were analysed using a hierarchical clustering algorithm to identify the different existing profiles, as well as a neural network, which was subsequently subjected to a post-hoc analysis to study the main drivers of student performance. From this study, it has been concluded that there are three student profiles when it comes to approaching these subjects. Regarding the variables that influence performance in statistics courses, it was not possible to determine the statistically significant impact of factors such as gender, the type of high school education completed, the major studied in university, study time or the use of electronic devices in comparison to pen-and-paper. The only factor shown to have a significant association was that lower perceived difficulty is linked to higher self-reported relative performance.

Key Words

Non-STEM majors, Statistics, Gender, High School, Study method

Graphical abstract

Pregunta de investigación	
¿Qué variables caracterizan a los estudiantes de asignaturas de estadística en grados no-STEM en relación con su rendimiento académico?	
Revisión de la literatura	Hipótesis de investigación
8 variables identificadas: - Género - Especialidad - Grado - Método de Estudio - Tiempo de Estudio - Uso de Papel - Dificultad Percibida - Dificultad Percibida de la Programación	H1: los alumnos se distribuyen en tres grupos, en cuanto a su forma de enfocar la asignatura de estadística. H2: las mujeres tienen un rendimiento promedio en la asignatura de estadística mejor que el de los varones. H3: existe una mayor variabilidad de los hombres en el rendimiento en la asignatura de estadística, de tal forma que éstos se encuentran sobrerrepresentados entre aquellos alumnos con mejor y peor rendimiento. H4: los alumnos provenientes de un bachillerato científico tienen un mejor rendimiento en las asignaturas de estadística. H5: el rendimiento en las asignaturas de estadística es diferente en las distintas carreras de Analytics ofertadas en ICADE. H6: el enfoque de estudio basado en la práctica de ejercicios está asociado a mejores resultados académicos. H7: el empleo de soportes digitales en la preparación de las asignaturas de estadística se asocia con mejores resultados en las mismas.
Estimación del modelo	Resultados del modelo
- Encuesta a una muestra de 151 alumnos de Analytics de ICADE. - Modelo de clustering jerárquico y red neuronal con análisis de sensibilidad <i>post-hoc</i>.	H1: se confirma. H2: permanece incierta. H3: no se confirma. H4: permanece incierta. H5: permanece incierta. H6: permanece incierta. H7: permanece incierta.

Índice de Contenido

Introducción	8
Motivación del Trabajo	8
Justificación.....	9
Objetivo. Pregunta de Investigación	10
Metodología	10
Revisión de la Literatura	11
Principales variables explicativas	12
Género.....	12
Especialidad	16
Grado.....	18
Método de Estudio	18
Tiempo de Estudio	20
Uso de Papel	21
Dificultad Percibida	22
Dificultad Percibida de la Programación.....	22
Rendimiento Relativo Declarado.....	24
Hipótesis	24
Material y Métodos.....	25
La Muestra	25
Preprocesado de la Muestra.....	26
El Modelo de Clustering	27
Variables del Modelo	27

Definición del Modelo.....	37
El Modelo de Red Neuronal con análisis post-hoc.....	37
Variables del Modelo	37
Definición del Modelo.....	38
Resultados.....	39
Clustering jerárquico.....	39
Red Neuronal con análisis post-hoc	46
Discusión	49
Género.....	50
Especialidad	51
Grado.....	51
Método de Estudio	51
Tiempo de Estudio	52
Uso de Papel	53
Dificultad Percibida	53
Dificultad Percibida de la Programación.....	53
Conclusiones	54
Declaración respecto al uso de Chat GPT u otras herramientas de IAG	56
Referencias.....	57
Anexo I: Encuesta.....	62

Índice de Tablas

Tabla 1. Cuadro-resumen de las variables	27
Tabla 2. Dos primeras componentes principales del dataset	40
Tabla 3. Promedio de las variables en cada clúster	41
Tabla 4. Distribución de los clústers por Género.....	43
Tabla 5. Distribución de los clústers por Grado.....	43
Tabla 6. Distribución de los clústers por Especialidad	44
Tabla 7. Rendimiento Relativo Declarado por clúster	44
Tabla 8. Rendimiento Relativo Declarado por Género	45
Tabla 9. Rendimiento Relativo Declarado por Grado	45
Tabla 10. Rendimiento Relativo Declarado por Especialidad.....	46
Tabla 11. Coeficientes asociados a las variables en la red neuronal.	46
Tabla 12. Coeficientes asociados a las variables en la regresión lineal	48

Índice de Figuras

Figura 1. Distribución de la muestra por Género	28
Figura 2. Distribución de la muestra por Especialidad	29
Figura 3. Distribución de la muestra por Grado	30
Figura 4. Distribución de la muestra según el método de estudio.....	31
Figura 5. Distribución de la muestra según el tiempo de estudio.....	32
Figura 6. Distribución de la muestra según el uso de papel.....	33
Figura 7. Distribución de la muestra según la dificultad percibida.....	34
Figura 8. Distribución de la muestra según la dificultad percibida de la programación	35
Figura 9. Distribución de la muestra según el Rendimiento Relativo Declarado	36
Figura 10. Dendrograma con distancia euclídea y enlace de Ward	39
Figura 11. Clústers.....	40
Figura 12. Biplot del dataset.....	41
Figura 13. Visualización de las variables explicativas del Rendimiento Relativo Declarado	47
Figura 14. Pesos de las variables explicativas del Rendimiento Relativo Declarado	47
Figura 15. Distribución de las variables explicativas del Rendimiento Relativo Declarado .	47

Introducción

Motivación del Trabajo

La estadística se ha consolidado en los últimos tiempos como una herramienta fundamental en múltiples disciplinas, trascendiendo el ámbito tradicional de las carreras de ciencias (lo que en el ámbito anglosajón se ha conocido como “Grados STEM”). Así, en grados no-STEM, como ADE, economía, sociología o psicología, el manejo de herramientas estadísticas resulta cada vez más importante para la toma de decisiones informadas, el análisis de datos y la comprensión de fenómenos económicos y sociales (Sharma y Srivastav, 2021). A pesar de ello, la enseñanza de la estadística en estos grados enfrenta desafíos específicos, relacionados con el perfil de los estudiantes y los factores que influyen en su rendimiento.

Es por ello por lo que uno de los principales intereses de este trabajo es comprender qué perfiles de estudiantes cursan estadística en grados no-STEM, ya que ello resulta fundamental para que las distintas metodologías docentes se orienten a maximizar el rendimiento de los distintos perfiles de alumnos.

En esta misma línea, resulta crucial identificar cuáles son los principales factores que explican el rendimiento en las asignaturas de estadística dentro de estos grados. Aspectos como el género, el método de estudio, la modalidad de bachillerato cursada previamente, el tiempo dedicado al estudio o incluso el soporte en el que éste se realiza pueden resultar determinantes en relación con el rendimiento del alumnado. Entender estos elementos permitirá no solo mejorar la enseñanza de la estadística en grados no-STEM, sino también fomentar una mayor competencia en el análisis de datos, una habilidad cada vez más valorada en el mundo profesional actual.

Este estudio busca aportar conocimiento sobre estos aspectos, con el objetivo de contribuir a la mejora de la formación estadística en ámbitos no-STEM y, en última instancia,

optimizar la capacitación de estos futuros profesionales en el manejo de las distintas herramientas cuantitativas a su disposición.

Este TFG forma parte de un proyecto de investigación que consta de 4 TFGs diferentes, cada uno evaluando un tipo de materias en grados no-STEM. Se engloba en el marco del proyecto de innovación docente “*Non-Binary Approaches to Higher Education: Quantitative Analysis and Fact-Based Research*” (Ref. 221, U. Sevilla) y la Cátedra Santalucía de Analytics for Educación de la Universidad Pontificia Comillas.

Justificación

La estadística se ha convertido en una herramienta esencial para los profesionales de múltiples disciplinas, incluidas aquellas que no pertenecen al ámbito STEM. En áreas como el mundo de los negocios, la economía o la psicología, el análisis de datos es clave para la toma de decisiones estratégicas, la evaluación de tendencias y la validación de hipótesis. Además de esto, en un contexto profesional cada vez más basado en la información, contar con competencias estadísticas no es solo una ventaja competitiva, sino una necesidad para interpretar correctamente la realidad y fundamentar decisiones en evidencia cuantitativa.

No obstante, y a pesar de su relevancia, la estadística es una asignatura que genera tradicionalmente altos niveles de ansiedad entre los estudiantes. Basta, a este respecto, con constatar la extensísima doctrina existente que ha abordado la cuestión de la “ansiedad estadística” (*statistics anxiety*). Como han puesto de manifiesto todos estos estudios, esta ansiedad puede afectar a la motivación y el rendimiento académico de los alumnos, limitando el aprovechamiento de una disciplina que será fundamental en el futuro profesional de éstos (Hedges, 2017).

En este sentido, resulta imprescindible comprender qué perfiles de estudiantes hay entre los alumnos de estadística en grados no-STEM, así como cuáles son los principales factores que inciden en su rendimiento. Identificar estos elementos permitirá desarrollar

metodologías de enseñanza más efectivas, que promuevan un verdadero aprovechamiento de estos estudiantes en las asignaturas de estadística.

Así, el objeto de este trabajo será establecer qué perfiles de estudiantes resulta posible identificar entre el alumnado de las asignaturas de estadística en grados no-STEM, y valorar cuáles son las principales variables determinantes de su rendimiento.

Objetivo. Pregunta de Investigación

Habiendo contextualizado debidamente la relevancia y motivación de este estudio, se plantea el siguiente interrogante, que constituirá la guía a lo largo de esta investigación:

¿Qué variables caracterizan a los estudiantes de asignaturas de estadística en grados no-STEM en relación con su rendimiento académico?

Con ánimo de responder a esta pregunta con el rigor exigido por un estudio de esta naturaleza, se comenzará haciendo un análisis de la literatura, para determinar si (i) ésta ha apreciado la existencia de perfiles diferentes de alumnos entre quienes cursan asignaturas de estadística en grados no-STEM, así como (ii) cuáles son las principales variables que caracterizan a estos alumnos, y que resultan determinantes para su rendimiento académico.

A continuación, se implementará un modelo de *clustering* para buscar estos perfiles dentro de una muestra de alumnos real, y se entrenará también una red neuronal (que será interpretada en un análisis *post-hoc*), con el propósito de descubrir qué variables influyen decisivamente en el rendimiento académico de los alumnos de la muestra obtenida.

Metodología

Con ánimo de responder adecuadamente a la pregunta de investigación formulada, tras la oportuna revisión de la literatura, se ha implementado un modelo basado en datos reales de encuestas. Para ello, se han llevado a cabo a cabo los siguientes pasos:

- Selección de las variables relevantes: en base a la revisión de la literatura llevada a cabo, se han seleccionado aquellas variables que, de conformidad con la doctrina mayoritaria, resultan más aptas para responder a las cuestiones planteadas.
- Elaboración de la encuesta: con el propósito de obtener una base de datos reales de alumnos con la que trabajar, se elaboró una encuesta entre alumnos de todos los grados en *Analytics* de ICADE. El cuestionario puede consultarse en el **Anexo I**. Se trata de una encuesta ómnibus, compartida por 4 TFGs como ya se ha mencionado anteriormente, de modo que algunas preguntas (como las variables sociodemográficas) son comunes a todos ellos, y otras son específicas de cada TFG.
- Limpieza de los datos: en base a los datos obtenidos en la encuesta, se llevó a cabo una primera interpretación, un preprocesado de los mismos y una imputación de valores faltantes, conforme a las especificaciones técnicas que se detallan más adelante.
- Construcción del modelo de clustering: con todo lo anterior, se desarrolló, en primer lugar, un modelo de clustering jerárquico, con el propósito de obtener los distintos perfiles de alumnos existentes a la hora de afrontar las asignaturas de estadística.
- Construcción de la red neuronal: igualmente, se entrenó una red neuronal que posteriormente fue interpretada con un análisis *post-hoc* basado en el cálculo de sensibilidades, con el propósito de conocer, de entre las estudiadas, qué variables resultaban más relevantes en relación con el rendimiento académico de los alumnos.
- Extracción de conclusiones: finalmente, se detallan las conclusiones de los dos modelos construidos, contrastándolas con las conclusiones de la revisión de la literatura llevada a cabo.

Revisión de la Literatura

Como se ha señalado, el objeto de este estudio es analizar los distintos perfiles de alumnos existentes a la hora de afrontar las distintas asignaturas de estadística de los planes de estudios en grados no-STEM. Como también se ha indicado, será especialmente

relevante, a este respecto, estudiar cuáles son los principales factores que impulsan el rendimiento de los alumnos.

No obstante, antes de entrar a analizar la posición de la doctrina en relación con cada una de las posibles variables explicativas, hemos de ver si existe literatura acerca de la mera existencia de perfiles de alumnos diferentes a la hora de afrontar las asignaturas de estadística.

Con carácter general, la doctrina parece entender que existen tres grupos de alumnos a la hora de afrontar la asignatura de estadística (si bien, los distintos estudios difieren en cuanto a su tamaño relativo).

A este respecto, un primer estudio llevado a cabo en Indonesia entre alumnos de carreras de ciencias sociales concluyó que, en un extremo, un 48% de los alumnos no son capaces de entender y aplicar los conceptos básicos de la estadística. En contraposición a estos, otro 4% de alumnos dominarían plenamente estos conceptos, siendo capaces de aplicarlos sin dificultades a la resolución de problemas (Jauhari et al., 2021).

Como decimos, esta estructura de tres clústers parece repetirse a lo largo de los experimentos que se han llevado a cabo, apareciendo claramente asociada tanto al rendimiento como a la motivación de los alumnos (Tosunoğlu, 2017).

Por tanto, siendo ésta la opinión de la doctrina mayoritaria, será algo que se habrá de tener en cuenta a la hora de desarrollar nuestro modelo de *clustering*.

Principales variables explicativas

Género

El rendimiento de los alumnos en asignaturas de estadística en función de su género ha sido un debate que ha suscitado una enorme polémica, y una consiguiente gran cantidad de estudios a nivel mundial. Estos estudios se han pronunciado en todos los sentidos posibles, siendo posible encontrar (a) aquellos que señalan que los hombres tienen un mejor desempeño que las mujeres, (b) aquellos que concluyen que el rendimiento en asignaturas

de estadística es independiente del género de los alumnos y (c) aquellos otros que atribuyen mejores resultados a las mujeres.

Entre aquellos que atribuyen un mejor desempeño a los hombres en las asignaturas de estadística, podemos destacar un primer estudio llevado a cabo entre alumnos de ADE en Estados Unidos, en el que se identificó al género como una variable significativa a la hora de predecir el rendimiento de los alumnos en la asignatura de estadística – mostrando, a este respecto, los hombres, un mejor desempeño que las mujeres (Mitra, 2023).

También en un contexto europeo, los profesores Förster y Maur (2015) descubrieron que las mujeres experimentaban una inseguridad mucho mayor a la hora de afrontar la estadística, y consecuentemente, su desempeño final terminaba siendo peor que el de sus compañeros varones.

En cuanto a aquellos otros estudios que entienden que el género es irrelevante a estos efectos, podemos destacar, en primer lugar, un estudio llevado a cabo en una universidad de Malasia, en el que se buscó analizar el impacto de diversos factores en el rendimiento de los alumnos en la asignatura de estadística. Se concluyó, a este respecto, que al contrario de lo que ocurría, por ejemplo, con la raza, el género carecía de impacto alguno en el rendimiento del alumnado (Mohamed et al., 2011).

En esta misma línea, algunos de estos autores constataron también que, con carácter general, los factores demográficos (como puede ser el género) tenían un impacto mucho más limitado en el rendimiento de los alumnos en comparación, por ejemplo, con los factores personales (Judi et al., 2011).

Otro estudio similar, llevado a cabo entre alumnos de ADE en Estados Unidos, concluyó recientemente, en términos muy similares, que no existían diferencias de género en el rendimiento en las asignaturas de estadística (Ngamsiriudom et al., 2022).

También en Estados Unidos, en un estudio llevado a cabo, en este caso, entre alumnos de rama sanitaria, se concluyó que, frente a los estudios previos que sí indicaban

un impacto de género en el rendimiento de los alumnos en la asignatura de estadística, no se apreciaban diferencias en las notas finales de la asignatura entre hombres y mujeres (Lindsay, 2021).

Finalmente, resulta posible incluso encontrar aquellos otros experimentos que indican que el rendimiento de las mujeres en asignaturas de estadística es superior al de los varones. Así, en un estudio realizado entre alumnos de ADE en Uzbekistán, se descubrió, por parte de los autores, que el rendimiento promedio de las mujeres en la asignatura de estadística resultaba incluso superior a aquel de los hombres. No obstante, estos mismos autores concluyeron que esto podía deberse a un sesgo de selección, dado que las mujeres apenas constituían el 30% del alumnado de estos grados (Akimov et al., 2024).

Quienes sí se pronuncian sin fisuras respecto de un mejor desempeño de las mujeres en las asignaturas de estadística son varios profesores, que llevaron a cabo un estudio muy similar a todos estos que se han analizado en una universidad de Malasia, concluyendo que, pese a lo que señalaba la mayor parte de la literatura, las mujeres mostraban un mejor rendimiento que los hombres en las distintas asignaturas de estadística (Norshahidi et al., 2021).

Con todo, se ha de tener en cuenta que todos estos estudios están valorando el rendimiento promedio de los estudiantes de cada uno de los géneros (lo cual no es equivalente a afirmar que ambos géneros sigan la misma distribución en cuanto a su rendimiento en las asignaturas de estadística). Y es que, en diversos campos de estudio (y también en el de la educación) existen investigaciones que han llegado a la conclusión de que la distribución del talento masculino es más dispersa, de tal modo que, con independencia de las medias, resulta más probable encontrarse con varones -en este caso- con un excelente o un pésimo rendimiento.

Los factores biológicos subyacentes que parecen explicar esta hipótesis de la “mayor variabilidad masculina” son, por un lado, el sistema de selección sexual de los varones, y por

otro, su composición cromosómica (XY) – factores que conducirían, en principio, a una mayor asunción de riesgos y una mayor variabilidad en su consumo basal de energía (Halsey et al., 2022).

Como se comentaba, esta hipótesis de la mayor variabilidad masculina cuenta con una amplia acogida en el campo de la educación. A título de ejemplo, en el campo de los grados no STEM objeto de este estudio, la doctrina señala que, aunque por lo general las mujeres tienden a obtener en promedio mejores calificaciones que sus compañeros varones, la mayor variabilidad masculina determina que haya una mayor proporción de varones entre aquellos alumnos que obtienen mejores y peores calificaciones, siendo esta mayor variabilidad más acusada en asignaturas “prácticas” como la estadística, que en asignaturas “teóricas” como la historia (O’Dea et al., 2018).

No obstante, hay también autores que señalan que, si bien esta hipótesis de la mayor variabilidad masculina rige plenamente en el ámbito de la educación, y probablemente determina la sobrerrepresentación masculina en los grados STEM, esta parece diluirse con el tiempo (Baye y Monseur, 2016).

Además de la puramente biológica, otra de las teorías que se han elaborado para explicar estas diferencias entre hombres y mujeres es aquella que señala que, con carácter general, los hombres tienden al exceso de confianza. Un exceso de confianza que les llevaría a prepararse peor para los exámenes y, finalmente, obtener peores calificaciones (Hossain y Tsigaris, 2015).

En el otro extremo de la doctrina, hay también quienes entienden que el mejor rendimiento de las mujeres se deriva, precisamente, del hecho de que tienden a esforzarse más (Mansor et al., 2020).

Así, se puede concluir que no existe en la doctrina un consenso acerca de la vinculación entre el género y el rendimiento de los alumnos en las asignaturas de estadística,

ni tampoco en relación con los factores que pudieran explicar las diferencias entre ambos sexos. Es por ello que los resultados de este trabajo revestirán especial interés en este punto.

Especialidad

Al igual que el género, otro factor que ha sido extensamente discutido en relación con el rendimiento de los alumnos en la asignatura de estadística ha sido su “Especialidad”, entendida esta como su bagaje cuantitativo-matemático previo, determinado mayoritariamente, en el caso de los alumnos objeto de este estudio, por la modalidad de bachillerato que cursaron.

A este respecto tenemos por un lado aquellos autores que entienden que haber cursado un bachillerato científico o tener un mayor acervo matemático determina un mejor rendimiento en la asignatura de estadística, ya sea por la vía directa (mayores habilidades cuantitativas) o indirecta (mayor confianza en las habilidades cuantitativas propias).

En esta línea, en un experimento realizado en España, varios profesores concluyeron que la experiencia previa en este tipo de cuestiones determinaba una menor ansiedad, y consiguientemente, un mejor rendimiento en la asignatura de estadística (Peiró-Signes et al., 2021). A la misma conclusión se llegó en otro estudio similar llevado a cabo en Alemania (Förster y Maur, 2015).

En el mismo sentido, un estudio realizado en Italia identificó también una correlación directa entre las habilidades matemáticas de los alumnos de un grado non-STEM y su desempeño en la asignatura de estadística (Fabbri et al., 2023).

Resultados en este sentido han sido obtenidos, además, fuera de un contexto europeo, apreciándose en una universidad de Malasia una fuerte correlación entre el bagaje matemático de los alumnos y su rendimiento en la asignatura de estadística (Kien-Kheng et al., 2016).

Igualmente en Malasia, otro estudio trató de comparar el rendimiento en la asignatura de estadística de alumnos con un perfil científico (*analytics*) frente al desempeño de alumnos

que carecían de este perfil (dentro de que todos ellos pertenecían un grado no-STEM). A este respecto, se concluyó que las diferencias identificadas resultaban significativas, pero únicamente en aquellas destrezas más “elevadas”, como la aplicación de lo aprendido a problemas reales y novedosos (Mustaffa y Zulkifli, 2019).

Y probablemente sea aquí donde reside la principal diferencia entre el bachillerato tecnológico y el bachillerato de ciencias sociales: en que el primero dota a los alumnos de una serie de habilidades que les permite distinguirse de sus compañeros cuando se enfrentan en la asignatura de estadística a problemas más complicados (Rahayu et al., 2022).

No obstante, también resulta posible identificar otros estudios que concluyan que no existe correlación entre la formación matemática adquirida durante el bachillerato y el desempeño de los alumnos en las asignaturas de estadística de la universidad (Arroyo-Barrigüete et al, 2023; Dupuis et al., 2012).

Otro estudio realizado en Malasia, que evaluaba los factores que explicaban el rendimiento en la asignatura de estadística, concluyó, en términos muy similares, que la modalidad de bachillerato cursada no era un elemento que permitiera explicar este desempeño (Mohamed et al., 2011).

La explicación a esto, según señalan otros estudios similares, es que existen diferencias importantes entre la estadística que se estudia en la asignatura de matemáticas del bachillerato, y la que finalmente se da en la carrera, de tal forma que los estudiantes que cursaron la modalidad científica, aun teniendo una mejor preparación matemática, carecerían de ventajas competitivas frente a sus compañeros de ciencias sociales (Bourne, 2014).

Con todo, resulta posible concluir que, respecto del impacto de los estudios de bachillerato en el rendimiento en la asignatura de estadística, la doctrina se encuentra, nuevamente, dividida – razón por la cual pueden resultar especialmente interesantes las conclusiones que arroje este estudio.

Grado

A la hora de estudiar el impacto en el rendimiento en la asignatura de estadística del grado cursado de los alumnos, se ha de tener en cuenta que la oferta de grados varía enormemente de universidad en universidad (no digamos ya de país en país). Es por ello por lo que la literatura que se ha examinado no se centrará tanto en grados concretos (cuestión imposible), sino más bien en si el rendimiento en la asignatura de estadística varía significativamente de carrera en carrera, sin especificar cuáles son éstas.

A este respecto, existen estudios que concluyen que la respuesta a este interrogante es afirmativa. Un estudio llevado a cabo en Estados Unidos concluyó que, efectivamente, había cambios significativos en el rendimiento en la asignatura de estadística, aun quedándonos en el marco de los grados no-STEM, entre carrera y carrera. Así, grados como Finanzas o Sistemas de Información aparecían asociados a una mayor motivación y un mejor rendimiento en la asignatura. En cambio, otros grados como Marketing o *Management* se correlacionaban con menor motivación y menor desempeño (Ngamsiriudom et al., 2022).

Igualmente, otros estudios apreciaron rendimientos distintos en función de la carrera, en este caso, en el ámbito biosanitario. No obstante, los autores de este estudio entendieron que estas diferencias se debían a que aquellas carreras más exigentes generaban un sesgo de selección del alumnado, que filtraba a aquellos con un mejor rendimiento, también en la asignatura de estadística (Lindsay, 2021).

Así, se concluye que, de acuerdo con la doctrina, la carrera estudiada sí puede tener un impacto en el desempeño en las asignaturas de estadística (aunque se habrá de tener en mente que gran parte de este efecto podrá deberse al sesgo de selección generado por las diferentes carreras). En todo caso, se habrá de estar a los resultados de nuestro experimento para poder apreciar si esos efectos se dan entre los grados en *Analytics* de ICADE.

Método de Estudio

Antes de entrar a desarrollar la postura de la doctrina en relación con el impacto de esta variable en el rendimiento del alumnado en asignaturas de estadística, se ha de clarificar

que con “Método de Estudio” se hace referencia al enfoque adoptado por los alumnos a la hora de afrontar el estudio de la asignatura, pudiendo consistir, en un extremo, en únicamente estudiar teoría, y en el otro, en practicar ejercicios, exclusivamente.

Respecto de esta cuestión, si bien no ha sido la más extensamente tratada por la doctrina, sí parece que de los estudios realizados se pueda extraer la conclusión de que, en promedio, los métodos de estudio con un enfoque más práctico aparecen asociados a un rendimiento más satisfactorio por parte del alumnado.

Así, en un experimento en el que estudiaron los factores que impulsaban el rendimiento en la asignatura de estadística de alumnos de ADE en Estados Unidos, Mitra y Le (2019) concluyeron que las metodologías prácticas aparecían asociadas a una mayor probabilidad de aprobar la asignatura de estadística.

Estas mismas conclusiones han sido alcanzadas en un contexto europeo. Así, en un estudio con alumnos de sociología de una universidad de Alemania, el profesor Grönemann (2024) concluyó que, dado que el aprendizaje de los alumnos es más fructífero si se realiza de forma activa, y dado que la interiorización de los conceptos es algo que lleva tiempo, el rendimiento se optimiza con un enfoque práctico, razón por la cual incluso propugna por dedicar un mayor tiempo durante la clase para la realización de ejercicios.

Esta idea de que la práctica de ejercicios constituye una forma “activa” de aprendizaje, mientras que el estudio teórico-memorístico es algo “pasivo” es una idea aparentemente consolidada en la doctrina. Así, en una revisión sistemática sobre técnicas de estudio, se concluyó que la mera lectura de los apuntes como metodología de trabajo aparecía asociada a un menor rendimiento académico, siendo necesario, de cara a aumentar dicho rendimiento, implementar una metodología que implicara una autoevaluación de su conocimiento por parte del alumno, como podría serlo la práctica de ejercicios (Xu et al., 2021).

Es precisamente por este motivo por el que muchos profesores han optado por adaptar sus métodos de enseñanza, de tal forma que estos favorezcan los enfoques

prácticos, en tanto que estos aparecerían, en principio, asociados a una mayor comprensión de la materia (Bologna, 2019). Se buscaría, así, que los alumnos fueran más allá de la mera memorización de conceptos, y fueran capaces de aplicarlos y, sobre todo, extraer conclusiones de los datos (Saputra et al., 2018).

Así, se puede sintetizar que la posición de la doctrina en relación con el impacto del método de estudio en el rendimiento de los alumnos es que, con carácter general, un enfoque más práctico a la hora de estudiar la asignatura de estadística vendría asociado a un mejor desempeño en la misma.

Tiempo de Estudio

En cuanto al impacto del tiempo dedicado a la asignatura de estadística en relación con el rendimiento en la misma, aunque *a priori* parecería que debiera existir una fuerte correlación positiva, la doctrina no se ha puesto de acuerdo a la hora de afirmar que un mayor tiempo de estudio implique necesariamente un mejor rendimiento en la asignatura.

Entre quienes sostienen que no existe correlación entre estas dos variables se encuentran los profesores Hammonds y Mariano (2015), quienes realizaron un estudio entre alumnos de psicología de Estados Unidos, en el que no hallaron correlación alguna entre el tiempo de estudio y los resultados de los exámenes de la asignatura de estadística.

En esta misma línea, una revisión sistemática sobre el impacto de las técnicas de estudio empleadas por los alumnos descubrió que el número de días de antelación con los que éstos empezaban a preparar un examen no influía, aparentemente, en los resultados que posteriormente obtendrían (Xu et al., 2021).

No obstante, también hay dentro de la doctrina quien ha resaltado la importancia del tiempo de cara a asentar los conceptos en asignaturas como estadística (Mitra y Le, 2019), determinando, en consecuencia, el tiempo de estudio, un mejor rendimiento en las asignaturas de estadística.

Es por ello por lo que, en este punto, se puede concluir que, en lo que respecta a la correlación entre el tiempo de estudio dedicado a la asignatura de estadística y los resultados obtenidos, no hay aparentemente ningún estudio que lleve a la doctrina a posicionarse en favor de una u otra postura. Es por ello por lo que el resultado de nuestro experimento revestirá especial interés.

Uso de Papel

Antes de entrar a desarrollar la postura de la doctrina en relación con el impacto de esta variable en el rendimiento del alumnado en asignaturas de estadística, se ha de clarificar que con “Uso del Papel” se hace referencia al empleo por parte de los alumnos de soportes manuscritos a la hora de preparar la asignatura (en contraposición al empleo de soportes digitales, tales como ordenadores o tabletas). Asimismo, se hace referencia al empleo de estos soportes tanto durante la clase como a la hora de estudiar la asignatura en casa y hacer las tareas que se propongan.

En relación con esta variable, los estudios tienden a señalar que la realización del trabajo empleando soportes digitales tiende a ser irrelevante en relación con el rendimiento de los alumnos en asignaturas de estadística (pudiendo incluso tener un impacto positivo, como consecuencia de ciertos efectos colaterales asociados a la metodología *online*).

Así, en un estudio realizado en Islandia, varios profesores concluyeron que el hecho de exigir a los alumnos que realicen sus tareas a través de plataformas digitales (en lugar de entregárselos en papel al profesor) no impacta en el rendimiento de los alumnos (salvo en el caso de aquellos alumnos de peor rendimiento, cuyo desempeño sí parece verse favorecido por el empleo de soportes digitales). La explicación que estos profesores darían a esta ligera ventaja de los sistemas digitales sobre los sistemas “analógicos” parecería estar, más bien, vinculada a factores externos asociados al soporte, pero distintos del mismo. Y es que, según explican, el empleo de soportes digitales permite una corrección “automática” de las tareas, que facilita a los alumnos tomar conciencia de su situación, obrar en consecuencia, y finalmente obtener mejores resultados. Esta hipótesis sería, además, congruente con el

hecho de que el empleo de sistemas digitales tenga un mayor impacto en aquellos alumnos con peor desempeño (Jonsdottir et al., 2017).

Resulta, en todo caso, sorprendente, que estos resultados que otorgan una ligera ventaja al empleo de soportes digitales, se reproduzcan cada vez que se realizan experimentos de esta naturaleza. Así, en otro experimento similar realizado en Estados Unidos, se concluyó que resultaba irrelevante en relación con los resultados finales, que los deberes se hagan a través de plataformas digitales, o se entreguen al profesor en papel, resultando, en este caso, una ligera ventaja para quienes lo hicieron en papel, si bien los autores concluyen que no resulta estadísticamente relevante (Williams, 2012).

Siendo estos los resultados de los estudios realizados al respecto, parece prudente concluir que, en opinión de la doctrina, el soporte empleado para el estudio tiene un impacto limitado en el rendimiento de los estudiantes (sin perjuicio de que una u otra metodología pudieran tener implicaciones colaterales de las que sí se pudieran derivar impactos significativos).

Dificultad Percibida

A la hora de comprender esta variable, se ha de entender que, tal y como se expone a la hora de desarrollar las características técnicas del estudio, éste se ha realizado en base a encuestas completadas con posterioridad a haber cursado las respectivas asignaturas de estadística. Es por ello por lo que, como resulta lógico, la dificultad percibida de la asignatura (evaluada con posterioridad a haberla cursado) tendrá una fuerte correlación inversa con el rendimiento en la misma.

Dificultad Percibida de la Programación

En relación con la correlación entre la dificultad que, según los alumnos, entraña la parte de programación en las asignaturas de estadística, y el rendimiento total de estos alumnos, la doctrina ha entendido, en términos generales, que una mayor facilidad para la parte de programación aparece asociada con un mejor rendimiento por parte de los alumnos.

Así, en un experimento llevado a cabo entre estudiantes de medicina en Irán, se descubrió que existía una muy fuerte correlación entre el desempeño total en la asignatura de estadística y el desempeño en la parte de programación. Una correlación que, además, no se debía simplemente a una mayor competencia global, ya que aquellos alumnos clasificados en el clúster correspondiente al mayor desempeño académico, mostraban simultáneamente un rendimiento en la parte de programación desproporcionadamente alto como para explicarse únicamente por su mayor rendimiento general en la asignatura (Saki et al., 2017).

En Europa, los resultados obtenidos en otros experimentos que se han realizado han sido muy similares. Así, se ha podido observar que, en términos generales, en grueso de los alumnos tienden a rehuir la parte de programación en las asignaturas de estadística (que perciben como más difícil), centrándose en el resto del examen en aras de optimizar su nota (Grönemann, 2024).

Resulta también interesante señalar cómo, según un estudio llevado a cabo en el Reino Unido, mientras que el nivel de matemáticas de los alumnos tiene un impacto importante (al menos al principio) en su rendimiento en asignaturas de estadística, éste bagaje matemático resulta irrelevante a la hora de evaluar el desempeño de los alumnos en la parte de programación (Bourne, 2014). Ello induce a pensar que este rendimiento en la parte de programación tiene un valor explicativo “propio”, al margen del perfil más o menos “científico” del alumnado.

Así, resulta posible concluir que, en opinión de la doctrina, la dificultad percibida, y consecuentemente el rendimiento en la parte de programación de la asignatura de estadística aparece fuertemente correlacionado con el rendimiento global en la asignatura, contando además esta variable con sustantividad propia frente a otras que pudieran influir en la misma y explicar, en última instancia, sus efectos.

Rendimiento Relativo Declarado

En relación con este factor, se ha de tener en cuenta que se trata de uno de los principales componentes que se busca evaluar a la hora de llevar a cabo la clasificación de los perfiles del alumnado en la asignatura de estadística. Es por ello por lo que, siendo esta la variable endógena de nuestro modelo, la explicación acerca de la misma será la expuesta a la hora de estudiar el resto de variables, a las que hemos atribuido la condición de exógenas.

En todo caso, cabe anticipar como una de las principales limitaciones de nuestro estudio el fuerte sesgo positivo que esta variable experimenta. Y es que, ante la imposibilidad de acceder a los expedientes de los alumnos, la evaluación de su rendimiento debe hacerse en base al rendimiento que estos declaran, en función de propia estimación de su situación en comparación con la de sus compañeros. Esto, naturalmente, entraña un importante grado de incertidumbre, y sesga positivamente la variable como consecuencia de que (i) los estudiantes ignoran el rendimiento exacto de sus compañeros, (ii) los estudiantes tienden a sobreestimar su propio rendimiento (*over-placement*) y (iii) en última instancia, los estudiantes mienten acerca de su rendimiento y tienden a inflarlo.

Estos problemas ya han sido reportados en ocasiones en estudios muy similares, en los que, pese a esta limitación, los resultados se han considerado, en cualquier caso, conclusivos (Berenson et al., 2015). Es por ello por lo que, salvo que los datos empleados nos lleven a conclusiones manifiestamente ilógicas, será posible que se continúe trabajando con esta variable.

Hipótesis

A partir de la revisión de la literatura llevada a cabo, resulta posible formular las siguientes hipótesis, que serán objeto de contraste en nuestro experimento:

- Hipótesis 1: los alumnos se distribuyen en tres grupos, en cuanto a su forma de enfocar la asignatura de estadística.

- Hipótesis 2: las mujeres tienen un rendimiento promedio en la asignatura de estadística mejor que el de los varones.
- Hipótesis 3: existe una mayor variabilidad de los hombres en el rendimiento en la asignatura de estadística, de tal forma que éstos se encuentran sobrerrepresentados entre aquellos alumnos con mejor y peor rendimiento.
- Hipótesis 4: los alumnos provenientes de un bachillerato científico tienen un mejor rendimiento en las asignaturas de estadística.
- Hipótesis 5: el rendimiento en las asignaturas de estadística es diferente en las distintas carreras de *Analytics* ofertadas en ICADE.
- Hipótesis 6: el enfoque de estudio basado en la práctica de ejercicios está asociado a mejores resultados académicos.
- Hipótesis 7: el empleo de soportes digitales en la preparación de las asignaturas de estadística se asocia con mejores resultados en las mismas.

Material y Métodos

La Muestra

Con el propósito de obtener una muestra de datos real en base a la cual llevar a cabo el estudio, se ha confeccionado un cuestionario cuyo contenido exacto puede consultarse en el **Anexo I**, en base al cual, se ha recopilado información relativa a las diferentes variables que han sido objeto de análisis en la revisión de la literatura.

El referido cuestionario ha sido respondido por alumnos de ICADE pertenecientes a los diferentes grados de *Analytics* ofertados por la universidad, a saber, E-2 Analytics (Doble Grado en ADE y Business Analytics), E-3 Analytics (Doble Grado en Derecho y Business Analytics) y E-6 Analytics (Doble Grado en Relaciones Internacionales y Business Analytics). Dentro de estos grados, y con el propósito de excluir a aquellos alumnos que todavía no han terminado de cursar todas las asignaturas de estadística, únicamente han respondido la encuesta aquellos alumnos que se encontraran en sus estudios a partir del segundo

cuatrimestre de tercero de carrera. Además, y con el propósito de restringir el estudio a ICADE, se ha excluido de la encuesta a aquellos alumnos extranjeros que se encontraban de intercambio en ICADE.

En cuanto a la metodología de compilación de datos, se ha optado, con el propósito de maximizar la calidad de las respuestas, por una recolección de datos íntegramente presencial y en papel.

Finalizada la encuesta, se han obtenido para el estudio un total de 151 observaciones.

Preprocesado de la Muestra

Recopilados los datos de la encuesta, se ha llevado a cabo un primer ejercicio de limpieza y preprocesado de los mismos.

Concretamente, se han convertido las variables categóricas indicadas para transformarlas en variables tipo factor, se han eliminado variables que carecían de contenido informativo o que contenían información errónea (como la variable EvAU, ya que algunos encuestados respondieron sobre 10 y otros sobre 14), y se ha llevado a cabo una imputación de valores faltantes.

Para esta interpolación de N/As, en relación con cada una de las variables con valores faltantes (Estadistica_Papel y Genero), se probaron diferentes modelos, para ver cuál ofrecía el mejor rendimiento en cuanto a la predicción de los valores faltantes.

En primer lugar, en relación con Estadistica_Papel, al ser esta una variable numérica, se probó un algoritmo de regresión lineal (OLS), un algoritmo de KNN, un algoritmo de Random Forest, y una aproximación estadística tomando como referencia la mediana de los alumnos de cada Grado. En este caso, el Random Forest demostró ser el mejor estimador, de tal forma que este fue el algoritmo empleado para la interpolación final.

En segundo lugar, en relación con Genero, al ser esta una variable categórica, se probó un algoritmo de regresión logística (*logit*), un algoritmo de KNN, un algoritmo de

Random Forest, y una aproximación estadística tomando como referencia la moda de los alumnos de cada Grado. Nuevamente, el Random Forest demostró ser el mejor estimador, de tal forma que este fue el algoritmo empleado para la interpolación final.

Realizado este preprocesado, el camino quedaba expedito para la implementación de los algoritmos referidos.

El Modelo de Clustering

Variables del Modelo

A continuación, se desgranará la distribución de cada una de las variables de interés de la encuesta, y se explorará su implementación en el modelo de clustering. Para la facilidad del lector, se muestra en la Tabla 1 un cuadro-resumen con las distintas variables que se emplearán.

Tabla 1. Cuadro-resumen de las variables

Nombre de la variable	Breve descripción
Género	Variable categórica dicotómica (Hombre/Mujer)
Especialidad	Variable categórica nominal (Ciencias/Ciencias Sociales/Bachillerato Internacional/Otras)
Grado	Variable categórica nominal (E-2 Analytics/ E-3 Analytics/ E-6 Analytics)
Método de Estudio ¹	Numérica discreta (1-6)
Tiempo de Estudio ²	Numérica discreta (0-10)

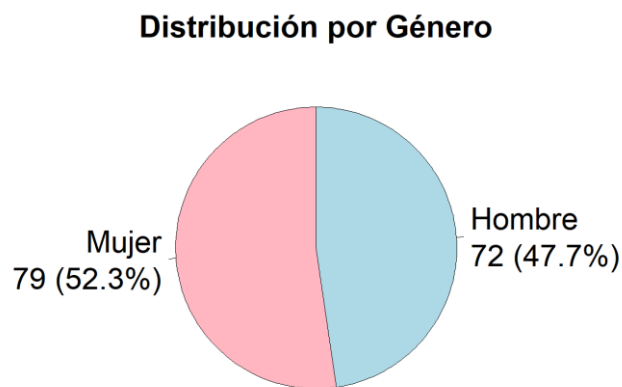
¹ PREGUNTA: En asignaturas de Estadística del doble grado (Estadística y Probabilidad, y Métodos Estadísticos) ¿aproximadamente qué porcentaje del tiempo de estudio dedicaste a estudiar teoría y que porcentaje a practicar problemas? Selecciona lo que corresponda.

² PREGUNTA: Comparado con otras asignaturas de tu doble grado (de todas, no solo las de Analytics) ¿Cuánto tiempo dedicaste a estudiar las asignaturas de Estadística (Estadística y Probabilidad, y Métodos Estadísticos)? Valóralo en una escala del 0 (mucho menos) al 10 (mucho más).

Uso de Papel ³	Numérica discreta (1-5)
Dificultad Percibida ⁴	Numérica discreta (0-10)
Dificultad Percibida en la Programación ⁵	Numérica discreta (0-10)
Rendimiento Relativo Declarado ⁶	Numérica discreta (0-10)

Género. En la Figura 1, puede observarse la distribución de la muestra en función del género.

Figura 1. Distribución de la muestra por Género



Como se ha señalado a la hora de llevar a cabo la revisión de la literatura, esta es una variable que, según muchos autores, tiene un impacto significativo en relación con el

³ PREGUNTA: En asignaturas de Estadística del doble grado (Estadística y Probabilidad, y Métodos Estadísticos), cuando estudiabas la asignatura ¿lo hacías usando dispositivos electrónicos o con papel y bolígrafo? Selecciona lo que corresponda.

⁴ PREGUNTA: ¿Cuál consideras que es el grado de dificultad de las asignaturas de Estadística (Estadística y Probabilidad, y Métodos Estadísticos)? Valóralo en una escala del 0 (extremadamente fácil) al 10 (extremadamente difícil).

⁵ PREGUNTA: Centrándonos solo en la parte de programación en R, ¿Cuál consideras que es el grado de dificultad de LA PARTE DE PROGRAMACIÓN de las asignaturas de Estadística (Estadística y Probabilidad, y Métodos Estadísticos)? Valóralo en una escala del 0 (extremadamente fácil) al 10 (extremadamente difícil).

⁶ PREGUNTA: ¿Cuál consideras que ha sido tu rendimiento (nota obtenida) en las asignaturas de Estadística (Estadística y Probabilidad, y Métodos Estadísticos)? Valóralo en una escala del 0 (muy por debajo de la media de mi clase) al 10 (muy por encima de la media de mi clase).

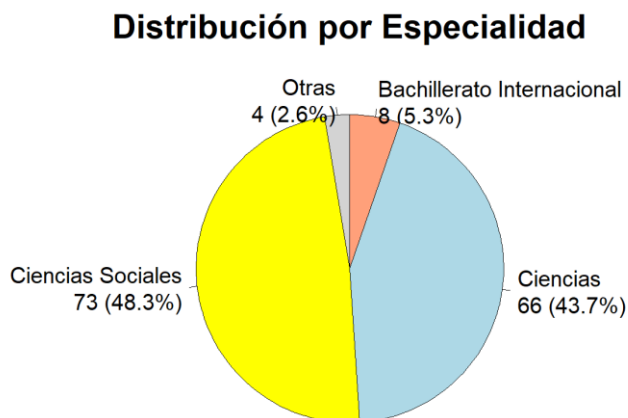
rendimiento de los alumnos en las asignaturas de estadística (y consecuentemente, es posible que también resulte relevante a la hora de segmentar a éstos en función de sus perfiles).

En relación con la muestra obtenida, puede verse que se encuentra bien equilibrada, lo que será positivo en aras de poder extraer conclusiones fiables de los modelos.

Asimismo, en relación con la implementación de la variable en el modelo, dado que lo que se quiere establecer en el clustering son perfiles de alumnos, se limitará la elaboración de los clústers a las variables numéricas, denotativas del perfil psicográfico-académico de los alumnos. Siendo el género, en cambio, una variable demográfica, la utilidad que se le dará será la de, una vez elaborados los clústers, ver qué perfil de género tienen éstos.

Especialidad. En la Figura 2, puede observarse la distribución de la muestra en función de la especialidad de bachillerato cursada por los alumnos.

Figura 2. *Distribución de la muestra por Especialidad*



Como se ha señalado en la revisión de la literatura, la modalidad de bachillerato cursada por los alumnos y el consiguiente acervo matemático parecen ser una variable relevante a la hora de perfilar su rendimiento en las asignaturas de estadística de la carrera. En este caso concreto, y a los efectos de nuestro análisis, parece razonable simplemente prescindir de aquellas modalidades de bachillerato distintas de las tradicionales modalidades de ciencias y ciencias sociales del bachillerato español (ya que éstas, además, permiten

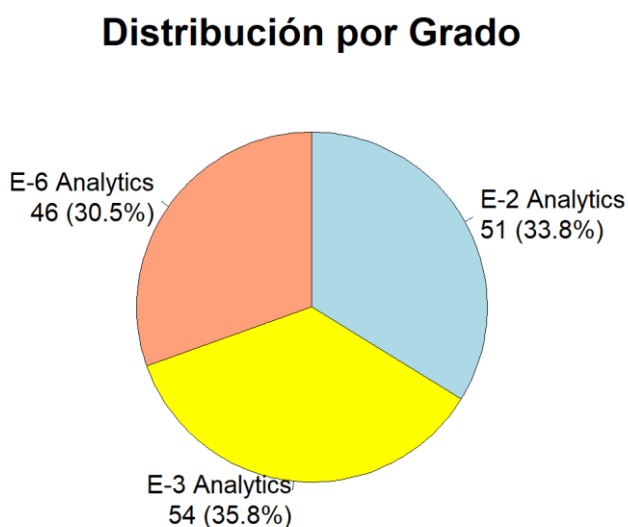
plantear un dualismo muy interesante a efectos de nuestro estudio, siendo además, las otras dos categorías cuantitativamente marginales y cualitativamente inciertas).

Nuevamente, al igual que ocurría con el género, la muestra obtenida resulta equilibrada en función de las distintas especialidades de bachillerato, lo cual, como se ha señalado, facilitará la extracción de conclusiones en este estudio.

Finalmente, señalar que, por ser esta una variable demográfica, al igual que se hizo con el género, se evitará introducirla directamente en el modelo a la hora de hacer los clústers, y simplemente se empleará a posteriori, para ver el perfil de los distintos clústers resultantes en función de las distintas modalidades de bachillerato estudiadas por los alumnos.

Grado. En la Figura 3, puede observarse la distribución de la muestra en función del grado estudiado por los alumnos.

Figura 3. *Distribución de la muestra por Grado*



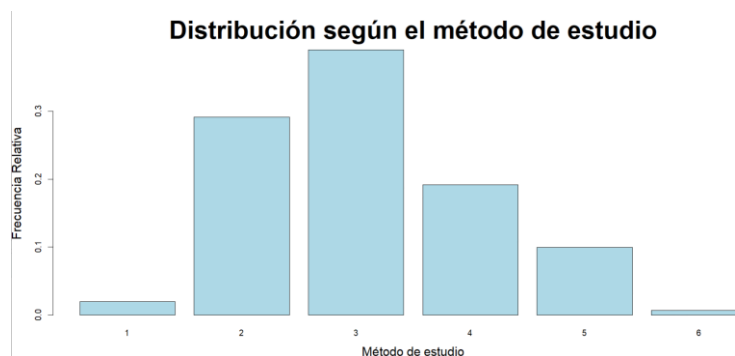
Como se señaló en la revisión de la literatura, el rendimiento de los alumnos en la asignatura de estadística puede variar en función del grado que estudian. En este caso, será especialmente interesante analizar qué distinto rendimiento tienen en las asignaturas de estadística los alumnos de las distintas carreras de *Analytics* de ICADE.

Una vez más, la muestra se encuentra equilibrada en cuanto a los grados cursados por los alumnos encuestados, lo que resultará positivo de cara a la generalización de conclusiones.

Finalmente, señalar que, nuevamente, por ser esta una variable demográfica, no se introducirá directamente en el modelo a la hora de hacer los clústers, sino que se empleará a posteriori, para ver el perfil de los distintos clústers resultantes en función de los grados que los componen.

Método de Estudio. En la Figura 4, puede observarse la distribución de la muestra en función de la metodología de estudio empleada por los alumnos (correspondiéndose los valores más altos con una metodología más teórica, y los valores más bajos con una metodología de estudio más orientada a hacer ejercicios).

Figura 4. *Distribución de la muestra según el método de estudio*



Según se ha visto en la revisión de la literatura, las diferentes metodologías empleadas por los alumnos a la hora de preparar la asignatura parecen ser relevantes, tanto en relación con la categorización de los alumnos, como a la hora de perfilar su rendimiento.

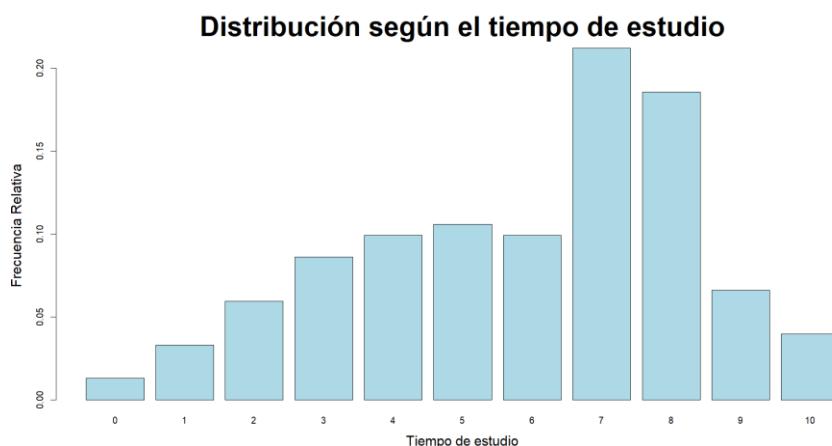
En el caso de esta muestra, se ha empleado una escala entre el 1 y el 6, representando el 1 un enfoque consistente en únicamente hacer ejercicios, y el 6 otro consistente en únicamente estudiar teoría.

Como puede observarse, la muestra se encuentra bastante equilibrada, con una media de 3,08, una desviación típica de 1 y una cierta asimetría positiva de 0,46. Se concluye,

por tanto, que resulta habitual entre los encuestados adoptar un enfoque mixto entre el estudio de la teoría y la práctica de ejercicios.

Tiempo de Estudio. En la Figura 5, puede observarse la distribución de la muestra en función del tiempo dedicado por los alumnos al estudio de las asignaturas de estadística (correspondiéndose los valores más altos con un mayor tiempo dedicado a estas asignaturas).

Figura 5. *Distribución de la muestra según el tiempo de estudio*



Según lo analizado en la revisión de la literatura, el tiempo dedicado al estudio de la estadística resulta relevante tanto para conocer el rendimiento de los alumnos como, sobre todo, para perfilar a éstos según sus características a la hora de abordar las asignaturas de estadística.

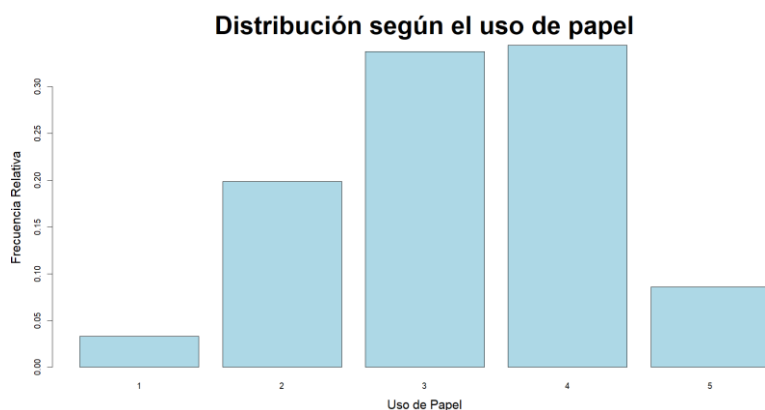
En el caso de esta muestra, se ha empleado una escala entre el 0 y el 10, representando el 0 no haberle dedicado casi tiempo a las asignaturas de estadística, y el 10 el haber empleado la práctica totalidad del tiempo disponible para el estudio en asignaturas de estadística.

Como se puede observar, la muestra parece bastante homogénea a lo largo de la escala, salvo para los números 7 y 8, en los que podemos identificar un considerable pico que concentra al 40% de los encuestados. Parecería, por tanto, que es entre estos valores entre los que se concentra el grueso del alumnado. En cualquier caso, los estadísticos

observados son una media de 5,89, una desviación típica de 2,39 y una asimetría negativa de -0,45.

Uso de Papel. En la Figura 6, puede observarse la distribución de la muestra en función del soporte empleado por los alumnos para el estudio de la asignatura de estadística (correspondiéndose los valores más altos con un mayor uso de soportes físicos en lugar de digitales).

Figura 6. *Distribución de la muestra según el uso de papel*



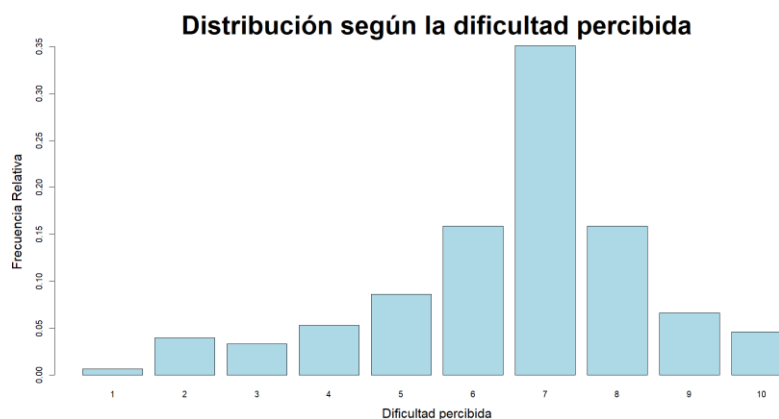
En relación con esta variable, se ha podido ver en la revisión de la literatura que su impacto no ha sido todavía plenamente abordado por la doctrina, y por tanto permanece, en gran medida, incierto.

En este caso, se ha distribuido la muestra según una escala entre 1 y 5, en la que el 1 representa el empleo exclusivo de soportes electrónicos para el estudio, mientras que el 5 representa a aquellos alumnos que utilizan únicamente soportes en papel.

A este respecto, la muestra parece equilibrada, con una media de 3,25, una desviación típica de 0,98, y una ligera asimetría negativa de -0,18. Es por ello por lo que, a este respecto, se puede concluir que es posible encontrar de todo entre el alumnado, empleando la mayoría de los estudiantes una combinación de ambos soportes.

Dificultad Percibida. En la Figura 7, puede observarse la distribución de la muestra en función de la dificultad percibida por los alumnos (correspondiéndose los valores más altos con una mayor dificultad percibida).

Figura 7. *Distribución de la muestra según la dificultad percibida*



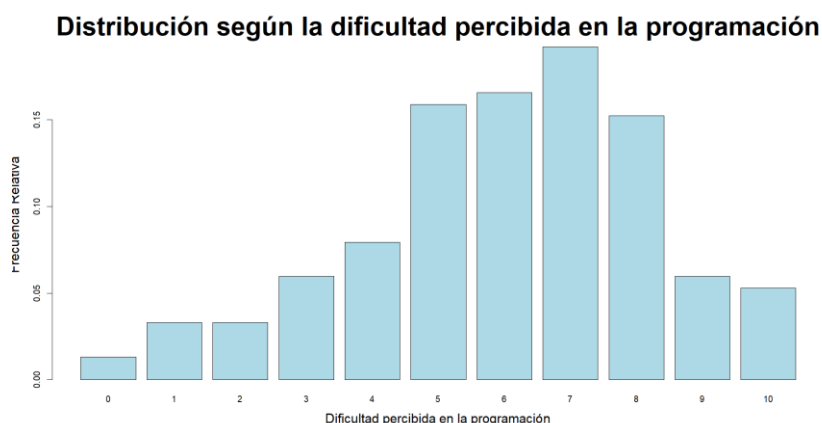
Como se ha visto, la distribución percibida a posteriori de la asignatura de estadística debería aparecer correlacionada inversamente con el rendimiento en estas materias.

En nuestra encuesta, se ha categorizado la dificultad percibida según una escala entre 0 y 10, de menor a mayor dificultad.

En relación con los resultados, puede verse que éstos se distribuyen de manera relativamente simétrica en torno al 7 (que acapara por sí solo el 35% de las observaciones). En cuanto a los estadísticos, se observa una media de 6,57, una desviación típica de 1,84 y una asimetría negativa de -0,7.

Dificultad Percibida en la Programación. En la Figura 8, puede observarse la distribución de la muestra en función de la dificultad percibida en la parte de programación de las asignaturas de estadística (correspondiéndose los valores más altos con una mayor dificultad percibida).

Figura 8. Distribución de la muestra según la dificultad percibida de la programación



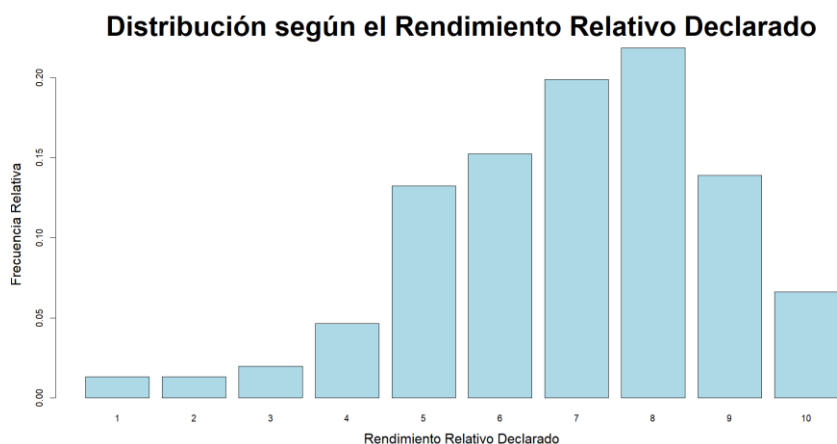
Según se ha visto a lo largo de la revisión de la literatura, existe una significativa e interesante correlación inversa entre la dificultad percibida de la parte de programación y el rendimiento global en la asignatura.

En esta encuesta, nuevamente, se ha medido esta variable según una escala entre el 0 y el 10, en la que el 10 representa la máxima dificultad percibida en la parte de programación.

A este respecto, resulta especialmente interesante comparar esta distribución con la de la dificultad percibida analizada previamente. Y es que, aunque nuevamente, las percepciones de la dificultad aparecen centradas en torno al 7, puede observarse ahora una dispersión mucho mayor, evidenciando ello una mucho mayor divergencia de opiniones en relación con la parte de programación. En cuanto a los estadísticos, la media se sitúa ahora en un 6,01, la desviación típica es de 2,24, y aparece nuevamente una asimetría negativa de -0,47.

Rendimiento Relativo Declarado. En la Figura 9, puede observarse la distribución de la muestra en función del Rendimiento Relativo Declarado por los alumnos en las asignaturas de estadística (correspondiéndose los valores más altos con un mayor Rendimiento Relativo Declarado).

Figura 9. Distribución de la muestra según el Rendimiento Relativo Declarado



Como se ha señalado al llevar a cabo la revisión de la literatura, parte del propósito de este trabajo es explicar cuáles son los principales factores que influyen en el rendimiento de los alumnos en las asignaturas de estadística, razón por la cual ésta será la variable de referencia, que se explicará a través de las demás (en referencia, naturalmente, a la red neuronal con interpretación *post-hoc*, ya que el clustering, al ser un modelo de aprendizaje no supervisado, carece de variable objetivo).

En todo caso, ha sido evaluada en la encuesta a través de una escala entre 0 y 10, representando el 10 un rendimiento autopercebido en las asignaturas de estadística muy superior al promedio de la clase (y viceversa).

Como puede verse, los resultados de la encuesta se distribuyen en torno al 7 y el 8, descendiendo abruptamente una vez pasamos la barrera del 5 (eso es, cuando se pasa a declarar un rendimiento inferior al promedio de la clase). En esta línea, la media se sitúa en un 6,91, con una desviación típica de 1,89 y una asimetría negativa de -0,62.

Como resulta evidente, estos datos difícilmente pueden corresponderse con la realidad, ya que es improbable estadísticamente hablando que tal cantidad de observaciones se sitúen por encima de la media. Las causas de este fenómeno ya son algo más discutible, pudiendo deberse a una infraestimación del rendimiento de los demás, a una sobreestimación

del rendimiento propio (*over-placement*), o simplemente a una declaración falsa a la hora de cumplimentar la encuesta.

En todo caso, esto tampoco tiene por qué conducir a desechar completamente estos datos. Y es que, aunque se ha visto que están muy probablemente inflados, lo cierto es que estos parecen seguir mostrando una distribución relativamente campaniforme en torno a una media (una distribución que sería la esperable, sin perjuicio de que esta media en torno a la que se agrupan los datos pueda estar sobreestimada). Además, como se vio en la revisión de la literatura, este es un problema al que se ha enfrentado habitualmente la doctrina, sin que aparentemente haga a los autores replantearse la validez de sus conclusiones.

Es por ello por lo que, en cualquier caso, se descarta eliminar esta variable.

Definición del Modelo

Se ha construido un modelo de clustering jerárquico, utilizando R como lenguaje de programación, y *factoextra* como librería de referencia.

Para la implementación del modelo, en tanto que todas las variables son numéricas, se ha empezado por escalar las variables, ya que no todas ellas se encontraban distribuidas conforme a una misma escala. A continuación, se ha construido la matriz de distancias, empleando para ello la distancia euclídea.

Seguidamente, se ha implementado el clustering jerárquico, utilizando como criterio de enlace el enlace de Ward (*ward.D2*). En base a este clustering jerárquico, se han elaborado las visualizaciones que se mostrarán a continuación.

El Modelo de Red Neuronal con análisis post-hoc

Variables del Modelo

En relación con las variables que han sido empleadas para entrenar la red neuronal, dado que estas ya han sido analizadas *in extenso* al estudiar el clustering jerárquico, bastará con remitirnos a este apartado para su análisis.

No obstante, y a diferencia de lo que se hizo en el clustering jerárquico, en la red neuronal sí se han introducido como variables explicativas la totalidad de las mismas (tanto numéricas como categóricas). Esto, no obstante, ha requerido eliminar las observaciones que contuviesen modalidades de bachillerato distintas de las tradicionales de Ciencias y Ciencias Sociales, ya que al contener cada una de ellas muy pocas observaciones, se generaban distorsiones indeseadas en la red neuronal.

En cuanto a la variable dependiente, se buscará predecir (con las reservas expuestas anteriormente) el Rendimiento Relativo Declarado, por ser esta variable el mejor *proxy* del éxito o fracaso académico de los alumnos.

Asimismo, a la hora de trabajar con variables categóricas, se ha hecho *one-hot-encoding* de cara a su implementación en la red neuronal. Ello implicará crear una variable nueva por cada categoría existente, marcando con un 0 o un 1 la concurrencia o no de la característica en la observación. Asimismo, se habrá de eliminar una de estas variables dicotómicas resultantes, para evitar los referidos problemas de multicolinealidad perfecta.

Finalmente, con el propósito de determinar si las variables son o no significativas, se ha construido un modelo de regresión para obtener los p-valores correspondientes a los coeficientes de las variables. Esto ha permitido, adicionalmente, comparar las betas obtenidas en el modelo de red neuronal con análisis *post-hoc*, con las betas de esta regresión lineal.

Definición del Modelo

Se ha construido una red neuronal acompañada de un análisis de sensibilidad *post-hoc*, utilizando R como lenguaje de programación, y *caret* y *NeuralSens* como librerías de referencia para la red neuronal y el análisis de sensibilidad *post-hoc*, respectivamente. Asimismo, la variable que se ha buscado predecir mediante la red neuronal ha sido el Rendimiento Relativo Declarado.

Como se ha señalado, para la implementación del modelo, se ha empezado escalando las variables numéricas y haciendo *one-hot-encoding* respecto de las variables categóricas. Seguidamente, se ha entrenado la red neuronal, haciendo un mallado para optimizar los hiperparámetros del modelo (a saber, el número de capas ocultas y el factor de caída). Así, probando entre una y cuatro capas ocultas, y un factor de caída de entre 0,1 y 10^{-7} , se ha concluido que los hiperparámetros óptimos eran una única capa oculta y un factor de caída de 0,1. Asimismo, se han programado un número máximo de 250 iteraciones y el RMSE como métrica de evaluación.

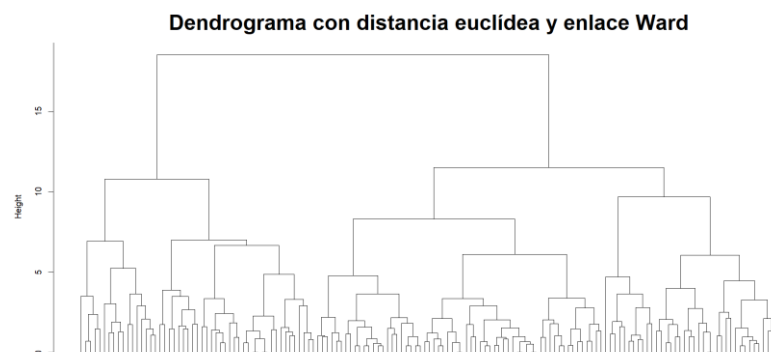
Por su parte, el análisis de sensibilidad *post-hoc* se ha realizado empleando la función disponible a tal efecto en la librería *NeuralSens*. Finalmente, para obtener los p-valores, se ha empleado un modelo de regresión OLS.

Resultados

Clustering jerárquico

En relación con el clustering jerárquico, se visualizará, en primer lugar, el dendrograma (**Figura 10**), para tener un panorama general de las agrupaciones existentes en los datos.

Figura 10. Dendrograma con distancia euclídea y enlace de Ward

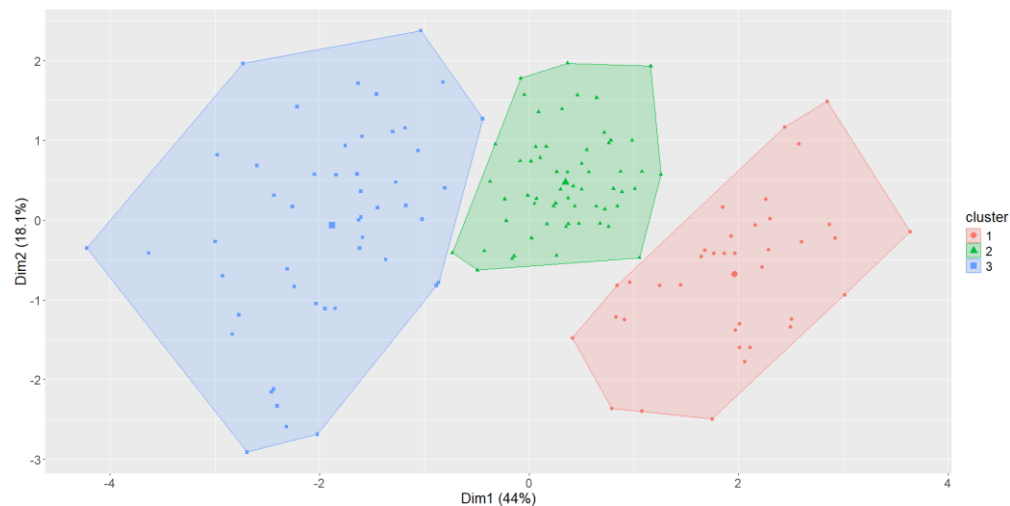


Si se atendiera exclusivamente a este dendrograma, pareciera que lo ideal sería hacer dos clústers. No obstante, atendiendo tanto a la revisión de la literatura como a la

interpretación de los propios clústers, la opción de tres clústers ha demostrado ser la mejor, atendiendo al propósito de este trabajo.

Así, se visualizarán a continuación los tres clústers resultantes (**Figura 11**).

Figura 11. Clústers



En relación con la interpretación de los clústers, se estudiarán, en primer lugar, las variables que conforman las componentes principales del dataset (para poder interpretar los ejes de la **Figura 11**), y a continuación, se estudiará el promedio de las variables en cada uno de los clústers, para mejorar la interpretación de los mismos.

Así, la composición de los ejes del gráfico, en función de las distintas variables del dataset es la que se muestra en la **Tabla 2**.

Tabla 2. Dos primeras componentes principales del dataset

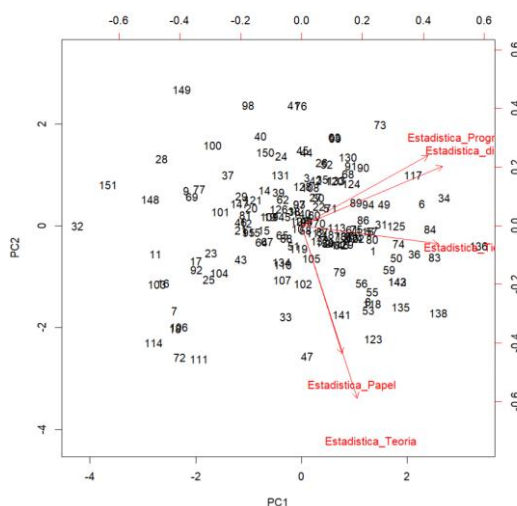
	PC1	PC2
Estadística_Teoría	0.2296456	-0.7362410
Estadística_Tiempo	0.5617590	-0.0767042
Estadística_Papel	0.1656862	-0.5456838
Estadística_dificultad	0.5789136	0.2547221
Estadística_Program	0.5187455	0.2990175

Como puede observarse, la primera componente principal (que se correspondería en la **Figura 11** con el eje de abscisas) es directamente proporcional al tiempo dedicado a la asignatura y la dificultad percibida tanto en el conjunto de la asignatura como en la parte de

programación. Por su parte, la segunda componente principal (que se correspondería en la **Figura 11** con el eje de ordenadas) es directamente proporcional a un método de estudio basado en la práctica de ejercicios y el empleo de medios electrónicos como soporte de trabajo.

De esta forma, los valores más escorados a la derecha de la **Figura 11** se corresponderán con alumnos a quienes la asignatura les cuesta más, y que a su vez le dedican más tiempo (y viceversa). Por su parte, los valores más escorados hacia arriba en la **Figura 11** se corresponderán con alumnos que preparan la asignatura haciendo ejercicios y empleando mayoritariamente soportes digitales (y viceversa). Para una mejor comprensión al respecto, puede observarse la **Figura 12**.

Figura 12. *Biplot del dataset*



A continuación, se analizará el promedio de las variables en cada uno de los clústers, para delimitar qué características singularizan cada uno de éstos. El resultado de este cálculo puede observarse en la **Tabla 3**.

Tabla 3. *Promedio de las variables en cada clúster*

Estadística_Teoría	Estadística_Tiempo	Estadística_Papel	Estadística_dificultad	Estadística_Program	cluster_A	prop
1.1197428	0.6917590	0.1726283	0.7023428	0.4500763	1	0.2516556
-0.4459325	0.4343614	0.1051513	0.2240421	0.2090847	2	0.4105960
-0.2922042	-1.0434755	-0.2564560	-0.7956792	-0.5895324	3	0.3377483

Como puede verse, los resultados son plenamente concordantes con la interpretación de los ejes de la **Figura 11**.

Así, el clúster 1 (el más escoreado hacia la derecha en la **Figura 11**), que agrupa a un 25% de la muestra, se caracteriza por un estudio fundamentalmente teórico de las asignaturas de estadística, por dedicarles bastante tiempo en términos relativos, y por percibir las asignaturas relativamente complicadas, tanto en su conjunto como en lo que respecta a la parte de programación.

Asimismo, el clúster 2 (situado a mitad del gráfico ligeramente hacia arriba en la **Figura 11**), que agrupa a un 41% de la muestra, destaca por que el método de estudio de sus integrantes consiste en hacer ejercicios. Pese a ello, los miembros de este clúster siguen percibiendo las asignaturas de estadística como materias relativamente complejas (en conjunto y en la parte de programación), y en consecuencia, le siguen dedicando un tiempo considerable.

Finalmente, el clúster 3 (situado a la izquierda del gráfico en la **Figura 11**), que agrupa a un 34% de la muestra, destaca por que, al contrario que los otros dos clústers, considera que las asignaturas de estadística son materias sencillas (en conjunto, y en particular, en lo que respecta a la parte de programación), y en consecuencia, les dedica relativamente poco tiempo. Además, aunque su enfoque sigue siendo práctico, tiene un mayor componente teórico que el del clúster 2.

En conjunto, resulta posible afirmar también que el empleo de medios electrónicos o en papel no parece tener un gran valor explicativo en nuestra muestra.

Para terminar el análisis de estos resultados, se verá qué perfil tienen cada uno de estos clústers desde el punto de vista demográfico y del rendimiento relativo declarado.

Tabla 4. Distribución de los clústers por Género

Cluster	Hombre	Mujer
1	0.5789474	0.4210526
2	0.4354839	0.5645161
3	0.4509804	0.5490196

Se ha de recordar, a este respecto, que la muestra contenía un 52% de mujeres y un 48% de hombres. Por tanto, según se observa en la **Tabla 4**, el único impacto de género significativo que parece haber dentro de los clústers es que los hombres se encuentran sobrerrepresentados en el clúster 1.

A continuación, se muestra la composición de los clústers por Grados en la **Tabla 5**.

Tabla 5. Distribución de los clústers por Grado

Cluster	E2+BA	E3+BA	E6+BA
1	0.4736842	0.1842105	0.3421053
2	0.4032258	0.1935484	0.4032258
3	0.1568627	0.6862745	0.1568627

Se ha de recordar, a este respecto, que la muestra estaba equilibrada en cuanto a los grados estudiados por los encuestados. Por tanto, lo que puede observarse en la distribución de los clústers es que los alumnos de E-3 Analytics se encuentran enormemente sobrerrepresentados en el clúster 3 (integrando el 68% del mismo pese a ser solo un tercio de la muestra). Por su parte, parece haber también una mayor proporción de alumnos de E-6 Analytics en el clúster 2, y una sobrerrepresentación de E-2 Analytics en el clúster 1.

Finalmente, se muestra la composición de los clústers por Especialidad de Bachillerato en la **Tabla 6**.

Tabla 6. Distribución de los clústers por Especialidad

Cluster	Bach_Internacional	Ciencias	Ciencias_Sociales	Otras
1	0.07894737	0.3947368	0.5263158	0.00000000
2	0.01612903	0.3709677	0.5645161	0.04838710
3	0.07843137	0.5490196	0.3529412	0.01960784

Poniendo el foco en las diferencias entre los alumnos procedentes de un bachillerato de Ciencias y los alumnos procedentes de un bachillerato de Ciencias Sociales, puede verse en la **Tabla 6** que, mientras que los alumnos de Ciencias Sociales son preponderantes en los clústeres 1 y 2, los alumnos procedentes de un bachillerato de Ciencias aparecen sobrerrepresentados en el clúster 3.

Para terminar el análisis de los resultados del clustering, veremos el rendimiento relativo declarado en cada uno de los clústers, junto a su relación con cada una de las variables demográficas estudiadas. Así, en la **Tabla 7** se muestra el Rendimiento Relativo Declarado por clúster.

Tabla 7. Rendimiento Relativo Declarado por clúster

Cluster	Rendimiento Relativo Declarado Promedio
1	6.394737
2	6.709677
3	7.549020

Como puede observarse en la **Tabla 7**, los clústers aparecen ordenados de peor a mejor rendimiento académico (con un rendimiento notablemente superior en el clúster 3).

Seguidamente, en la **Tabla 8**, se muestra el Rendimiento Relativo Declarado por Género.

Tabla 8. Rendimiento Relativo Declarado por Género

Genero	Rendimiento Relativo Declarado Promedio
Hombre	6.875000
Mujer	6.949367

Como puede observarse en la **Tabla 8**, apenas hay diferencias entre hombres y mujeres en cuanto a su Rendimiento Relativo Declarado. Estos resultados parecen concordantes con los perfiles de los clústers.

Seguidamente, en la **Tabla 9**, se muestra el Rendimiento Relativo Declarado por Grado:

Tabla 9. Rendimiento Relativo Declarado por Grado

Grado	Rendimiento Relativo Declarado Promedio
E2+BA	6.941176
E3+BA	7.222222
E6+BA	6.521739

Como puede observarse en la **Tabla 9**, el Rendimiento Relativo Declarado en la muestra es mayor entre los alumnos de E-3 Analytics, y menor entre los alumnos de E-6 Analytics. Esto resulta sorprendente, si se contrasta con los resultados del *clustering*. Y es que, por un lado, la diferencia de E-3 Analytics respecto del resto de carreras no parece ser tan acusada. Por su parte, E-2 Analytics tiene un mejor promedio que E-6 Analytics, lo que no resulta consistente con el hecho de que E-2 Analytics aparece mayormente asociado al clúster 1, que se correspondería con un menor rendimiento académico.

Finalmente, en la **Tabla 10**, se muestra el Rendimiento Relativo Declarado por Especialidad de Bachillerato:

Tabla 10. Rendimiento Relativo Declarado por Especialidad

Especialidad	Rendimiento Relativo Declarado Promedio
Bach_Internacional	6.875000
Ciencias	7.136364
Ciencias_Sociales	6.780822
Otras	5.750000

Poniendo el foco, nuevamente, en las diferencias entre los bachilleratos de Ciencias y Ciencias Sociales, puede verse en la **Tabla 10** que haber cursado un bachillerato científico aparece asociado a un mejor rendimiento académico (lo que resulta coherente con los resultados del *clustering*).

Red Neuronal con análisis post-hoc

En relación con la red neuronal con análisis *post-hoc*, el propósito del trabajo ha sido identificar los principales *drivers* explicativos del Rendimiento Relativo Declarado.

En primer lugar, si se analizan los resultados de la red neuronal, puede verse que los pesos obtenidos son los que se muestran en la **Tabla 11**.

Tabla 11. Coeficientes asociados a las variables en la red neuronal.

	mean	std	meanSensSQ
Estadistica_Teoria	0.012774886	0.017683965	0.021763974
Estadistica_Tiempo	-0.323524566	0.447847206	0.551173599
Estadistica_Papel	0.004011136	0.005552519	0.006833584
Estadistica_dificultad	-0.255380929	0.353517623	0.435080487
Estadistica_Program	0.076373865	0.105722488	0.130114564
Genero_Mujer	0.243164652	0.336606927	0.414268190
Especialidad_Ciencias_Sociales	-0.172550441	0.238857388	0.293966077
\`Grado_E3+BA\`	-0.073510176	0.101758353	0.125235832
\`Grado_E6+BA\`	0.096811435	0.134013720	0.164933092

Estos mismos resultados pueden observarse de manera gráfica en la **Figura 13**, la **Figura 14** y la **Figura 15**.

Figura 13. Visualización de las variables explicativas del Rendimiento Relativo Declarado

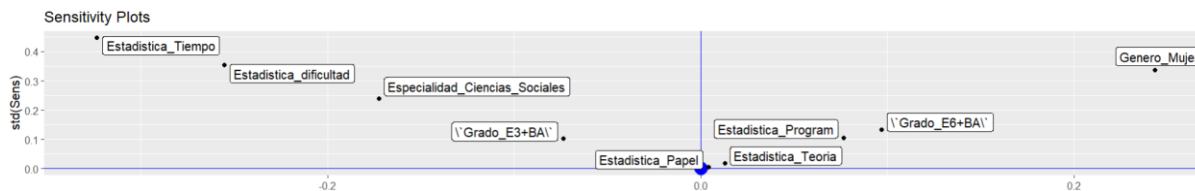


Figura 14. Pesos de las variables explicativas del Rendimiento Relativo Declarado

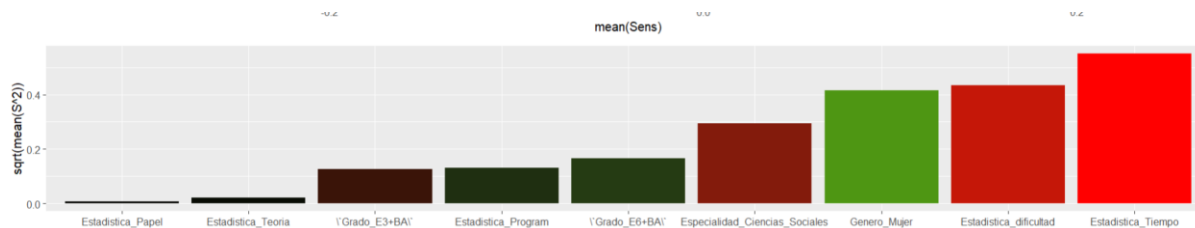


Figura 15. Distribución de las variables explicativas del Rendimiento Relativo Declarado



A la hora de interpretar estos resultados, se ha de entender que el potencial explicativo de cada una de las variables viene determinado por el valor absoluto de su media (una media que no es sino el promedio de las pendientes analizadas en el análisis *post-hoc*, respecto de la curva dibujada por la red neuronal). Así, el signo positivo o negativo de cada una de las medias de las variables indica la correlación positiva o negativa entre dicha variable explicativa y la variable estudiada (el Rendimiento Relativo Declarado). Por su parte, el valor absoluto de la media reflejaría, como se ha señalado, el potencial explicativo de la variable en cuestión.

Por otro lado, es importante dejar claro que, tal y como se muestra en la **Figura 15**, la distribución de las sensibilidades de cada una de las variables dibuja una curva casi campaniforme (no aparecen distribuciones multimodales), y la desviaciones típicas son pequeñas, lo que es indicativo de que no hay efectos no lineales relevantes.

No obstante, se ha de señalar, a modo de reserva, que en vistas de los resultados del *clustering*, algunas de las variables pueden estar correlacionadas (como el tiempo de estudio y la dificultad percibida tanto en el conjunto de la asignatura como en la parte de programación). Es por ello por lo que sus efectos podrían llegar a ser confundidos por el modelo (multicolinealidad imperfecta).

Finalmente, habiendo hecho también un modelo de regresión lineal para obtener los p-valores asociados a los coeficientes de las variables, los resultados son los que se muestran en la **Tabla 12**.

Tabla 12. Coeficientes asociados a las variables en la regresión lineal

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.02509759	0.20974468	0.1196578	0.904987319
Estadistica_Teoría	-0.03846610	0.09463960	-0.4064482	0.685256561
Estadistica_Tiempo	-0.17248254	0.13934571	-1.2378030	0.218603259
Estadistica_Papel	-0.06180482	0.09331273	-0.6623407	0.509232207
Estadistica_dificultad	-0.39332347	0.12037012	-3.2676172	0.001474184
Estadistica_Program	0.12059821	0.10653625	1.1319922	0.260266204
Genero_Mujer	0.21446872	0.19952288	1.0749079	0.284927553
Especialidad_Ciencias_Sociales	0.04978193	0.19017416	0.2617702	0.794021375
``Grado_E3+BA``	-0.18805589	0.29589386	-0.6355519	0.526478092
``Grado_E6+BA``	-0.26502759	0.25004514	-1.0599190	0.291660649

Lo primero que debe observarse es la similitud que guardan los coeficientes obtenidos en esta regresión lineal con los obtenidos en la red neuronal (lo cual es indicativo de que esta se ha construido adecuadamente).

Asimismo, como puede observarse, tomando un 5% de umbral de significación, la única variable respecto de la cual resulta posible afirmar que se han obtenido resultados significativos, es la dificultad percibida.

Con todo, el análisis *post-hoc* de los resultados de la red neuronal conduce a las siguientes conclusiones acerca de las variables:

- **Género:** se aprecia una correlación positiva entre ser mujer y tener un mejor Rendimiento Relativo Declarado. El efecto parece ser coherente (si bien tal vez un poco sobredimensionado) en relación con los resultados del *clustering*. En cualquier caso, los resultados no pueden considerarse significativos.
- **Método de estudio:** no se encuentra correlación con el Rendimiento Relativo Declarado. Esto podría ser consistente con que tanto en el clúster 1 como en el clúster 3 haya una mayor tendencia al estudio teórico que en el clúster 2. En cualquier caso, los resultados no pueden considerarse significativos.
- **Tiempo de estudio:** se aprecia una correlación negativa, de tal forma que un menor tiempo de estudio aparecería vinculado a un mayor Rendimiento Relativo Declarado. Esto resulta coherente con los resultados del *clustering*. En cualquier caso, los resultados no pueden considerarse significativos.
- **Uso de Papel:** no se encuentra correlación con el Rendimiento Relativo Declarado. Esto resulta coherente con los resultados del *clustering*. En cualquier caso, los resultados no pueden considerarse significativos.
- **Dificultad Percibida:** se aprecia una correlación negativa, de tal forma que una menor dificultad percibida vendría asociada a un mayor Rendimiento Relativo Declarado. Ello resulta coherente con los resultados del *clustering*.
- **Dificultad Percibida en la Programación:** se aprecia cierta correlación positiva con el Rendimiento Relativo Declarado, lo cual resultaría inconsistente con los resultados del *clustering*. En cualquier caso, los resultados no pueden considerarse significativos.

Discusión

En primer lugar, se ha de señalar que, tal y como indican algunos autores dentro de la doctrina (Jauhari et al., 2021; Tosunoğlu, 2017), la muestra de este estudio ha dado lugar a una distribución conforme a tres clústers. Como se señalaba en la revisión de la literatura, los estudios discrepan en cuanto al tamaño relativo de los clústers, y en este caso, se han identificado tres clústers que, ordenados por su rendimiento académico y metodología de

estudio (según se ha visto), representan el 25%, el 41% y el 34% de la muestra, respectivamente.

Género

Como se ha visto en la revisión de la literatura, no existe consenso en la doctrina acerca de cuál es el impacto del género sobre el rendimiento de los alumnos en las asignaturas de estadística, así como que, más allá del promedio, la distribución del rendimiento pudiera ser distinta en función del género.

En la muestra estudiada, se ha visto, en primer lugar, que el rendimiento de las mujeres es ligeramente superior al de los hombres (Rendimiento Relativo Declarado de 6,95 frente a 6,88). Además de ello, se ha visto que las mujeres están ligeramente sobrerrepresentadas entre los perfiles de alumnos con mejor rendimiento (clústers 2 y 3) y significativamente infrarrepresentadas en el grupo de alumnos con peor rendimiento (clúster 1). No obstante, se ha visto también que no se han identificado impactos significativos del Género a la hora de explicar el Rendimiento Relativo Declarado.

Es por ello que los resultados de este estudio se alinean con los obtenidos por Mohamed et al. (2011), Judi et al. (2011), Ngamsiriudom et al. (2022) o Lindsay (2021). No obstante, dadas las características técnicas de este estudio, sería conveniente ampliar el tamaño y diversidad de la muestra, para ver si se pueden alcanzar conclusiones significativas al respecto.

Asimismo, en relación con la hipótesis de la mayor variabilidad masculina, no parece que esta muestra la confirme, ya que, si bien los hombres predominan entre el alumnado con peor rendimiento, entre aquellos alumnos con mejor desempeño, son las mujeres, por el contrario, las que aparecen sobrerrepresentadas.

Especialidad

Como se ha visto en la revisión de la literatura, existe discusión en la doctrina acerca de si el hecho de haber cursado un bachillerato científico tiene algún impacto en el rendimiento de los alumnos en las asignaturas de estadística.

Atendiendo a los resultados de este estudio, no parece que tal afirmación pueda sostenerse con rotundidad, ya que, si bien es cierto que la muestra tiende a apreciar un mejor rendimiento de los alumnos del bachillerato de Ciencias, lo cierto es que se ha visto que estas diferencias no resultan estadísticamente significativas.

Por ello, se ha de coincidir con los estudios llevados a cabo por Arroyo-Barrigüete et al. (2023), Dupuis et al. (2012), Mohamed et al. (2011) y Bourne (2014).

Sería interesante, pese a todo, de cara a futuros estudios, analizar si estas diferencias en función de la Especialidad pueden ser significativas, y confirmado este punto, si son homogéneas entre todos los contenidos de las asignaturas de estadística, o únicamente se limitan a los aspectos cuantitativos y no conceptuales, o a aquellas habilidades más “elevadas”, como señalaban algunos autores.

Grado

Como se ha visto en la revisión de la literatura, existen autores que defienden que el rendimiento en la asignatura de estadística no es homogéneo entre las distintas carreras universitarias. En este estudio, no obstante, no se ha observado que el Grado sea un factor significativo a la hora de explicar el Rendimiento Relativo Declarado de los alumnos, y en consecuencia, no resulta posible realizar la afirmación tan extendida de que el rendimiento difiere según el Grado cursado.

Método de Estudio

Como se vio en la revisión de la literatura, la doctrina tiende a concluir que la metodología consistente en la práctica de ejercicios aparece generalmente asociada a un mejor rendimiento académico.

En el caso de este estudio, no obstante, tal conclusión no resulta del todo clara. Y es que, como se vio, aunque los alumnos del clúster con peor rendimiento sí tienen una tendencia generalizada al estudio teórico, lo cierto es que esta tendencia hacia el estudio de conceptos también se observa entre aquellos con mejor desempeño (de tal forma que aquellos alumnos que más tienden a la práctica de ejercicios son aquellos situados en la “zona media”).

Ello probablemente se deba a que estos alumnos con mejor rendimiento tienen la facilidad de aprehender los conceptos sin necesidad de practicar tantos ejercicios, resultando interesante realizar un estudio al respecto.

En cualquier caso, se ha observado que la variable del método de estudio carece de ningún impacto significativo a la hora de explicar el Rendimiento Relativo Declarado de los alumnos, y en consecuencia, debe tenerse por no probado que afecte a tal variable.

Tiempo de Estudio

Como se vio en la revisión de la literatura, la doctrina parece asociar un mayor tiempo de estudio con un mejor rendimiento académico. Y ello tiene sentido, ya que, en términos generales, un mayor esfuerzo por parte de los alumnos conduce a mejores resultados.

En la muestra de este estudio, en cambio, se observa todo lo contrario: que es el menor tiempo de estudio el que se asocia a un mayor Rendimiento Relativo Declarado. Aunque ello pudiera parecer contraintuitivo, puede tener una explicación relativamente sencilla, que no es otra que el hecho de que ICADE es una universidad extremadamente selectiva.

Esto conduce a un fuerte sesgo de selección de sus alumnos, que lleva a que, quienes dedican un menor tiempo de estudio a una asignatura no lo hacen por dejadez o falta de esfuerzo, sino porque no requieren de más tiempo para su estudio personal. Eso lleva a que sean precisamente los alumnos con menores capacidades intelectuales los que tengan que dedicar un mayor tiempo al estudio, al contrario de lo que ocurriría en otras universidades.

Es por ello que resultaría interesante, de cara a futuras investigaciones, llevar a cabo estudios similares en otros entornos, para comprobar si este efecto se reproduce o no.

En cualquier caso, se ha de señalar, una vez más, que los efectos apreciados en cuanto a esta variable no pueden considerarse estadísticamente significativos, de tal forma que se habrá de estar a estudios posteriores para confirmar o rechazar estos efectos.

Uso de Papel

Como se ha visto en la revisión de la literatura, la mayoría de estudios no encuentran efectos significativos en el rendimiento académico asociados al empleo de uno u otro soporte para el estudio (atribuyendo, si acaso, unos mejores resultados a la metodología digital).

Estos resultados se reproducen en este estudio, en el que no se han identificado diferencias significativas en cuanto al rendimiento, respecto de si se utilizan para el estudio papel y lápiz o dispositivos electrónicos. Así, se habrán de suscribir las tesis de Williams (2012).

Dificultad Percibida

Como se ha visto en la revisión de la literatura, la doctrina tiende a asociar una menor dificultad percibida con un mejor rendimiento académico en las distintas asignaturas de estadística.

Tal ha sido el caso en este estudio pues, como resulta lógico, la sensación de una menor dificultad se deriva directamente de un mayor dominio de la materia – un dominio que se asocia directamente a un mejor rendimiento en estas asignaturas. Así se ha demostrado tanto en el *clustering* (siendo el clúster 3 aquel con la menor dificultad percibida) como en el análisis de las variables determinantes del rendimiento académico, siendo los efectos al respecto estadísticamente significativos.

Dificultad Percibida de la Programación

Como se ha visto en la revisión de la literatura, una menor dificultad percibida en la parte de la programación de las asignaturas de estadística aparece asociada a un mejor

rendimiento en estas asignaturas (dejando incluso de lado el efecto de una menor dificultad percibida en el conjunto de la asignatura).

Si bien los resultados del *clustering* parecieran seguir esta línea, los resultados de la red neuronal interpretable con análisis *post-hoc* resultan diametralmente opuestos. No obstante, y como se ha visto, estos resultados no han sido estadísticamente significativos, de tal forma que, por muy ilógico que pudiera parecer este efecto, podría explicarse, sencillamente, por una cuestión de puro azar.

Es por ello que sería preciso realizar ulteriores estudios sobre la materia para confirmar o descartar estas tesis.

Conclusiones

A lo largo de este trabajo, se ha analizado el estudio de la estadística en grados no-STEM, y con tal propósito, se ha analizado cuáles son los diferentes perfiles de alumnos a la hora de enfocar las asignaturas de estadística en estas carreras, así como cuáles son los principales factores que influyen en su rendimiento.

Para ello, previa revisión de la literatura al respecto, se ha trabajado con una muestra de 151 alumnos de tres dobles grados de Analytics de ICADE, a quienes se ha hecho la encuesta que puede encontrarse en el **Anexo I**. Los resultados de la encuesta han sido objeto de un preprocesado, en el que se ha hecho una imputación de valores faltantes mediante la técnica que mejores resultados aportaba en cada variable en concreto, de entre una regresión lineal o logística, un KNN, un Random Forest o una aproximación estadística.

El dataset resultante ha sido objeto de dos técnicas de machine learning: por un lado, un clustering jerárquico del que se han extraído tres perfiles de alumnos, y por otro lado, una red neuronal con análisis *post-hoc*, de la que se ha extraído el impacto de las diferentes variables estudiadas en el rendimiento de los alumnos.

En virtud de estos resultados, y en base a las hipótesis planteadas, tenemos que:

Hipótesis 1. La hipótesis se confirma, ya que podemos observar tres clústers cuantitativamente diferenciados y cualitativamente significativos.

Hipótesis 2. La hipótesis permanece incierta, ya que el mayor Rendimiento Relativo Declarado de las mujeres no resulta significativo.

Hipótesis 3. La hipótesis no se confirma, pues se ha observado en la muestra que hay una mayor proporción de mujeres que de hombres en el clúster de alumnos con mayor rendimiento.

Hipótesis 4. La hipótesis permanece incierta, ya que el mayor Rendimiento Relativo Declarado de los alumnos procedentes de un bachillerato científico no resulta significativo.

Hipótesis 5. La hipótesis permanece incierta, ya que no se han podido apreciar diferencias estadísticamente significativas por Grados en cuanto al Rendimiento Relativo Declarado.

Hipótesis 6. La hipótesis permanece incierta, ya que, aunque es cierto que los alumnos con peor rendimiento académico estudian con un enfoque mayoritariamente teórico, los alumnos con mejor rendimiento tienen también cierta tendencia al estudio teórico-conceptual, y en todo caso, no se han apreciado diferencias estadísticamente significativas al respecto.

Hipótesis 7. La hipótesis permanece incierta, pues la ausencia de impacto del soporte empleado en el Rendimiento Relativo Declarado de los alumnos que se ha apreciado, no ha resultado estadísticamente significativa.

Como puede verse, únicamente se han podido extraer conclusiones respecto de dos de las siete hipótesis planteadas. Esto probablemente sea consecuencia de las tres principales limitaciones del trabajo. En primer lugar, en reducido tamaño de la muestra, respecto del número de variables empleadas (que ha dificultado la extracción de conclusiones significativas respecto de estas variables, aumentando en exceso la dimensionalidad del

dataset). En segundo lugar, la falta de diversidad de la muestra, al haberse realizado la encuesta únicamente entre alumnos de ICADE. Y en tercer lugar, el empleo como métrica del Rendimiento Relativo Declarado en lugar del rendimiento real, que como se ha visto, ha derivado en problemas como consecuencia de que sus resultados no eran ajustados a la realidad (con independencia de que ello pudiera deberse a declaraciones deliberadamente falsas, o a percepciones erróneas acerca del rendimiento ajeno).

Precisamente como consecuencia de estas limitaciones, a modo de futuras líneas de investigación, resultaría interesante replicar este estudio utilizando, por un lado, datos reales del rendimiento de los alumnos, y por otro, una muestra más variada que incluya a alumnos de otras universidades. Resultaría también interesante profundizar en algunas de las cuestiones que han ido surgiendo a lo largo del estudio, como el debate sobre el impacto del método de estudio en el rendimiento de los alumnos.

Declaración respecto al uso de Chat GPT u otras herramientas de IAG

Por la presente, yo, Gonzalo Villarías Vozmediano, estudiante de E-3 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “El estudio de la estadística en grados no STEM”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
2. **Interpretador de código:** Para realizar análisis de datos preliminares.
3. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
4. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han

dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Referencias

- Akimov, A., Malin, M., Sargsyan, Y., Suyunov, G. y Turdaliev, S. (2024). Student success in university in a first-year statistics course: do students' characteristics affect their academic performance? *Journal of Statistics and Data Science Education*, 32(1), 24-35. <https://doi.org/10.1080/26939169.2023.2184435>.
- Arroyo Barrigüete, J. L., Carabias López, S., Hernández Estrada, A., y Segura, M. (2023). Efecto de la especialidad en bachillerato en el rendimiento matemático en la universidad: un estudio comparativo en grados de Administración de Empresas. *Revista de Educación*, 402, 115-140. <https://doi.org/10.4438/1988-592X-RE-2023-402-597>.
- Baye, A. y Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-scale Assessments in Education*, 4(1). <https://doi.org/10.1186/s40536-015-0015-x>.
- Berenson, M. L., Ramnarayanan, R. y Oppenheim, A. (2015). Assessing the disconnect between grade expectation and achievement in a business statistics course. *Journal of Education for Business*, 90(2), 72-80. <https://doi.org/10.1080/08832323.2014.973828>.
- Bologna, E. L., Vaiman, M. y Alfonso, M. A. (14 de agosto de 2019). *Aligning evaluation with achievement objectives: automated exams based on bloom's taxonomy* [Resumen de presentación de la conferencia]. IASE 2019 Satellite Conference: Kuala Lumpur, Malaysia. Recuperado el 21 de febrero de 2025 de <https://cutt.ly/lreauldk>.

- Bourne, V. J. (2014). To what extent is mathematical ability predictive of performance in a methodology and statistics course? Can an action research approach be used to understand the relevance of mathematical ability in psychology undergraduates? *Psychology Teaching Review*, 20(2), 14-27. Recuperado el 21 de febrero de 2025 de <https://cutt.ly/CresCIUL>.
- Dupuis, D. N., Medhanie, A., Harwell, M., Lebeau, B., Monson, D. y Post, T. R. (2012). A multi-institutional study of the relationship between high school mathematics achievement and performance in introductory college statistics. *Statistics Education Research Journal*, 11(1), 4-20. <https://doi.org/10.52041/serj.v11i1.337>.
- Fabbricatore, R., Parola, A., Pepicelli, G. y Palumbo, F. (2023). Successful factors in statistics learning for non-STEM courses students: a PLS-PM approach. *Statistica Applicata – Italian Journal of Applied Statistics*, 34(2). <https://doi.org/10.26398/IJAS.0034-010>.
- Förster, M. y Maur, A. (2015). Statistics anxiety and self-concept of beginning students in the social sciences – a matter of gender and socio-cultural background? *Zeitschrift für Hochschulentwicklung*, 10(4), 67-90. <https://doi.org/10.3217/zfhe-10-04/04>.
- Grönemann, M. (2024). Learning statistics by doing sociology. Applying inquiry-based learning un undergraduate methods course. *HINT – Heidelberg Inspirations for Innovative Teaching*, 5(1), 93-114. <https://doi.org/10.11588/hint.2024.1.108590>.
- Halsey, L. G., Careau, V., Pontzer, H., Ainslie, P. N., Andersen, L. F. Anderson, L. J., Arab, L., Baddou, I., Bedu-Addo, K., Blaak, E. E., Blanc, S., Bonomi, A. G., Bouten, C. V. C., Bovet, P., Buchowski, M. S., Butte, N. F., Camps, S. G. J. A., Close, G. L., Cooper, J. A.,...Speakman, J. R. (2022). Variability in energy expenditure is much greater in males than females. *Journal of Human Evolution*, 171(1), 103229. <https://doi.org/10.1016/j.jhevol.2022.103229>.

- Hammonds, F. y Mariano, G. (2015). Student test grades in college: a study of possible predictors. *International Journal of Teaching and Learning in Higher Education*, 27(1), 114-118. Recuperado el 21 de febrero de 2025 de <https://cutt.ly/YreamlQX>.
- Hedges, S. (2017). Statistics student performance and anxiety: comparisons in course delivery and student characteristics. *Statistics Education Research Journal*, 16(1), 320-336. <https://doi.org/10.52041/serj.v16i1.233>.
- Hossain, B. y Tsigaris, P. (2012). Are grade expectations rational? A classroom experiment. *Education Economics*, 23(2), 199-212. <https://doi.org/10.1080/09645292.2012.735073>.
- Jauhari, A. L. R., Ariany, R. L., Fardillah, F. y Ayu, A. (2021). Profile of students' statistical capabilities in introductory social statistics courses. *Journal of Physics: Conference Series*, 1764(1), 12118. <https://dx.doi.org/10.1088/1742-6596/1764/1/012118>.
- Jonsdottir, A. H., Bjornsdottir, A. y Stefansson, G. (2017). Difference in learning among students doing pen-and-paper homework compared to web-based homework in an introductory statistics course. *Journal of Statistics Education*, 25(1), 12-20. <https://doi.org/10.1080/10691898.2017.1291289>.
- Judi, H. M., Ashaari, N. S., Mohamed, H. y Tengku Wook, T. M. (2011). Students profile based on attitudes towards statistics. *Procedia Social and Behavioral Sciences*, 18(1), 266-272. <https://doi.org/10.1016/j.sbspro.2011.05.038>.
- Kien-Kheng, F., Azlan. N., Ahmad, S. N. D., Leong, N. L. H. y Mohamed, I. (2016). Relationship between cognitive factors and performance in an introductory statistics course: a Malaysian case study. *Malaysian Journal of Mathematical Sciences*, 10(3), 269-282. Recuperado el 22 de febrero de 2025 de <https://cutt.ly/2relGNV7>.

- Lindsay, K. G. (2021). Predicting success in a statistics course geared toward Allied Health Students. *International Journal of Research in Education and Science*, 7(2), 339-350. <https://doi.org/10.46328/ijres.1545>.
- Mansor, R., Zaini, B. J., Yusof, Z. M., Gabda, D. y Wong, W. (2020). Undergraduate students' profile based on attitudes towards statistics. *ASM Science Journal*, 13(1). [https://doi.org/10.32802/asmscj.2020.sm26\(2.23\)](https://doi.org/10.32802/asmscj.2020.sm26(2.23)).
- Mitra, S. (2023). How are students learning in a Business Statistics course? Evidence from both direct and indirect assessment. *INFORMS Transactions on Education*, 23(2), 95-103. <https://doi.org/10.1287/ited.2022.0270>.
- Mitra, S. y Le, K. (2019). The effect of cognitive and behavioral factors on student success in a bottleneck statistics course via deeper analytics. *Communications in Statistics – Simulation and Computation*, 51(6), 2779–2808. <https://doi.org/10.1080/03610918.2019.1700279>.
- Mohamed, H., Ashaari, N. S., Judi, H. M. y Tengku Wook, T. S. M. (2011). Factors affecting FTSM students' achievement in statistics course. *Procedia – Social and Behavioral Sciences*, 59(1), 125-129. <https://doi.org/10.1016/j.sbspro.2012.09.255>.
- Mustaffa, N. A. y Zulkifli, M. (2019). The influence of students' numeracy skills on the attainment of course learning outcome in introductory statistics course. *AIP Conference Proceedings*, 2138(1), 50022. <https://doi.org/10.1063/1.5121127>.
- Ngamsiriudom, W., Devkota, M. L. y Menon, M. K. (2022). Can gender and major explain college students' performance in Business Statistics? *American Business Review*, 25(2), 253-269. <https://doi.org/10.37625/abr.25.2.253-269>.
- Norshahidi, N. D., Ismail, N. H. S., Abidin, S. Z. y Razak, N. F. A. (2021). Profile identification of students' attitude towards statistics course: a case study in UiTM Pahang Branch

(Raub Campus). *Mathematical Sciences and Informatics Journal*, 2(1), 63-69.

Recuperado el 22 de febrero de 2025 de <https://cutt.ly/0redgex9>.

O'Dea, R. E., Lagisz, M., Jennions, M. D. y Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications*, 9(1), 3777. <https://doi.org/10.1038/s41467-018-06292-0>.

Peiró-Signes, Á., Trull, O., Segarra-Oña, M. y García-Díaz, J. C. (2021). Anxiety towards statistics and its relationship with students' attitudes and learning approach. *Behav. Sci.*, 11(3), 32. <https://doi.org/10.3390/bs11030032>.

Rahayu, R., Kartono, Dwijanto, Agoestanto, A., Waluya, S. T. B. y Dewi, N. R. Profile of statistical problem-solving ability based on the mathematical disposition. *AIP Conference Proceedings*, 2577(1), 20051. <https://doi.org/10.1063/5.0096117>.

Saki, A., Tabesh, H., Yousefi, R. y Khalili, S. (2017). Clustering the attitudes towards statistics and technology among medical post graduate students. *British Journal of Applied Science & Technology*, 18(4), 1-11. <https://doi.org/10.9734/BJAST/2016/29673>.

Saputra, K. V. I., Cahyadi, L. y Sembiring, U. A. (2018). Assessment of statistical education in Indonesia: preliminary results and initiation to simulation-based inference. *IOP Conference Series: Journal of Physics: Conference Series*, 948(1), 12033. <https://dx.doi.org/10.1088/1742-6596/948/1/012033>.

Sharma, A. M. y Srivastav, A. (2021). Study to assess attitudes towards statistics of Business School students: an application on the SATS-36 in India. *International Journal of Instruction*, 14(3), 207-222. <https://doi.org/10.29333/iji.2021.14312a>.

Tosunoğlu, N. (2017). Adaptive neuro-fuzzy inference system (ANFIS) approach for modelling the effect of achievement in statistics to students' attitudes towards statistics. *Cappadocia Academic Review*, 1(2), 38-53. Recuperado el 22 de febrero de 2025 de <https://cutt.ly/hrexgToG>.

Williams, A. (2012). Online homework v. Traditional homework: statistics anxiety and self-efficacy in an educational statistics course. *Technology Innovations in Statistics Education*, 6(1). <https://doi.org/10.5070/T561011683>.

Xu, J., Ong, J., Tran, T., Kollar, Y., Wu, A., Vujicic, M. y Hsiao, H. (2021). The impact of study and learning strategies on post-secondary student academic achievement: a mixed-methods systematic review. *PsyArXiv*. <https://doi.org/10.31234/osf.io/7ng5y>.

Anexo I: Encuesta

Fecha:

Grado (señala la opción que corresponda): E2 + Analytics / E3 + Analytics / E6 + Analytics

Curso (señala la opción que corresponda): 3º / 4º / 5º

Soy mayor de edad y acepto que mis respuestas, tratadas de forma absolutamente anónima, sean empleadas con la finalidad de investigación académica

☐ Sí

☐ No

En asignaturas de **Estadística** del doble grado (Estadística y Probabilidad y Métodos Estadísticos) consideras que no tuviste que aprender nada de memoria (marca el 0) o lo aprendiste todo de **memoria** (marca el 10). Selecciona lo que corresponda.

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

En asignaturas de **Estadística** del doble grado (Estadística y Probabilidad y Métodos Estadísticos) ¿aproximadamente qué porcentaje del tiempo de estudio dedicaste a estudiar teoría y qué porcentaje a practicar problemas? Selecciona lo que corresponda.

☐ 0% teoría – 100% problemas (1)

☐ 20% teoría – 80% problemas (2)

- ☐ 40% teoría – 60% problemas (3)
- ☐ 60% teoría – 40% problemas (4)
- ☐ 80% teoría – 20% problemas (5)
- ☐ 100% teoría – 0% problemas (6)

Comparado con otras asignaturas de tu doble grado (de todas, no solo las de Analytics) ¿Cuánto **tiempo** dedicaste a estudiar las asignaturas de **Estadística** (Estadística y Probabilidad y Métodos Estadísticos)? Valóralo en una escala del 0 (mucho menos) al 10 (mucho más).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

En asignaturas de **Estadística** del doble grado (Estadística y Probabilidad y Métodos Estadísticos), cuando estudiabas la asignatura ¿lo hacías usando dispositivos electrónicos o con papel y bolígrafo? Selecciona lo que corresponda.

- ☐ Solo uso dispositivos electrónicos (ordenador, tablet, etc.) (1)
- ☐ La mayoría del estudio lo hago con dispositivos electrónicos (ordenador, tablet, etc.) (2)
- ☐ La mitad en papel y la mitad con dispositivos electrónicos (3)
- ☐ La mayoría del estudio lo hago en papel (4)
- ☐ Todo el estudio lo hago en papel (5)

¿Cuál consideras que es el grado de **dificultad** de las asignaturas de **Estadística** (Estadística y Probabilidad y Métodos Estadísticos)? Valóralo en una escala del 0 (extremadamente fácil) al 10 (extremadamente difícil).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Comparado con la **dificultad que esperabas** ¿Cuál consideras que es el grado de dificultad de las asignaturas de **Estadística** (Estadística y Probabilidad y Métodos Estadísticos)? Valóralo en una escala del 0 (extremadamente más fácil de lo que esperaba al comenzar la asignatura) al 10 (extremadamente más difícil de lo que esperaba al comenzar la asignatura).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Centrándonos solo en la parte de **programación** en R, ¿Cuál consideras que es el grado de dificultad de LA PARTE DE PROGRAMACIÓN de las asignaturas de **Estadística** (Estadística y Probabilidad y Métodos Estadísticos)? Valóralo en una escala del 0 (extremadamente fácil) al 10 (extremadamente difícil).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Si volvieras a empezar tu doble grado ¿Cuál de estas dos opciones crees que sería más adecuada de cara a tu **carrera profesional** que comenzarás en breve?

- ☐ Estudiar un único lenguaje de programación para conocerlo en mayor profundidad (1)
- ☐ Estudiar varios lenguajes de programación para tener una visión general de todos (2)

¿Cuál consideras que ha sido tu **rendimiento** (nota obtenida) en las asignaturas de **Estadística** (Estadística y Probabilidad y Métodos Estadísticos)? Valóralo en una escala del 0 (muy por debajo de la media de mi clase) al 10 (muy por encima de la media de mi clase).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

¿Qué **especialidad cursaste en bachillerato**?

☐Ciencias ☐Ciencias Sociales ☐Otras _____

¿Cuál es tu **género**?

☐Hombre ☐Mujer ☐Otros _____

¿Qué nota obtuviste en la **EvAU** (fase general, escala del 0 al 10)? _____

¿Por qué elegiste un doble grado que incluía Analytics? Puedes marcar más de una:

- ☐ Me gustan las matemáticas y/o la programación
- ☐ Se me dan bien las matemáticas y/o la programación
- ☐ Buenas salidas profesionales
- ☐ No era mi primera opción, pero no entré en la titulación que quería
- ☐ Influencia de padres y/o amigos
- ☐ Está de moda
- ☐ Tenía una nota muy alta en EvAU y quería aprovecharla
- ☐ Otras (especificar): _____