



Facultad de Ciencias Económicas y Empresariales

# Análisis de la relación entre la situación socioeconómica de los padres y rendimiento académico de los hijos en España

Autor: Jesús Dueñas Huecas

Director: Antonio Rúa Vieites

MADRID

Junio 2020

## Resumen

En este Trabajo de Fin de Grado se analiza la relación entre la situación socioeconómica de los padres y el rendimiento académico de sus hijos en España desde el prisma de las Ciencias Económicas. Para ello se revisa profundamente la bibliografía actual sobre el asunto, a fin de poder establecer unas hipótesis de investigación que aporten valor y nueva información. Se hace uso de la base de datos de 29.253 observaciones referentes a alumnos recogida por el Instituto Nacional de Evaluación Educativa, y se emplea la herramienta Rstudio para hallar conclusiones empíricas y sustentadas por la estadística. Se corrobora que factores tales como acudir a un centro privado, un mayor nivel de estudios de los padres y un mayor nivel socioeconómico y cultural, se ven relacionados con un mayor rendimiento académico.

Palabras clave: capital humano, desarrollo económico, educación, rendimiento académico, nivel educativo.

## Abstract

In this Final Degree Project, the relationship between the socioeconomic situation of parents and the academic performance of their children in Spain is analyzed from the perspective of Economic Sciences. The current bibliography on the subject is thoroughly reviewed in order to establish research hypotheses that provide value and new information. A database of 29,253 observations referred to Spanish students collected by the National Institute of Educational Evaluation, and Rstudio tool are used to find empirical and supported by statistics conclusions. It is found that factors such as going to a private school, a higher level of education of the parents and a higher socioeconomic and cultural level, are related to higher academic performance.

Key words: human capital, economic development, education, academic performance, educational level.

# Índice

<b>1. Objetivo y utilidad del trabajo:</b>	<b>4</b>
<b>2. Metodología empleada:</b>	<b>6</b>
<b>3. Revisión de la bibliografía:</b>	<b>7</b>
La Educación en España	8
El Rendimiento Académico	9
<b>4. Formulación de las hipótesis de investigación:</b>	<b>13</b>
<b>5. La base de datos:</b>	<b>15</b>
<b>6. Análisis de la muestra:</b>	<b>19</b>
Tratamiento de la base de datos	19
Comprobación hipótesis de investigación	22
1. Los alumnos cuyos padres tengan más estudios, desempeñan un mejor rendimiento académico.	24
2. Los alumnos que acuden a un centro privado, desempeñan un mejor rendimiento académico.	38
3. Los alumnos que gozan de un mayor apoyo familiar en referencia a los estudios, desempeñan un mejor rendimiento académico.	44
4. A mayor tiempo dedicado a los deberes, los alumnos muestran un mejor desempeño académico.	51
5. Los alumnos que históricamente presentaban un peor R.A. tienen un peor desempeño académico.	55
6. A mayor nivel socioeconómico y cultural, los alumnos desempeñan un mejor rendimiento académico.	60
<b>7. Conclusiones:</b>	<b>63</b>
<b>Bibliografía:</b>	<b>67</b>
<b>Anexo I: Resto de Variables de la Base de Datos.</b>	<b>69</b>
<b>Anexo II: Código empleado para el análisis en Rstudio.</b>	<b>73</b>

## 1. Objetivo y utilidad del trabajo:

El porqué del estudio y cómo pueden emplearse las conclusiones.

Debido al interés creciente que la Educación ha ido tomando en las sociedades desarrolladas; su estudio, análisis y formulación de teorías han ido ganando progresivamente un gran protagonismo. El objetivo ha sido en todo momento la optimización de recursos y la mejora de la Educación como institución, ya que ésta se comprende como un buen indicador del grado de desarrollo de las naciones (PNUD, 2008). El afán por comparar y cuantificar el grado de avance de la Educación en las sociedades del mundo, unido a la era del Big Data en la que vivimos, ha propiciado la creación de grandes bases de datos en referencia a este campo (OCDE, s.f.).

En este Trabajo de Fin de Grado se buscará analizar en qué forma y medidas guardan relación la situación socioeconómica de los progenitores con el rendimiento académico de sus hijos. Para ello se revisará la bibliografía escrita sobre el asunto, enfocándolo desde la perspectiva de las Ciencias Económicas y Sociales. Posteriormente, haciendo uso de principios econométricos y estadísticos y de los programas informáticos apropiados para el análisis de las bases de datos, se comprobará la validez de las hipótesis formuladas en relación al asunto. Para ello se utilizará una base de datos de alumnos españoles de Educación Secundaria Obligatoria, recogida por el Instituto Nacional de Evaluación Educativa, dependiente del Ministerio de Educación y Formación Profesional.

Se obtendrán las conclusiones probadas de que la familia guarda un gran efecto en el rendimiento académico del alumno. De forma que el mayor nivel de estudios de los padres, mayor apoyo al estudio por parte de la madre y mayor tiempo dedicado al estudio, así como no haber repetido ningún curso, provocan un significativo efecto positivo en el rendimiento académico del alumno. Mientras que el apoyo al estudio por parte del padre no produce efectos, y el estudiar más de una hora al día tampoco, ya que el rendimiento es igual que para aquellos alumnos que dediquen al estudio tres horas o más. En lo referente al plano familiar estrictamente económico, el acudir a un centro privado favorece un mejor rendimiento, y el nivel socioeconómico y cultural guarda una relación positiva con un mejor desempeño.

En cuanto a la utilidad del estudio, éste sirve como punto de partida para tratar de imitar las buenas conductas que producen efectos positivos en el rendimiento académico. Es decir, sirve para que se investiguen las diferencias entre la enseñanza pública y privada a fin de que, en la medida de la posible, la pública trate de imitar aquellas características que producen el mejor rendimiento que desempeñan los alumnos de centros privados. A su vez, las conclusiones del presente trabajo pueden ser útiles en cuanto a tratar de fomentar el apoyo materno al alumno dado que, como se ha comprobado, éste guarda una relación con un mejor rendimiento. En tercer lugar, el hecho de que los alumnos que estudian como mínimo una hora al día tengan un rendimiento significativamente mayor que aquellos que dedican menos tiempo, puede servir como base para tratar de fomentar el estudio en al menos una hora diaria. En la misma línea de lo mencionado, aquellas conclusiones que se han obtenido en el presente estudio pueden ser sumamente útiles en cuanto a fomentar (en la medida de lo posible) las prácticas que conllevan una mejora del rendimiento académico, para así producir una mejora en la educación en su conjunto, y de forma indirecta en la economía (Neira, 2007; OEHHA).

## 2. Metodología empleada:

### El procedimiento por el que se llegará a las conclusiones.

Con el fin de hallar unas conclusiones o *insights* veraces, significativas y relevantes, el estudio que lleva a las mismas se ha llevado a cabo siguiendo un proceso dictado por una metodología. Ésta está basada en un enfoque deductivo que se detallará en el presente apartado.

En primera instancia se realizará una revisión de la bibliografía, con objeto de ser conscientes de los estudios previos que otros autores han realizado sobre la materia, ya que servirán como base a la hora de plantear las hipótesis de investigación. A continuación mostraremos qué información nos confiere la base de datos empleada, de cara a poder trabajar con las variables que disponemos. Y finalmente haremos uso de una metodología basada en un enfoque deductivo, que se divide en tres secciones principales:

1. Formulación de la teoría: exposición de las hipótesis básicas de investigación y argumentación de su ser. Dichas hipótesis estarán en sintonía con los objetivos del Trabajo y utilizarán la información revelada anteriormente por otros autores. La cual quedará mostrada en el apartado 3, referente a la revisión de la bibliografía.
2. Recogida de datos: análisis de la base de datos mostrada (en este caso, será la base de datos utilizada para el informe PISA del año 2009). Con vistas a facilitar y simplificar su comprensión, se presentará y expondrá de forma esquematizada. En el apartado 5: “La Base de Datos”, será donde se muestre la mencionada base, con la que se trabajará en pos de verificar las hipótesis básicas de investigación.
3. Verificación de la teoría: haciendo uso mayoritariamente del programa estadístico informático RStudio, comprobaremos si se cumplen las hipótesis de investigación. A partir de entonces, con la información hallada, podremos mostrar las conclusiones del estudio, habiendo hecho uso de la estructura y metodología comentadas.

### 3. Revisión de la bibliografía:

Repaso de aquellos insights que otros autores han aportado sobre la materia.

En primer lugar, debido a la polisemia y las múltiples formas de concebir los conceptos a continuación mostrados, conviene homogeneizar la comprensión de algunas palabras clave. A su vez, en reiteradas ocasiones el texto se referirá a los mismos. Por ello, en pos de facilitar la comprensión y evitar confusiones, a continuación se muestra la definición a la nos referiremos en el presente Trabajo cada vez que se mencione alguna de las siguientes palabras:

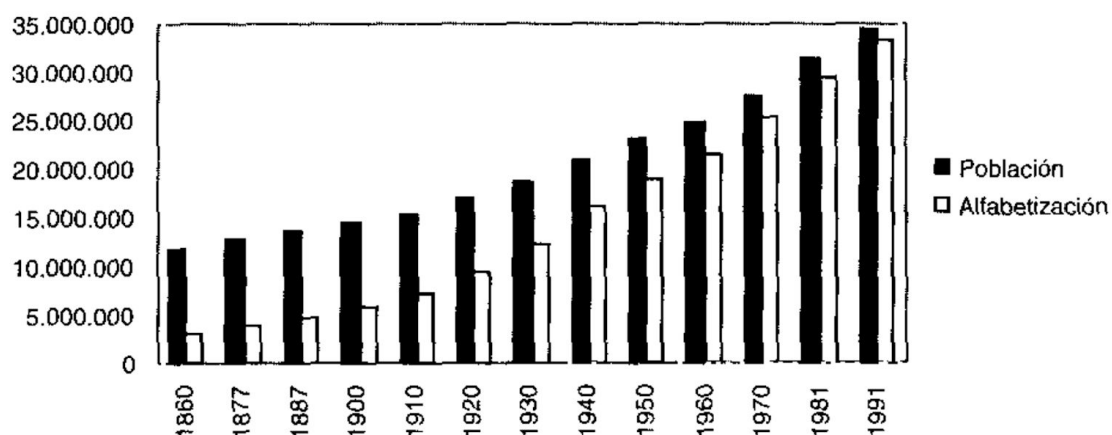
- **Alfabetización:** “acción de alfabetizar”. Alfabetizar: “enseñar a alguien a leer y escribir” (2ª acepción), según la Real Academia Española (RAE).
- **Capital Humano:** pese a que haya innumerables definiciones, una muy representativa es el capital humano como conjunto de capacidades productivas que un individuo adquiere por acumulación de conocimientos generales o específicos (Becker, 1964). En referencia al tema que nos atañe, Schultz (1961) señala que “invertir en escolarización, salud, en la forma de crianza de los niños, en profesionales, en investigación, es invertir en capital humano”. En este contexto, cuando se hace referencia al Capital Humano disponible de una sociedad, se refiere al conjunto del mismo presente en todos los individuos que la componen.
- **Crecimiento/Desarrollo Económico:** “incremento de la renta nacional o el PIB por persona de un país o una región o un grupo de países” (Expansión). De forma muy parecida lo expresa González (2008): “El crecimiento económico es una de las metas de toda sociedad, implica un incremento notable de los ingresos, y de la forma de vida de todos los individuos de una sociedad.” Y señala que el incremento del Producto Interior Bruto de un país a lo largo del tiempo, sirve para medir el Desarrollo Económico del mismo.
- **Educación:** “crianza, enseñanza y doctrina que se da a los niños y a los jóvenes” (2ª acepción), e “instrucción por medio de la acción docente” (3ª acepción), según la RAE.

- **Nivel Educativo:** “nivel de educación más alto que una persona ha terminado” ya sea Educación Primaria, Educación Secundaria, Bachillerato, u otros (OEHHA).
- **Rendimiento Académico/Escolar:** “nivel de conocimientos demostrado en un área o materia comparado con la norma de edad y nivel académico” (Jiménez, 2000), o desempeño de los alumnos, el cual suele quedar plasmado en las calificaciones (Pérez y Gardey, 2008; Erazo, 2012; Edel, 2003).

## La Educación en España

A través del tiempo, la Educación, comprendida como la Institución encargada de formar a las nuevas generaciones para que desempeñen un trabajo satisfactorio, ha ido ganando importancia de forma exponencial. En el ámbito nacional, si bien es cierto que en la Baja Edad Media, en consecuencia de las ideas humanistas del momento, ya se comenzaba a puntualizar en la enseñanza de la lectura y escritura, ésta se limitaba al estrato social del clero. El pueblo llano, cuya actividad económica principal se basaba en la agricultura y ganadería, tuvo que esperar a que las ideas ilustradas calasen en la población, para que de esta forma la Educación se fuese integrando en el conjunto de la sociedad (Kagan, 1981; Delgado, 1993; Viñao, 1993). Fue a partir de la segunda mitad del Siglo XIX cuando se comenzó la costosa y gran batalla contra el analfabetismo, de forma que en el año 2016 este se diese sólo en el 1,75% de la población española adulta (Datosmacro). Como podemos observar en el Gráfico 1, en el que se relaciona la población y la alfabetización en España desde 1860, esta creció de forma exponencial:

Gráfico 1: Alfabetización en España 1860-1991



Fuente: De Gabriel, N. (1997).



Por ende, la Educación ha ido ganando peso en la sociedad española a lo largo de la historia reciente. Sin embargo, el motor social que subyace tras el impulso de la Educación en un país es la mejora del Capital Humano, el cual desemboca en una mejora de la economía nacional, como señala Neira (2007). Mientras que el motor individual que promueve la mejora de la educación de los ciudadanos es una mayor remuneración, ligada al mayor nivel educativo (OEHHA). Por ello, una de las bases de los estados de la OCDE es un alto nivel educativo de sus ciudadanos. En base a este hecho, dicha organización elabora el informe PISA (Programme for International Student Assessment), en el que se evalúa el Rendimiento Académico en las áreas de la ciencia, lectura y matemáticas (PISA), con un carácter meramente cuantitativo (Popkewitz, 2013). El Nivel Educativo se entiende como un factor tan determinante a la hora de evaluar el nivel de desarrollo de un país, que es uno de los tres índices sobre los que se calcula el Índice de Desarrollo Humano, elaborado cada año por el Programa de las Naciones Unidas para el Desarrollo.

Por ello, al ser tan relevante la Educación en la Economía, es muy importante poder cuantificar su nivel de desempeño por parte de los alumnos. Mientras que el Nivel Educativo hace referencia a la última etapa educativa superada (Educación Primaria, Educación Secundaria, Bachillerato, etc.), Jiménez (2000) define el Rendimiento Académico como el “nivel de conocimientos demostrado en un área o materia comparado con la norma de edad y nivel académico”, por lo que nos basaremos en este último para el estudio que nos atañe, dada la posibilidad de observarse de forma cuantitativa.

## El Rendimiento Académico

El Rendimiento Académico o Escolar hace referencia al propio desempeño de los alumnos, el cual suele quedar plasmado en las calificaciones (Pérez y Gardey, 2008; Erazo, 2012; Edel, 2003). Ha habido una gran cantidad de estudios cuyo objetivo ha sido hallar una explicación probada, razonable y veraz de los factores de los que depende este Rendimiento Académico (R.A. a partir de ahora). Erazo (2012) señala que “el Rendimiento Académico no es un producto que sólo se centra en el estudiante o el docente o en su interacción, sino que es el resultado de múltiples variables de tipo personal y social, que provoca a los profesionales en educación, ciencias sociales y psicología educativa a analizar, describir y evaluar.” Por lo que

nos lleva a pensar que hay un gran número de variables que afectan en el R.A., ya sean de tipo cualitativo o cuantitativo.

En lo referente a su estudio en el ámbito temporal, un gran número de estudios e informes han probado que el mejor predictor del rendimiento académico futuro es el rendimiento anterior (Goberna y otros, 1987; House, Hurst y Keely 1996; Jiménez, 1987; Wilson y Hardgrave, 1995). Y en línea con esta hipótesis, de Miguel (1988) concluye su obra con la afirmación de que hay un “elevadísimo grado de correspondencia entre las puntuaciones de rendimiento curso a curso”, señalando que “la probabilidad de éxito o fracaso a lo largo de la E.G.B. [...] queda asociada al rendimiento obtenido durante el primer curso”.

Sin embargo, focalizando el estudio en aras de tratar de explicar el R.A. en función de variables exógenas, es importante ser conscientes de que otros tantos autores han tratado de identificar las variables más explicativas en este campo. Ruiz de Miguel (2001) señala que “queda suficientemente probada [...] la hipótesis de que el rendimiento académico del alumno no se debe exclusivamente a la labor desempeñada en el centro educativo, sino que sobre él ejerce una poderosa influencia el entorno familiar”. En la misma línea, Erazo (2012) indica que “En el factor personal se especifican características de tipo orgánico, cognitivo, estrategias y hábitos de aprendizaje, motivación, autoconcepto, emoción y conducta y en el factor social están las características de tipo familiar, escolar, socioeconómico y cultural que se muestran como variables que atraviesan el acto educativo y su resultado en el rendimiento y la nota académica”. Este último autor nos comunica que en base a su estudio hay dos tipos de variables que influyen en el R.A.: el “factor personal” y el “factor social”.

En referencia al “factor social” Feldman et al (2008) indican que “las condiciones favorables de salud mental están asociadas con un mayor apoyo social y un menor estrés académico”. Y en el estudio de Lastre, López y Alcázar (2018) “se determinó una relación estadísticamente significativa entre el nivel de rendimiento y apoyo familiar. Esto indica que aquellos padres que acompañan, retroalimentan y están pendientes de la vida escolar, sus hijos muestran mejores niveles de desempeño escolar.” Asimismo, focalizando el estudio en el impacto del entorno social familiar en el R.A., Pérez-Díaz, Rodríguez y Sánchez (2001) reflejan que hay un gran consenso en que las familias deben ser las responsables de educar, como refleja el Gráfico 2:

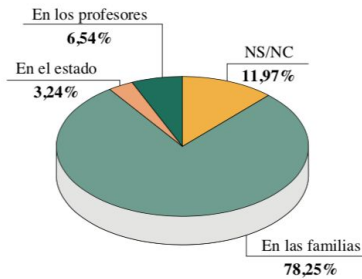
Gráfico 2: Encuesta sobre la responsabilidad de educar

---

¿En quién recae la principal responsabilidad de educar a la generación siguiente?

N = 300

---



---

Fuente: Encuesta ASP 00.030.

Fuente: Pérez-Díaz, V., Rodríguez, J.C. & Sánchez, L. (2001).

A su vez, estos tres autores llevaron a cabo un estudio con 4833 alumnos de primaria o ESO en 1999-2000, y concluyeron también que en relación al “progenitor que más se ocupa de la educación de los hijos en cada hogar, la distribución de la muestra por sexos está muy escorada del lado femenino: 82% de mujeres”. En referencia a la misma hipótesis de que las madres suelen tener un mayor peso en la educación de los hijos, Mella y Ortiz (1999) señalaron que “las variables contextuales en torno a la madre adquieren mayor relevancia. Especialmente las expectativas de la madre respecto a la carrera educacional de su hijo/a son decisivas”.

Es por ello que otros tantos autores han tratado de establecer la influencia del entorno familiar, y en particular de los padres, en el R.A. de sus hijos. Lorenzo, Santos y Godás (2012), en base al Informe PISA de 2006, concluyeron que “cuando los estudios [de los padres] son superiores, el aumento de «notables» y «sobresalientes» es importante, sobre todo en las familias autóctonas”, y que “cuanto menor nivel de estudios de ambos padres, más número de «suspensos» en las cuatro materias (CCNN, L.Castellana, L.Gallega y Matemáticas) y más disminuye el número de «notables» y «sobresalientes»”. En base a esta última afirmación, añadieron que “esta apreciación se observa con más claridad y cuantía en las familias inmigrantes”, ya que estos investigadores también estudiaron el R.A. de los alumnos de esta condición, señalando en este campo que “los estudiantes inmigrantes desempeñan un peor rendimiento académico que los nativos”, sobre todo en los casos en los que los padres tienen menor nivel de estudios, como se ha señalado previamente.

En referencia a la forma en que desempeña su día a día el alumno en su casa, también se ha afirmado que el hecho de que los padres ayuden a los hijos a hacer los deberes, se relaciona con un menor número de suspensos; y que el hecho de que los padres mantengan una relación estrecha con el centro escolar provoca un mayor R.A. (Peralbo y Fernández, 2003). También los padres más jóvenes y con más estudios, ayudan más a sus hijos con los deberes; y a mayor tiempo dedicado a los deberes, mejor rendimiento (Pérez-Díaz, Rodríguez y Sánchez, 2001). En cuanto a la televisión y su consumo, el mismo estudio realizado por Pérez-Díaz, Rodríguez y Sánchez (2001), mostraba como conclusión que “el tener TV en el cuarto del niño empeora su rendimiento académico. Y a mayor consumo, menor rendimiento”.

Y por último, en referencia al sexo del estudiante, de Miguel (1988) señala que “las niñas tienden a mejores calificaciones que los varones en el áreas de lengua e inferiores en matemáticas”, aunque podemos inferir de alumnos estudiantes del grado de Administración y Dirección de Empresas el factor relevante de que “las alumnas tienen un rendimiento significativamente superior al de los alumnos” (Martínez et al, 2010).

Por ello, cabe pensar que de alguna forma u otra, la familia ejerce una gran influencia en el R.A. de los hijos, la investigación y trabajos realizados en este estudio se orientan hacia la resolución de esta gran pregunta sobre la que ya han arrojado cierta luz los autores mencionados: ¿cómo y cuánto influye la familia en el R.A. de los alumnos? En el presente Trabajo utilizaremos los estudios que han realizado otros autores en referencia a esta cuestión para así poder enfocarnos en tratar de comprobar la influencia del nivel socioeconómico de los padres en el R.A. de los hijos, además de otros aspectos relacionados muy relevantes.

#### 4. Formulación de las hipótesis de investigación:

Qué vamos a probar en este trabajo.

El objetivo de este estudio es analizar y evaluar la relación entre la situación socioeconómica de los padres y el rendimiento académico de los hijos en España, por ello, probaremos las siguientes hipótesis de investigación:

- *Los alumnos cuyos padres tengan más estudios, desempeñan un mejor rendimiento académico.*  
La base de esta hipótesis es la afirmación de que a mejores estudios de los padres, mejor R.A. (Santos y Godás, 2012).
- *Los alumnos que acuden a un centro privado, desempeñan un mejor rendimiento académico.*  
En la línea de la hipótesis anterior, la base es la trabajo de los padres, mejor R.A. (Santos y Godás, 2012), infiriendo que a mayores ingresos, mayor probabilidad de acudir a un centro privado.
- *Los alumnos que gozan de un mayor apoyo familiar en referencia a los estudios, desempeñan un mejor rendimiento académico.*  
Esta hipótesis está construída sobre la afirmación de que un apoyo familiar sano, favorece “mejores niveles de desempeño escolar” (Lastre, López y Alcázar, 2018; Pérez-Díaz, Rodríguez y Sánchez, 2001).
- *A mayor tiempo dedicado a los deberes, los alumnos muestran un mejor desempeño académico.*  
Como señalan Peralbo y Fernández (2003) y Pérez-Díaz, Rodríguez y Sánchez, (2001).
- *Los alumnos que históricamente presentaban un peor R.A., desempeñan un peor rendimiento académico.*  
Esta hipótesis servirá para comprobar la efectividad del modelo educativo que permite repetir cursos, además de probar la relevancia del pasado en el R.A. actual (de Miguel, 1988; Goberna y otros, 1987; House, Hurst & Keely 1996; Jiménez, 1987; Wilson & Hardgrave, 1995).
- *A mayor nivel socioeconómico y cultural, los alumnos desempeñan un mejor rendimiento académico.*

Dado que la familia tiene un gran peso en la educación del estudiante (Ruiz de Miguel, 2001), y el nivel socioeconómico de la misma influye positivamente (Lorenzo, Santos y Godás, 2012).

Una vez planteadas estas hipótesis de investigación, pasaremos a mostrar y explicar la base de datos con la que trabajaremos, así como las variables que contempla. Esta base tiene una gran relevancia, ya que será aquello sobre lo que comprobaremos las hipótesis mencionadas, para así hallar unas conclusiones veraces y que aporten valor a la ciencia.

## 5. La base de datos:

Cómo es la base de datos sobre la que construiremos el estudio.

Este Trabajo de Fin de Grado se construirá sobre bases de datos de alumnos españoles de último curso de Educación Secundaria Obligatoria recogidas por el Instituto Nacional de Evaluación Educativa, dependiente del Ministerio de Educación y Formación Profesional.

En particular se empleará la realizada en el año 2009, disponible en la página web del mencionado Ministerio. En este caso, previo procesamiento de la base de datos para retirar aquellas observaciones defectuosas, disponemos de 29.253 observaciones de alumnos, pertenecientes a 933 centros educativos distintos de toda España. El motivo por el que se recogió esta base de datos fue para utilizarla en el informe PISA de dicho año. Es por ello que, además de haber muchas variables en relación con la vida del estudiante, también encontramos los resultados de las cuatro pruebas que fueron evaluadas, referentes a los temas: Lingüística, Matemática, Ciencias y Mundo Físico, y Ciencias Social y Ciudadana.

Sin embargo, lo más relevante y el motivo principal de la utilización de esta base de datos es el hecho de que nos muestra variables en relación a cómo los alumnos desempeñan su estudio en casa, por lo que en relación a los resultados en las cuatro pruebas, servirán para estudiar el impacto de variables sociales como el entorno familiar. Cabe destacar que la presente base de datos nos brinda un gran número de variables. En el espacio destinado a los anexos, bajo el título de “Anexo I: Resto de Variables de la Base de Datos.”, podemos encontrar aquellas variables no referentes directamente al tema al que este Trabajo se enfoca. A continuación pasaremos a explicar brevemente aquellas que nos resultan útiles en nuestro estudio, agrupándolas en función de a qué aspecto del estudiante se refieren:

Variables **identificativas**: que nos servirán para identificar a cada alumno:

- N° de Centro y N° de Alumno.
- **Titularidad** del Centro: Pública o Privada.
- **Lengua**: Castellano, Catalán, Gallego, Valenciano, Vasco, o No contesta.
- **Sexo**: Hombre, Mujer, o No contesta.
- Fecha de Nacimiento: día, mes y año.

VARIABLES REFERENTES A LA **FAMILIA**:

- **Vive habitualmente** con la Madre, con el Padre, con Hermana/Hermano, con Otros familiares, con Otras situaciones: pudiendo responder Sí, No o No contesta a cada pregunta.
- **Lugar de nacimiento** del alumno, de la madre y del padre: pudiendo responder la Comunidad Autónoma (si es en España), el país, o No contesta.
- Si naciste en otro país, edad al llegar España.
- **Estudios** de la madre y del padre: pudiendo responder a cada caso Universitarios superiores, Universitarios medios, Técnico superior de FP, Bachillerato, Técnico FP grado medio, Estudios obligatorios (ESO,EGB), No completó estudios obligatorios, No lo sé, o No contesta.
- **Situación laboral** de la madre y del padre: pudiendo responder a cada caso Trabaja fuera de casa, En casa de forma remunerada, Está en paro, Está jubilado, Sólo las tareas de casa, No lo sé, o No contesta.
- **Trabajo** de tu madre y de tu padre: pudiendo responder a cada caso Miembros del poder ejecutivo, legislativo, personal directivo administración y empresas, Profesionales científicos e intelectuales, Técnicos y profesionales de nivel medio, Empleados de oficina, Trabajadores de los servicios y vendedores de comercios y mercados, Agricultores y trabajadores cualificados agropecuarios y pesqueros, Oficiales, operarios y artesanos de artes mecánicas y otros oficios, Operadores de instalaciones y máquinas y montadores, Trabajadores no cualificados, o No contesta.

VARIABLES REFERENTES AL **HISTORIAL ACADÉMICO**:

- Años al empezar el colegio o guardería: pudiendo responder Antes de los 2 años, A los 2 años, A los 3 años, A los 4 años, A los 5 años, A los 6 años o más, o No contesta.
- No repetido ningún curso: pudiendo responder Sí, No o No contesta a cada opción.
- Ha **repetido**:
  - 2º de Educación Primaria
  - 4º de Educación Primaria
  - 6º de Educación Primaria
  - 1º de ESO
  - 2º de ESOpudiendo responder Sí, No o No contesta a cada opción.



Actividades y deberes **fuera del horario escolar:**

- Tiempo dedicado a **estudiar** después de clase: Generalmente, no tengo deberes: pudiendo responder 1 hora o menos, De 1 a 2 horas, De 2 a 3 horas, Más de 3 horas, o No contesta.

- **Ayuda deberes:**

- De mis padres
- De mis hermanos
- De mis amigos
- De un profesor o profesora particular
- Asisto a una academia
- No recibo ayuda de nadie

pudiendo responder Sí, No o No contesta a cada opción.

- **Actividades** fuera del colegio

- Ver televisión, vídeos o DVD
- Escuchar música
- Practicar deporte
- Leer libros, cuentos, cómic
- Leer prensa o revistas
- Jugar con amigos
- Hablar por teléfono
- Jugar con videojuegos o con el ordenador
- Participar en chats o messenger
- Usar internet

pudiendo responder Nada de tiempo, Menos de 30 minutos, Entre 30 y 60 minutos, Entre 1 y 2 horas, Más de 2 horas, o No contesta a cada opción.

- Frecuencia utilización **ordenador:**

- En casa
- En el colegio
- En otros lugares

pudiendo responder Nunca o casi nunca, Una o dos veces al mes, Una o dos veces a la semana, Todos o casi todos los días, o No contesta a cada opción.

- **Internet** para:

- Buscar información para tus estudios
- Buscar información sobre deportes
- Buscar información sobre juegos
- Buscar información sobre música o cine
- Buscar información sobre otros temas
- Comunicarme con otras personas

pudiendo responder Nunca o casi nunca, Una o dos veces al mes, Una o dos veces a la semana, Todos o casi todos los días, o No contesta a cada opción.

### **Relación con los padres:**

- Relación con tu madre y con tu padre:
  - Mantenemos una buena relación y me comprende
  - Se preocupa por mis hábitos de trabajo y estudio
  - Me anima a estudiar
  - Me pregunta si tengo deberes
  - Comprueba que hago los deberes
  - Me ayuda a hacer los deberes y trabajos
  - Me pregunta cómo me ha ido en clase
  - Se preocupa por mis resultados y calificaciones escolares
  - Asiste a reuniones, entrevistas y fiestas del centro
  - Se preocupa por mi comportamiento y mi relación con los compañeros
  - Me acompaña cuando hago actividades de tiempo libre (deportes, música...)

pudiendo responder Nunca o casi nunca, Algunas veces, Casi siempre, Siempre, o No contesta a cada opción (para la madre y para el padre).

### **Resultados en las pruebas:**

- Índice Social, Económico y Cultural (**ISEC**): con media=1 y desviación típica con valor de 1, sirve “para medir diversos aspectos del origen social y familiar de los alumnos [...], refleja la ocupación profesional y el nivel educativo de los padres, así como los recursos disponibles en el hogar.” (INEE). En esta base de datos su valor oscila entre -3,22 y 1,82.
- **Puntuación Lingüística**: que viene dividida en cinco apartados.
- **Puntuación Matemática**: que viene dividida en cinco apartados.
- **Puntuación Ciencias y Mundo Físico**: que viene dividida en cinco apartados.
- **Puntuación Ciencias Social y Ciudadana**: que viene dividida en cinco apartados.

Como se puede comprobar, esta base de datos nos permite realizar innumerables estudios para intentar explicar de qué depende el R.A.. En el tema que nos atañe, la influencia del entorno familiar, hay un gran número de variables que nos van a servir para comprobar la validez de las hipótesis de investigación, como se explicará en el siguiente apartado.

## 6. Análisis de la muestra:

Comprobación del posible cumplimiento de las hipótesis de investigación.

### Tratamiento de la base de datos

En primer lugar, tendremos que modificar la base de datos para facilitar su comprensión y el trabajo con la misma, para ello, se ha hecho uso de la hoja de cálculo Excel. En esta primera etapa, el trabajo se ha fundamentado en nombrar directamente tanto las variables como los posibles valores que éstas puedan mostrar. En la base de datos (B.D. a partir de ahora), nos encontramos los valores y las variables según se puede observar en los gráficos 3, 4 y 5.

Gráfico 3:

S6M	S6P	S7M	S7P	S8	S10
4	4	5	1	9	6
8	8	2	1	9	6
8	8	5	1	99	7
8	8	1	1	5	7
8	8	5	1	99	6
6	6	1	1	5	7
8	8	1	1	5	6
7	6	5	1	9	6
6	8	1	1	9	6
8	6	5	1	99	8
8	8	1	1	6	6
8	8	5	1	9	6

Fuente: INEE. (2017).

Gráfico 4:

S6M	Estudios de la madre
S6P	Estudios del padre
S7M	Situación laboral madre
S7P	Situación laboral padre
S8	Trabajo de tu madre
S10	Trabajo de tu padre

Fuente: INEE. (2017).

Gráfico 5:

S6P	1	Universitarios superiores
	2	Universitarios medios
	3	Técnico superior de FP
	4	Bachillerato
	5	Técnico FP grado medio
	6	Estudios obligatorios (ESO,EGB)
	7	No completó estudios obligatorios
	8	No lo sé
	9 <sup>a</sup>	No contesta
S7M	1	Trabaja fuera de casa
	2	En casa de forma remunerada
	3	Está en paro
	4	Está jubilado
	5	Sólo las tareas de casa
	6	No lo sé
	9 <sup>a</sup>	No contesta

Fuente: INEE. (2017).

De esta forma, la B.D. se dividía en tres hojas, en la primera, de la que se ha extraído el Gráfico 3, se mostraban las variables y los valores que tomaba cada observación. La segunda mostraba la codificación que se había seguido para representar cada variable con números y letras, como podemos observar en el Gráfico 4. Y por último, la tercera hoja indicaba a qué valor representaba cada número, en caso de que dichos valores perteneciesen a variables categóricas (como ocurre en la mayoría de los casos).

Una vez sabiendo de qué forma se “codifica” la B.D., haciendo uso de fórmulas condicionales, BUSCARV, sustituir, y pegando como valores (y no como fórmulas), se ha logrado que ésta muestre un aspecto mucho más simple e intuitivo, como queda reflejado en el Gráfico 5.

Gráfico 5:

TITULAR	SEXO	ESTUDIOS MADRE	ESTUDIOS PADRE
PUBLICO	CHICO	Estudios obligatorios (ESO,EGB)	No lo sé
PUBLICO	CHICO	No lo sé	Estudios obligatorios (ESO,EGB)
PUBLICO	CHICO	No lo sé	No lo sé
PUBLICO	CHICO	No lo sé	No lo sé
PUBLICO	CHICA	Técnico FP grado medio	Técnico FP grado medio
PUBLICO	CHICO	Técnico superior de FP	Estudios obligatorios (ESO,EGB)
PUBLICO	CHICA	No completó estudios obligatorios	No completó estudios obligatorios
PUBLICO	CHICO	Técnico FP grado medio	Técnico FP grado medio
PUBLICO	CHICA	Técnico superior de FP	Universitarios medios
PUBLICO	CHICA	Estudios obligatorios (ESO,EGB)	Universitarios medios
PUBLICO	CHICO	Estudios obligatorios (ESO,EGB)	Estudios obligatorios (ESO,EGB)
PUBLICO	CHICA	No completó estudios obligatorios	No completó estudios obligatorios
PUBLICO	CHICA	No lo sé	No lo sé
PUBLICO	CHICO	Estudios obligatorios (ESO,EGB)	Estudios obligatorios (ESO,EGB)

Fuente: elaboración propia.

Como se manifiesta en este último gráfico, se ve de forma más intuitiva cada variable y cada valor. A pesar de ello, se sigue presentando la cuestión de los

valores “no encontrados”; es decir, aquellos que representan cuando no se ha respondido a una pregunta referente a cualquier variable o cuando no se ha realizado una de las cuatro pruebas. Sin embargo, este inconveniente lo resolveremos haciendo uso del programa estadístico Rstudio.

A continuación, cargamos la B.D. en el programa Rstudio, por lo que ya tenemos la B.D. disponible para trabajar con ella en el mencionado program (en el “Anexo II: Código empleado para el análisis en Rstudio”, queda plasmado todo el código ejecutado para el trabajo en su conjunto). Sin embargo, seguimos teniendo varias observaciones que contienen valores no disponibles, representados como “NA”. Para ello suprimimos todas las observaciones que presenten este tipo de valor en cualquier variable. En casos similares, hay que observar de qué forma se ve mermada la muestra al retirar este tipo de observaciones. En nuestro caso, contábamos con 29.253 observaciones, que tras ejecutar la mencionada función, se quedan en 26.039. Como podemos advertir, se desechan casi un 11% de las observaciones, lo que no es para nada insignificante. A pesar de ello, cabe destacar que la investigación habría sido mucho más compleja en caso de contar con tantos valores no disponibles en diferentes variables, por lo que, en aras de simplificar el trabajo, se ha decidido prescindir de tales observaciones.

Una vez tratada la B.D. y “depurada” de casos con valores no disponibles que entorpezcan y dificulten el estudio, trataremos cada una de las seis hipótesis por separado. Explicaremos brevemente qué variables tendremos en cuenta para cada contrastar hipótesis, y posteriormente comprobaremos su veracidad.

## Comprobación hipótesis de investigación

En este apartado trataremos cada hipótesis por separado, de cara a facilitar su explicación y poder establecer las futuras conclusiones. Analizaremos las seis hipótesis de investigación en el mismo orden en que éstas fueron planteadas en apartados anteriores, y finalizaremos el trabajo mediante la recopilación de las conclusiones que habremos extraído tras el análisis.

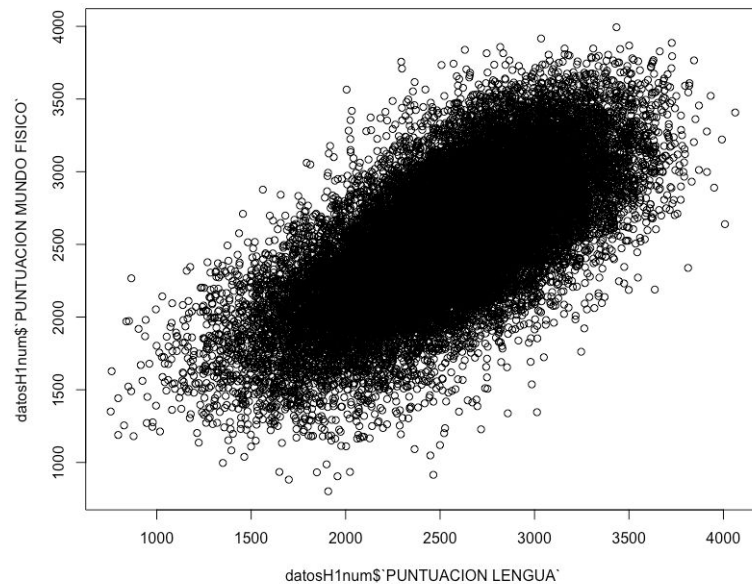
Cabe destacar que, dado que todas las hipótesis de investigación tratan el R.A., este será evaluado en función de los resultados que los alumnos obtengan en las cuatro pruebas de conocimientos. Sin embargo, para no tener que realizar cada análisis en función de cada una de las cuatro, se ha optado por reunir en una nueva variable todos los resultados. El motivo principal de ello es simplificar el estudio, sin embargo, el hecho que sustenta que se puedan unir los cuatro resultados sin pedir información es el siguiente:

```
> cor(as.matrix(coentrepruebas))
```

	PUNTUACION LENGUA	PUNTUACION MATEMATICA	PUNTUACION MUNDO FISICO	PUNTUACION SOCIAL Y CIUDADANA
PUNTUACION LENGUA	1.0000000	0.5396081	0.6840541	0.6680542
PUNTUACION MATEMATICA	0.5396081	1.0000000	0.5923228	0.5124827
PUNTUACION MUNDO FISICO	0.6840541	0.5923228	1.0000000	0.6406707
PUNTUACION SOCIAL Y CIUDADANA	0.6680542	0.5124827	0.6406707	1.0000000

Como podemos comprobar, la correlación entre las pruebas es media-alta, por lo que es esperable que a medida que se obtiene un mayor resultado en una prueba, este hecho se replique en el resto. De forma gráfica y para facilitar la comprensión, el gráfico 7 nos muestra de qué forma se correlacionan las puntuaciones de Lengua y Mundo Físico.

Gráfico 7: Diagrama de dispersión Puntuación Lengua - Puntuación Mundo Físico



Fuente: elaboración propia.

Por ello, una nueva variable llamada “Suma pruebas” será el resultado de sumar los cuatro resultados, y ésta será sobre la que trabajemos a la hora de evaluar el R.A. de cada alumno.

A continuación, pasaremos a comprobar la validez y veracidad de cada una de las seis hipótesis básicas centrales del presente trabajo:

- 1. Los alumnos cuyos padres tengan más estudios, desempeñan un mejor rendimiento académico.*
- 2. Los alumnos que acuden a un centro privado, desempeñan un mejor rendimiento académico.*
- 3. Los alumnos que gozan de un mayor apoyo familiar en referencia a los estudios, desempeñan un mejor rendimiento académico.*
- 4. A mayor tiempo dedicado a los deberes, los alumnos muestran un mejor desempeño académico.*
- 5. Los alumnos que históricamente presentaban un peor R.A., desempeñan un peor rendimiento académico.*
- 6. A mayor nivel socioeconómico y cultural, los alumnos desempeñan un mejor rendimiento académico..*

## ***1. Los alumnos cuyos padres tengan más estudios, desempeñan un mejor rendimiento académico.***

En esta primera hipótesis, evaluaremos la variable Estudios (del padre y de la madre) y la compararemos con la suma de las cuatro pruebas (S.C.P. en adelante), que representa el R.A.. Lo que evaluaremos será si la media de la S.C.P. es mayor para aquellos alumnos cuyos padres tienen más estudios. Debemos ser conscientes que mientras que la variable Suma de las pruebas es numérica, la variable Estudios es categórica. Esta última puede tomar los valores:

- 'No lo sé',
- 'No completó estudios obligatorios',
- 'Estudios obligatorios (ESO,EGB)',
- 'Técnico FP grado medio',
- 'Bachillerato',
- 'Técnico superior de FP',
- 'Universitarios medios' o
- 'Universitarios superiores'.

Sin embargo, antes de continuar, efectuamos la modificación de retirar todas aquellas observaciones que tomen como valor “No lo sé” en la variable Estudios. Esto es llevado a cabo debido a que las observaciones de este tipo, no sólo no aportan valor e información al estudio, sino que también dificultan su realización.

A continuación, para esquematizar los gráficos y clasificar de alguna forma los valores, lo hemos hecho en niveles, de forma que cada uno de los siguientes números corresponde al nivel del valor:

- 1: No completó estudios obligatorios
- 2: Estudios obligatorios (ESO,EGB)
- 3: Técnico FP grado medio
- 4: Bachillerato
- 5: Técnico superior de FP
- 6: Universitarios medios
- 7: Universitarios superiores

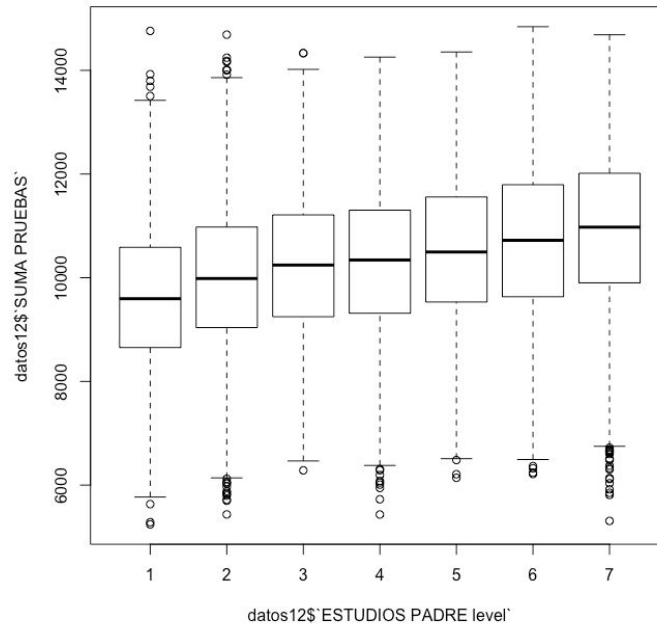
Debido a esta “transformación” en niveles, aquellos alumnos cuyo valor fuese “No contesta”, pasa a ser un valor nulo o NA, el cual eliminamos de nuevo. En



consecuencia, tenemos dos poblaciones con la variable estudios (una de la madre, y la otra del padre) y la S.C.P..

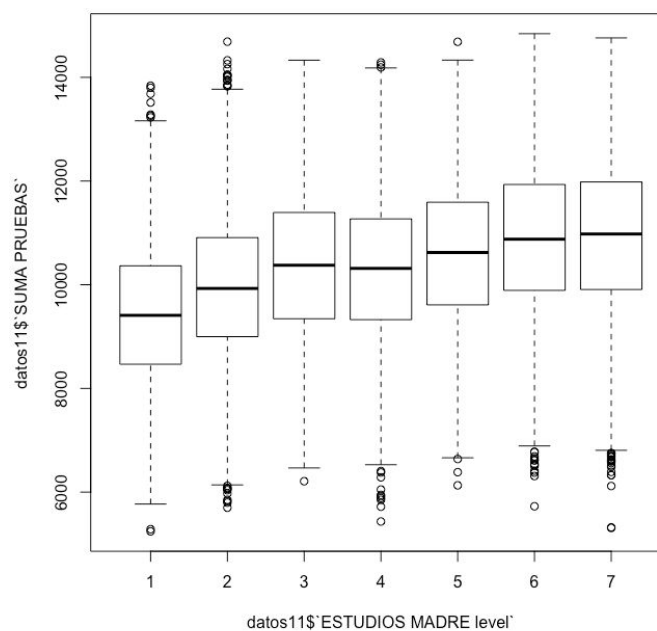
De forma gráfica y teniendo presente a qué nivel de estudios corresponde cada nivel (expuesto en el párrafo anterior), en los gráficos 8 y 9 podemos ver de qué forma se distribuye la S.C.P. en función del nivel de estudios del padre y de la madre:

Gráfico 8: Resultado S.C.P. en función del nivel de estudios del padre.



Fuente: elaboración propia.

Gráfico 9: Resultado S.C.P. en función del nivel de estudios de la madre.



Fuente: elaboración propia.

En primera instancia y de forma intuitiva, podríamos llegar a afirmar que a mayor nivel de estudios de los padres, mejor R.A.. Sin embargo esta afirmación carecería de validez, ya que para ello debe sustentarse en datos empíricos y numéricos, por ello llevaremos a cabo un Análisis de la Varianza, ampliamente conocido como ANOVA.

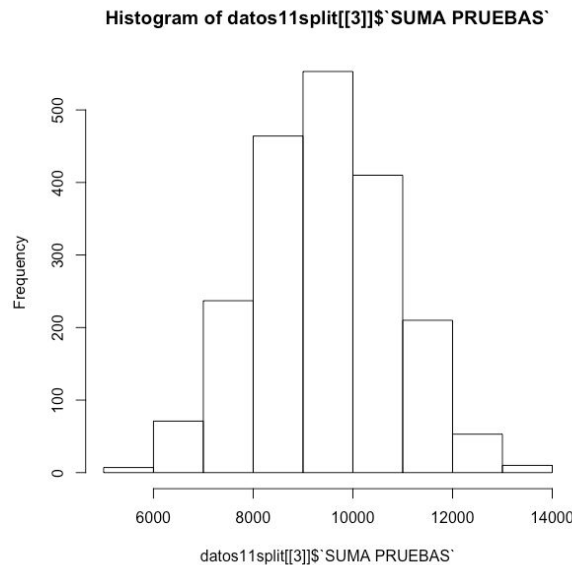
### **ANOVA: Estudios de la madre y S.C.P.**

El Análisis de la Varianza es sumamente útil a la hora de realizar un contraste de igualdad de medias para más de dos medias (aunque también se pueda utilizar así). Este análisis contrasta la igualdad de tantas medias como poblaciones haya, es decir, comprueba de forma estadística si éstas tienen la misma media. Su hipótesis básica es que todas las medias son iguales, mientras que la alternativa es que al menos una de las medias no es similar al resto. Como el ANOVA sólo nos concede la información referente a si todas las medias son estadísticamente similares o no; para saber de qué forma difieren emplearemos el Test de Tukey, que explicaremos posteriormente. De forma práctica, analizaremos en primer lugar el R.A. en función de los estudios de la madre, y posteriormente en función de los estudios del padre.

Para llevar a cabo el ANOVA, deben cumplirse unas hipótesis básicas, de forma que si esto sucede, los resultados del análisis serán vinculantes y veraces. Estas hipótesis son las siguientes:

- **Independencia:** la cual se cumple debido a que cada observación es independiente del resto, al tratarse de alumnos con desempeños académicos y padres distintos.
- **Normalidad:** para que se pueda desarrollar el estudio correctamente, cada población (en nuestro caso cada grupo de alumnos cuyos padres tengan el mismo nivel de estudios) deben distribuirse siguiendo una distribución normal. Podemos observar en el gráfico 10 cómo la S.C.P. de los alumnos cuyos padres no completaron los estudios parece que sigue una distribución normal, al menos en cuanto a la forma que presenta el histograma de frecuencias.

Gráfico 10: Histograma frecuencias S.C.P. con padres que no completaron estudios.



Fuente: elaboración propia.

Sin embargo, el hecho de que “parezca” cumplir la hipótesis de normalidad no lo garantiza, ya que debemos comprobarlo estadísticamente. Para ello llevamos a cabo un test de normalidad de Anderson-Darling, en el que la hipótesis nula es que la población cumple la hipótesis de normalidad.

```
> ad.test(datos11split[[3]]$`SUMA PRUEBAS`)
```

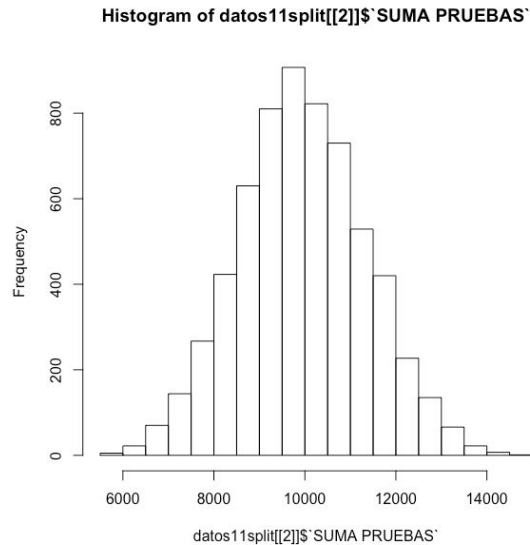
```
Anderson-Darling normality test
```

```
data: datos11split[[3]]$`SUMA PRUEBAS`  
A = 0.49936, p-value = 0.2092
```

Como podemos observar, al ser el P-valor mucho mayor que 0,01 (en el caso de que ese sea nuestro nivel de significación), por lo que efectivamente la S.C.P en alumnos cuyos padres no completaron los estudios, sí cumple la hipótesis de normalidad.

En cambio, en el caso de los alumnos cuyos padres completaron los estudios obligatorios, como podemos ver claramente en el gráfico 11, también parece seguir la S.C.P. una distribución normal.

Gráfico 11: Histograma frecuencias S.C.P. con padres que completaron los estudios obligatorios.



Fuente: elaboración propia.

Mientras que, a la hora de comprobarlo estadísticamente, comprobamos que no cumple la hipótesis de normalidad.

```
> ad.test(datos11split[[2]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos11split[[2]]$`SUMA PRUEBAS`
A = 1.1993, p-value = 0.003978
```

Y lo mismo ocurre con el resto de variables, mientras que gráficamente parecen seguir una distribución normal, algunas de ellas no lo cumplen cuando lo comprobamos con el test de Anderson-Darling.

A pesar de la ausencia de normalidad en base al test de Anderson-Darling, podríamos hacer uso del Teorema Central del Límite, para así suponer la normalidad de las poblaciones. Sin embargo, tanto en esta hipótesis como en las siguientes, en el caso de haber ausencia de normalidad, aplicaremos la característica del ANOVA que se explica a continuación.

Afortunadamente, el ANOVA es un modelo robusto frente a las desviaciones de la normalidad cuando las poblaciones son de tamaños similares. Este hecho significa que, en el caso de que las poblaciones tengan el mismo número de observaciones, se puede proseguir con el análisis a pesar de no cumplirse la hipótesis de normalidad (recordemos que, gracias al TCL, sí podríamos suponer con base racional y estadística el cumplimiento de la

hipótesis de normalidad, a pesar de ello, trabajaremos los ANOVAs como si no fuese así). Como se puede observar, a continuación se muestra el tamaño poblacional para cada valor de “estudios de la madre”:

▶ Bachillerato	list [3653 x 3]
▶ Estudios obligatorios (ESO,EGB)	list [6237 x 3]
▶ No completó estudios obligatorios	list [2015 x 3]
▶ Técnico FP grado medio	list [1023 x 3]
▶ Técnico superior de FP	list [1408 x 3]
▶ Universitarios medios	list [2151 x 3]
▶ Universitarios superiores	list [2980 x 3]

Las poblaciones tienen tamaños muy dispares, por lo que para conseguir homogeneizar el tamaño, realizamos un muestreo aleatorio sin reemplazo en el que extraemos 1.000 observaciones de cada lista, de forma que trabajemos con siete listas de igual tamaño, cada una correspondiente a un valor que pueda tomar la variable “estudios de la madre”. A continuación las unimos en una B.D. nueva, que será el que emplearemos para extraer los resultados del ANOVA. Esta nueva base tiene tamaño 7.000, y es propicia para realizar el análisis, gracias a la robustez del mismo.

- **Homocedasticidad:** las varianzas poblacionales deben ser homogéneas. Para comprobarlo, realizamos una prueba o test de Levene, que nos indicará si las varianzas poblacionales son similares o no. La hipótesis nula en este caso es que las varianzas son homogéneas.

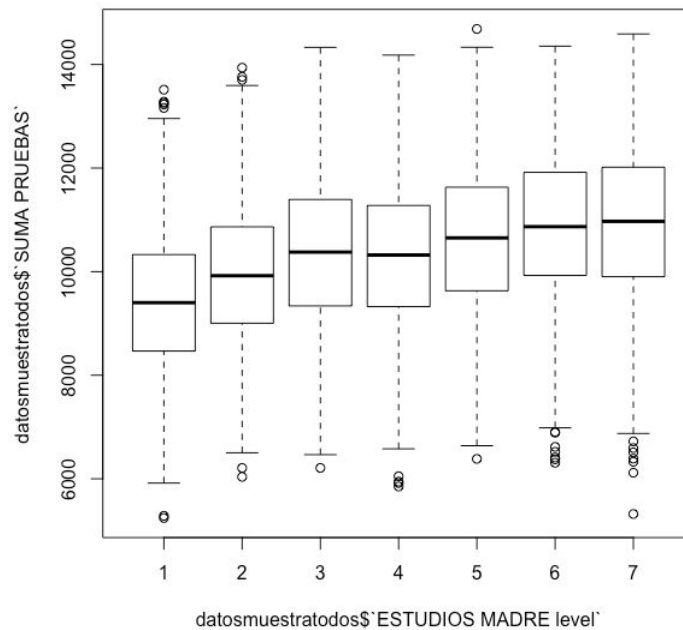
```
> leveneTest(datosmuestratodos$`SUMA PRUEBAS`,datosmuestratodos$`ESTUDIOS MADRE level`)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  6  2.0677 0.05364 .
      6993
```

Como se nos indica, en este caso el P-valor es 0,054. Empleando el mismo nivel de significación que en el caso anterior, 0,01, no podemos afirmar la hipótesis alternativa, por lo que se cumple la hipótesis nula. En otras palabras, se cumple la hipótesis básica de homocedasticidad.

Una vez siendo conscientes del cumplimiento de las hipótesis básicas, podemos llevar a cabo el ANOVA, que nos indicará si hay diferencias significativas entre las medias de cada población, correspondientes a cada valor que tome la variable

“estudios de la madre”. En el gráfico 12, vemos de qué forma se presentan los diagramas de caja para cada valor.

Gráfico 12: Resultado S.C.P. en función del nivel de estudios de la madre.



Fuente: elaboración propia.

El gráfico 12 guarda una gran similitud con el gráfico 9, sin embargo, el 12 hace referencia a la B.D. mencionada anteriormente, en la que se han tomado 1.000 observaciones de cada valor. De cara a tenerlo presente, mostramos de nuevo la “codificación” de la variable estudios, para poder visualizar de forma más rápida tanto el gráfico inmediatamente anterior como los posteriores:

- 1: No completó estudios obligatorios
- 2: Estudios obligatorios (ESO,EGB)
- 3: Técnico FP grado medio
- 4: Bachillerato
- 5: Técnico superior de FP
- 6: Universitarios medios
- 7: Universitarios superiores

Empleando la herramienta Rstudio, vemos los siguientes resultados del ANOVA, recordando que la hipótesis nula de la que partimos es que la media sea la misma para todas las poblaciones.

```
> anovaestmadre<-aov(datosmuestratodos$`SUMA PRUEBAS`~datosmuestratodos$`ESTUDIOS MADRE`)
> summary(anovaestmadre)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
datosmuestratodos\$`ESTUDIOS MADRE`	6	1.63e+09	271705093	133.8	<2e-16 ***
Residuals	6993	1.42e+10	2030328		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El P-valor es mucho menor que el nivel de significación de 0,01, por lo que podemos afirmar que las medias de la S.C.P. para los alumnos cuyos estudios de la madre sean distintos, es distinta. En otras palabras, el nivel de estudios de la madre implica diferente R.A..

Una vez conocido el hecho de que hay diferencias entre las medias, comprobaremos cómo son esas diferencias mediante el test de Tukey, que nos mostrará la forma en que difiere la media de la S.C.P. entre todos los valores que tome la variable “estudios de la madre”.

```
> TukeyHSD(anovaestmadre)
Tukey multiple comparisons of means
95% family-wise confidence level
```

```
Fit: aov(formula = datosmuestratodos$`SUMA PRUEBAS` ~ datosmuestratodos$`ESTUDIOS MADRE`)
```

```
$`datosmuestratodos$`ESTUDIOS MADRE``
```

	diff	lwr	upr	p adj
Estudios obligatorios (ESO,EGB)-Bachillerato	-323.4342	-511.36571	-135.5027	0.0000082
No completó estudios obligatorios-Bachillerato	-861.6202	-1049.55168	-673.6887	0.0000000
Técnico FP grado medio-Bachillerato	91.1365	-96.79501	279.0680	0.7855331
Técnico superior de FP-Bachillerato	325.3895	137.45803	513.3210	0.0000070
Universitarios medios-Bachillerato	559.7031	371.77164	747.6347	0.0000000
Universitarios superiores-Bachillerato	616.0074	428.07585	803.9389	0.0000000
No completó estudios obligatorios-Estudios obligatorios (ESO,EGB)	-538.1860	-726.11747	-350.2545	0.0000000
Técnico FP grado medio-Estudios obligatorios (ESO,EGB)	414.5707	226.63920	602.5022	0.0000000
Técnico superior de FP-Estudios obligatorios (ESO,EGB)	648.8237	460.89224	836.7552	0.0000000
Universitarios medios-Estudios obligatorios (ESO,EGB)	883.1374	695.20585	1071.0689	0.0000000
Universitarios superiores-Estudios obligatorios (ESO,EGB)	939.4416	751.51005	1127.3731	0.0000000
Técnico FP grado medio-No completó estudios obligatorios	952.7567	764.82516	1140.6882	0.0000000
Técnico superior de FP-No completó estudios obligatorios	1187.0097	999.07820	1374.9412	0.0000000
Universitarios medios-No completó estudios obligatorios	1421.3233	1233.39181	1609.2548	0.0000000
Universitarios superiores-No completó estudios obligatorios	1477.6275	1289.69602	1665.5590	0.0000000
Técnico superior de FP-Técnico FP grado medio	234.2530	46.32153	422.1845	0.0044670
Universitarios medios-Técnico FP grado medio	468.5667	280.63515	656.4982	0.0000000
Universitarios superiores-Técnico FP grado medio	524.8709	336.93935	712.8024	0.0000000
Universitarios medios-Técnico superior de FP	234.3136	46.38211	422.2451	0.0044511
Universitarios superiores-Técnico superior de FP	290.6178	102.68631	478.5493	0.0001057
Universitarios superiores-Universitarios medios	56.3042	-131.62730	244.2357	0.9750825

El test nos muestra a un nivel de confianza del 95% el intervalo en el que se hallará el valor medio de la variable S.C.P. en función de los estudios de la madre. Cabe destacar que este test nos indica si la diferencia entre cada media con el resto es significativa, por lo que nos encontramos (previsiblemente, dado el gráfico) que algunos valores no muestran diferencias significativas. Esto es entre las S.C.P. de cuyos estudiantes sus madres estudiaron técnico de FP de grado medio y

Bachillerato, así como para estudios universitarios superiores y universitarios medios. Podemos afirmarlo dado que su P-valor es mayor que el nivel de significación de 0,01, por lo que no se rechaza la hipótesis nula de que las medias son similares.

Mientras que las diferencias se nos muestran en términos absolutos, resulta interesante comprobar de qué forma varían en términos relativos. Por ello quedan reflejadas a continuación todas las medias.

Valor	Media S.C.P.
1: No completó estudios obligatorios	9418,77
2: Estudios obligatorios (ESO,EGB)	9956,96
3: Técnico FP grado medio	10371,53
4: Bachillerato	10280,39
5: Técnico superior de FP	10605,78
6: Universitarios medios	10840,10
7: Universitarios superiores	10896,40

Entonces, una vez siendo conscientes de las medias y las diferencias entre ellas, podemos sacar las conclusiones que se expresan a continuación (mostramos las más representativas). Debemos tener presente que mostramos la variación media de un intervalo de confianza del 95%, y que hacemos analogía entre la Suma del resultado de las Cuatro Pruebas (S.C.P.) y el Rendimiento Académico (R.A.). Asimismo, los porcentajes representan un término medio.

- Los alumnos cuyas madres no completaron los estudios obligatorios presentan un R.A. medio del 8,4% menor que aquellos cuyas madres tienen como última titulación el bachillerato.
- Respecto a los alumnos cuyas madres únicamente poseen los estudios obligatorios (ESO o EGB);
  - los alumnos cuyas madres tienen estudios universitarios medios desempeñan un R.A. medio del 8,9% mayor.
  - los que sus madres tienen estudios universitarios superiores tienen un R.A. medio del 9,4% mayor.



- En referencia a los alumnos cuyas madres no completaron los estudios obligatorios;
  - los que sus madres tienen el título de técnico FP de grado medio, tienen un R.A. medio del 10,1% mayor.
  - aquellos que sus madres tienen el título de técnico superior de FP, presentan un R.A. medio del 12,6% mayor.
  - los que sus madres gozan de estudios universitarios medios, tienen un R.A. medio del 15,1% mayor.
  - los alumnos cuyas madres tienen estudios universitarios superiores, presentan un R.A. medio del 15,7% mayor.

De esta forma se muestran las mayores diferencias de R.A. en función de los estudios de la madre del estudiante. En otras palabras, podríamos concluir con la siguiente afirmación:

***Si establecemos los siguientes cinco niveles en función de los estudios que tenga la madre del estudiante:***

- 1: no completó estudios obligatorios,***
- 2: estudios obligatorios (ESO,EGB),***
- 3: técnico FP grado medio y Bachillerato,***
- 4: técnico superior de FP, y***
- 5: universitarios medios y universitarios superiores.***

***Queda probado que, a medida que se aumenta de nivel, el rendimiento académico medio del alumno medio aumenta de forma significativa, llegando a haber una diferencia de más de 15 puntos porcentuales entre el primer y el último nivel.***

## ANOVA: Estudios del padre y S.C.P.

En este caso realizaremos el mismo análisis, pero utilizando la variable estudios del padre (y, por supuesto, la S.C.P.). Para realizar, el ANOVA, en primer lugar comprobaremos el cumplimiento de las tres hipótesis básicas:

- **Independencia:** la cual se cumple dado que cada observación no depende del resto.
- **Normalidad:** realizamos el test de Anderson-Darling, cuya hipótesis nula es la normalidad de la población en cuestión a evaluar.

```
> ad.test(datos12split[[3]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos12split[[3]]$`SUMA PRUEBAS`  
A = 0.28881, p-value = 0.6159
```

```
> ad.test(datos12split[[4]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos12split[[4]]$`SUMA PRUEBAS`  
A = 0.34752, p-value = 0.4784
```

En estos dos primeros casos, que hacen referencia a las observaciones en las que la variable “estudios del padre” tome los valores “no completó los estudios obligatorios” y “técnico FP de grado medio”; podemos ver cómo el P-valor toma valores mucho mayores que el nivel de significación de 0,01. Este hecho significa que en estos casos, se cumple la hipótesis de normalidad. Sin embargo no ocurre lo mismo en todos, como podemos ver a continuación.

```
> ad.test(datos12split[[6]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos12split[[6]]$`SUMA PRUEBAS`  
A = 2.158, p-value = 1.765e-05
```

```
> ad.test(datos12split[[7]]$`SUMA PRUEBAS`)
```

Anderson-Darling normality test

```
data: datos12split[[7]]$`SUMA PRUEBAS`  
A = 7.5136, p-value < 2.2e-16
```

En estos dos casos, referentes a la S.C.P. de los alumnos cuyos padres tiene estudios universitarios medios y superiores, respectivamente, se niega la hipótesis nula. Por ende, la hipótesis de normalidad no se cumple para todos los valores.

Por ello hacemos uso de la propiedad de robustez que caracteriza al ANOVA frente a las desviaciones de normalidad, en los casos en los que las poblaciones tomen tamaños similares. En este caso, observamos los siguientes tamaños poblacionales:

Bachillerato	list [2832
Estudios obligatorio...	list [5817
No completó estudio...	list [2269
Técnico FP grado me...	list [1116
Técnico superior de FP	list [1685
Universitarios medios	list [1846
Universitarios superi...	list [3205

Por lo que tomamos muestras sin reemplazamiento de tamaño 1000 observaciones, para seguir trabajando con el análisis y que los resultados sean estadísticamente veraces.

- **Homocedasticidad:** en cuanto a la homogeneidad de varianzas poblacionales, el test de Levene lo probará, teniendo en cuenta que la hipótesis nula es que las varianzas poblacionales son similares.

```
> leveneTest(datosmuestra12$`SUMA PRUEBAS`,datosmuestra12$`ESTUDIOS PADRE level`)  
Levene's Test for Homogeneity of Variance (center = median)  
  Df F value Pr(>F)  
group  6  1.0436 0.3947  
6993
```

Al ser el P-valor mucho mayor que el nivel de significación de 0,01, podemos seguir trabajando con el ANOVA, siendo conscientes de que se cumplen las hipótesis básicas.

Introducimos los comandos pertinentes en Rstudio, y obtenemos los siguientes resultados:

```
> anovaestpadre<-aov(datosmuestra12$`SUMA PRUEBAS`~datosmuestra12$`ESTUDIOS PADRE`)
> summary(anovaestpadre)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
datosmuestra12\$`ESTUDIOS PADRE`	6	9.977e+08	166286218	78.19	<2e-16 ***
Residuals	6993	1.487e+10	2126732		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Que prueban que hay diferencias significativas entre las medias de la S.C.P. de los alumnos para cada nivel educativo de sus padres. Sin embargo, hacemos uso del test de Tukey para ser conscientes de qué valores son los que difieren y cuáles no presentan diferencias significativas entre sí.

```
> TukeyHSD(anovaestpadre)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = datosmuestra12$`SUMA PRUEBAS` ~ datosmuestra12$`ESTUDIOS PADRE`)

$`datosmuestra12$`ESTUDIOS PADRE``
              diff             lwr             upr             p adj
Estudios obligatorios (ESO,EGB)-Bachillerato -287.01264 -479.35407 -94.67122 0.0002211
No completó estudios obligatorios-Bachillerato -606.03039 -798.37181 -413.68896 0.0000000
Técnico FP grado medio-Bachillerato -44.54732 -236.88874 147.79411 0.9935456
Técnico superior de FP-Bachillerato 157.87844 -34.46299 350.21986 0.1898799
Universitarios medios-Bachillerato 420.44429 228.10286 612.78572 0.0000000
Universitarios superiores-Bachillerato 593.48244 401.14102 785.82387 0.0000000
No completó estudios obligatorios-Estudios obligatorios (ESO,EGB) -319.01775 -511.35917 -126.67632 0.0002211
Técnico FP grado medio-Estudios obligatorios (ESO,EGB) 242.46533 50.12390 434.80675 0.0038175
Técnico superior de FP-Estudios obligatorios (ESO,EGB) 444.89108 252.54966 637.23251 0.0000000
Universitarios medios-Estudios obligatorios (ESO,EGB) 707.45693 515.11551 899.79836 0.0000000
Universitarios superiores-Estudios obligatorios (ESO,EGB) 880.49509 688.15366 1072.83651 0.0000000
Técnico FP grado medio-No completó estudios obligatorios 561.48307 369.14165 753.82450 0.0000000
Técnico superior de FP-No completó estudios obligatorios 763.90883 571.56740 956.25025 0.0000000
Universitarios medios-No completó estudios obligatorios 1026.47468 834.13325 1218.81610 0.0000000
Universitarios superiores-No completó estudios obligatorios 1199.51283 1007.17141 1391.85426 0.0000000
Técnico superior de FP-Técnico FP grado medio 202.42576 10.08433 394.76718 0.0315658
Universitarios medios-Técnico FP grado medio 464.99161 272.65018 657.33303 0.0000000
Universitarios superiores-Técnico FP grado medio 638.02976 445.68833 830.37119 0.0000000
Universitarios medios-Técnico superior de FP 262.56585 70.22443 454.90728 0.0011235
Universitarios superiores-Técnico superior de FP 435.60400 243.26258 627.94543 0.0000000
Universitarios superiores-Universitarios medios 173.03815 -19.30327 365.37958 0.1106125
```

El test de Tukey nos muestra que, con un nivel de significación de 0,01; no hay diferencias significativas entre el R.A. de los estudiantes cuyos padres tengan estudios:

- Técnico FP de grado medio, técnico superior de FP o Bachillerato.
- Universitarios medios o superiores.

Y en cuanto a las diferencias en términos relativos, nos basamos en la siguiente tabla que muestra las medias de R.A. o S.C.P. para cada valor que tome la variable “estudios del padre”:

Valor	Media S.C.P.
No completó estudios obligatorios	9692,68
Estudios obligatorios (ESO,EGB)	10011,70
Técnico FP grado medio	10254,16
Bachillerato	10298,71
Técnico superior de FP	10456,59
Universitarios medios	10719,15
Universitarios superiores	10892,19

En consecuencia, y basándonos en la premisa de que trataremos el resultado de la S.C.P como el R.A., podemos extraer la siguiente conclusión:

***Si establecemos los siguientes cinco niveles en función de los estudios que tenga el padre del estudiante:***

- 1: no completó estudios obligatorios,***
- 2: estudios obligatorios (ESO,EGB),***
- 3: técnico FP grado medio, Bachillerato y técnico superior de FP, y***
- 4: universitarios medios y universitarios superiores.***

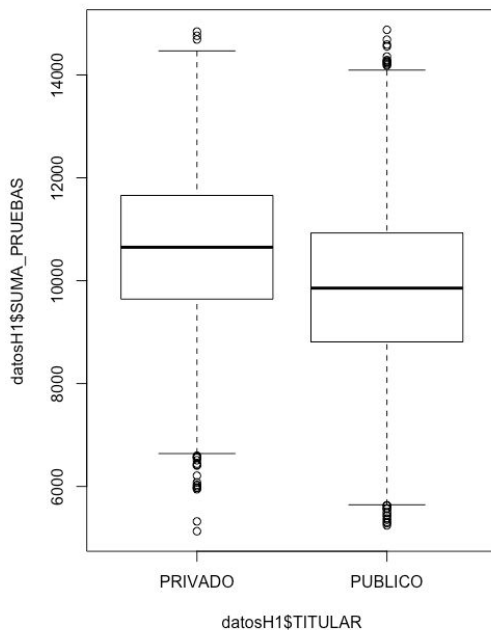
***Queda probado que, a medida que se aumenta de nivel, el rendimiento académico medio del alumno medio aumenta de forma significativa, llegando a haber una diferencia de más de 10 puntos porcentuales entre el primer y el último nivel.***

## 2. Los alumnos que acuden a un centro privado, desempeñan un mejor rendimiento académico.

En este caso, compararemos la S.C.P. de aquellos alumnos que acuden a un centro público y los que acuden a uno de titularidad privada. Para ello haremos uso de la mencionada variable, así como la variable titularidad, de carácter dicotómico, que nos indica si el centro es público o privado. Al tener que comparar el comportamiento de una variable en función de otra categórica, el ANOVA particularizado a dos grupos es el análisis que mejor nos va a ayudar a probar la hipótesis.

Antes de comenzar con el estudio, podemos observar cómo en el gráfico 13 se percibe que la variable S.C.P. tiene un valor mediano (y medio también, dado el comportamiento simétrico de la variable, que indica que media y mediana se situarán en valores muy próximos) mayor para los alumnos cuyo centro es de titularidad privada.

Gráfico 13: Resultado S.C.P. en función de la titularidad del centro.



Fuente: elaboración propia.

Gracias a este gráfico, en el caso de que no se cumpla la hipótesis nula del ANOVA, que indica que no hay diferencias significativas entre las medias, podemos suponer que la media de la S.C.P. será mayor en los alumnos que acuden a un centro privado.

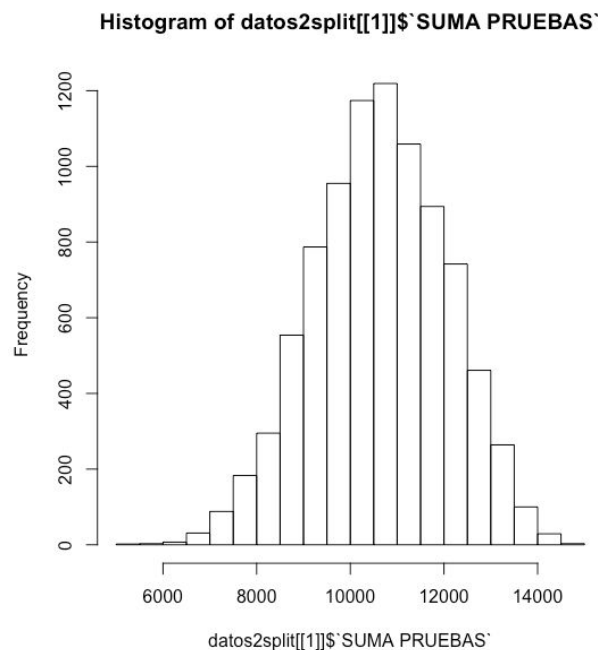
## ANOVA: Titularidad del centro y S.C.P.

Como se ha procedido en casos anteriores, comenzaremos por probar las hipótesis básicas, en pos de hallar unos resultados veraces estadísticamente hablando.

- **Independencia:** al tratarse de observaciones independientes unas de otras, se cumple.
- **Normalidad:** en este caso vuelve a ocurrir que, mientras que en términos gráficos parecen seguir una distribución normal, el test de Anderson-Darling nos indica que en realidad no es así.

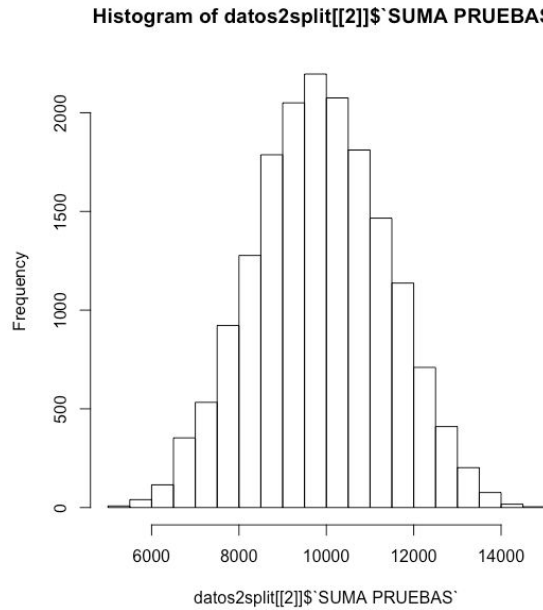
En los gráficos 14 y 15 podemos observar cómo los histogramas de frecuencias parecen seguir la antes mencionada distribución normal.

Gráfico 14: Histograma frecuencias S.C.P. alumnos que acuden a centro privado.



Fuente: elaboración propia.

Gráfico 15: Histograma frecuencias S.C.P. alumnos que acuden a centro público.



Fuente: elaboración propia.

Sin embargo, en el momento que realizamos el test de Anderson-Darling, cuya hipótesis nula indica que la variable sigue, una distribución normal, obtenemos los siguientes resultados:

```
> ad.test(datos2split[[1]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos2split[[1]]$`SUMA PRUEBAS`  
A = 3.2011, p-value = 5.096e-08
```

```
> ad.test(datos2split[[2]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos2split[[2]]$`SUMA PRUEBAS`  
A = 3.6728, p-value = 3.665e-09
```

Sabiendo que el primero corresponde a los resultados de la S.C.P. de los alumnos que acuden a un centro de titularidad privada, y el segundo a los que van a otro de titularidad pública. En ambos casos comprobamos que se niega la hipótesis nula de normalidad, dado que el P-valor es mucho menor que el nivel de significación de 0,01 que utilizaremos en esta segunda hipótesis del trabajo.



Por ello haremos uso de nuevo de la característica que posee esta clase de análisis, y es que es una técnica robusta contra las desviaciones de la normalidad, para aquellos casos en los que los tamaños poblacionales sean muy similares. En base a este hecho, y siendo conscientes del número de observaciones cuyo centro toma cada uno de los dos valores, público o privado; extraeremos muestras similares para crear un nuevo D.B.

▶ PRIVADO	list [8850 x 2]
▶ PUBLICO	list [17189 x 2]

Al tener estas dos listas, cada una correspondiente a cada valor de la variable titularidad, los tamaños mostrados, realizamos un muestreo aleatorio sin reemplazo en el que extraemos 4.000 observaciones de cada lista. Posteriormente uniremos ambas listas y trabajaremos con el nuevo D.B., con características idóneas para trabajar con el ANOVA (al menos en lo referente a la hipótesis básica de normalidad).

- **Homocedasticidad:** en cuanto a las varianzas de cada población, estas deben ser homogéneas. En otras palabras, deben ser similares. Realizando el test de Levene, cuya hipótesis nula es que hay homocedasticidad, obtenemos los siguientes resultados:

```
> leveneTest(datosmuestra2$`SUMA PRUEBAS`,datosmuestra2$TITULAR)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  6.2085 0.01273 *
      7998
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los cuales nos indican que, a un nivel de significación de 0,01, no se puede negar la hipótesis nula, suponiendo así la homocedasticidad de las poblaciones.

Habiéndose cumplido las hipótesis básicas, podemos pasar a realizar el análisis. Este nos indicará si se acepta la hipótesis nula de que no hay diferencias significativas entre las medias de la S.C.P. para los alumnos que acuden a centros de cada tipo de titularidad; o por el contrario encontramos diferencias significativas.

Por ende, ya estamos capacitados para llevar a cabo el ANOVA de forma estadísticamente veraz:

```
> anovaestmadre<-aov(datosmuestratodos$`SUMA PRUEBAS`~datosmuestratodos$`ESTUDIOS MADRE`)
> summary(anovaestmadre)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
datosmuestratodos\$`ESTUDIOS MADRE`	6	1.63e+09	271705093	133.8	<2e-16 ***
Residuals	6993	1.42e+10	2030328		

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Como queda plasmado, el P-valor es mucho menor que el nivel de significación de 0,01, pudiéndose afirmar que hay diferencias significativas entre las medias de los resultados de la S.C.P. entre aquellos alumnos que acuden a un centro de titularidad privada y los que acuden a uno de titularidad pública.

A continuación, mediante la realización en Rstudio del test de Tukey, veremos cómo es esa diferencia entre medias.

```
> TukeyHSD(anovaititularidad)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = datosmuestra2$`SUMA PRUEBAS` ~ datosmuestra2$TITULAR)

$`datosmuestra2$TITULAR`
      diff      lwr      upr p adj
PUBLICO-PRIVADO -804.0309 -867.7617 -740.3001 0
```

Este nos indica que, a un nivel de confianza del 95%, el valor medio de la variable S.C.P. en función de la titularidad del centro disminuirá entre 740 y 867 puntos si el centro es público en vez de privado. Esta conclusión es acorde al gráfico 13, que nos mostraba cómo se comportaba la variable S.C.P. para los alumnos cuyo centro es público o privado.

A la hora de comprender qué significa esta variación en términos relativos, extraemos las medias de la S.C.P. para las observaciones que tomen cada uno de los dos valores de la variable titularidad, de forma que son:

Valor	Media S.C.P.
Público	9837,23
Privado	10641,26

En consecuencia, podemos afirmar la conclusión que se mostrará a continuación en referencia a esta hipótesis sobre si los alumnos que acuden a un centro privado desempeñan un mejor rendimiento académico. Es importante ser conscientes de que haremos analogía entre la Suma del resultado de las Cuatro Pruebas (S.C.P.) y el Rendimiento Académico (R.A.), y que nos referimos a la variación media de un intervalo de confianza del 95%:

***Queda probado que los alumnos que acuden a un centro de titularidad privada presentan un rendimiento académico del 8,2% mayor que aquellos otros estudiantes que acuden a un centro de titularidad pública.***

### ***3. Los alumnos que gozan de un mayor apoyo familiar en referencia a los estudios, desempeñan un mejor rendimiento académico.***

Para comprobar la validez de esta hipótesis, utilizaremos como en casos anteriores la variable suma de los resultados de las cuatro pruebas para evaluar el R.A.; así como las variables “me anima a estudiar” (el padre o la madre).

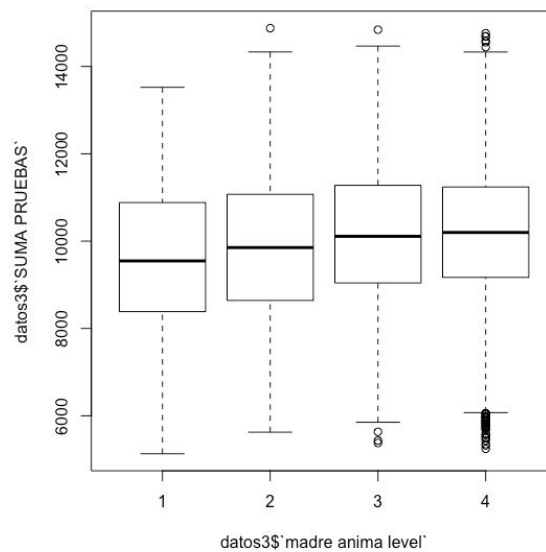
En primer lugar, trataremos de modificar los datos para que estos nos sean propicios para realizar el estudio. La variable referente al apoyo familiar es categórica, sin embargo la clasificaremos en niveles de forma que equivalgan a cada valor de la siguiente manera:

- 1: Nunca o casi nunca
- 2: Algunas veces
- 3: Casi siempre
- 4: Siempre

En el momento que se realiza esta transformación de las variables, dejamos aparte el valor “no contesta”, que se convertirá en “NA” o valor no encontrado en la nueva variable que refleja los niveles. Esto nos será útil porque una vez eliminados los NAs o valores no encontrados, sólo tendremos aquellas observaciones que nos muestran de forma completa el nivel de apoyo familiar, así como el resultado de las S.C.P..

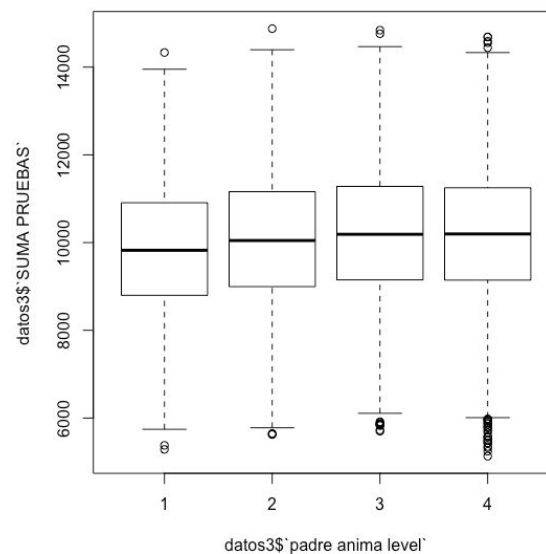
A continuación, en los gráficos 16 y 17, podremos ver de qué forma se comportan y distribuyen los valores de la variable que hace referencia al R.A., en función del ánimo a estudiar que reciben por parte del padre y de la madre.

Gráfico 16: Resultado S.C.P. en función del ánimo a estudiar por parte de la madre.



Fuente: elaboración propia.

Gráfico 17: Resultado S.C.P. en función del ánimo a estudiar por parte del padre.



Fuente: elaboración propia.

Cabe destacar que, en función del nivel de apoyo al estudio por parte de los padres (cuya equivalencia con los valores categóricos se ha señalado anteriormente), los estudiantes muestran cierta mejora en el valor de la S.C.P. mediano. Sin embargo, la diferencia que se presenta entre cada nivel es mucho menor que la que hemos observado en hipótesis anteriores, por lo que gracias al ANOVA comprobaremos si esta diferencia es significativa o no.

Como en casos anteriores, en primera instancia comprobaremos las hipótesis básicas que confieren a los resultados del ANOVA de validez estadística:

- **Independencia:** suponemos su cumplimiento, dado que cada familia se comporta de forma independiente.
- **Normalidad:** llevando a cabo el test de Anderson-Darling, previamente utilizado, se nos muestra que esta hipótesis no se cumple en los siguientes casos:

```
> ad.test(datos31split[[1]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos31split[[1]]$`SUMA PRUEBAS`  
A = 1.4113, p-value = 0.001194
```

```
> ad.test(datos31split[[2]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos31split[[2]]$`SUMA PRUEBAS`  
A = 2.6464, p-value = 1.137e-06
```

Estos dos primeros test, están realizados a los valores de la variable S.C.P. para aquellos casos en los que la variable “madre anima a estudiar” toma los valores “algunas veces” y “casi siempre”, respectivamente. Vemos cómo el p-valor, al ser menor que 0,01, indica que niega la hipótesis nula de normalidad, por lo que ésta hipótesis no se cumpliría para la variable “madre anima a estudiar”.

En la misma línea, ocurre lo mismo para el caso de la variable “padre anima a estudiar”, como queda plasmado a continuación:

```
> ad.test(datos32split[[1]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos32split[[1]]$`SUMA PRUEBAS`  
A = 1.1311, p-value = 0.005849
```

```
> ad.test(datos32split[[2]]$`SUMA PRUEBAS`)
```

Anderson-Darling normality test

```
data: datos32split[[2]]$`SUMA PRUEBAS`  
A = 1.3593, p-value = 0.001607
```

También se muestran los valores “algunas veces” y “casi siempre”, respectivamente, y también comprobamos cómo no se cumple la hipótesis de normalidad. Haremos uso, como en caso anteriores, de la característica de robustez del ANOVA frente a la hipótesis de normalidad cuando las poblaciones son de similar tamaño.

En este caso, el número de observaciones que toma cada valor es muy dispar, como vemos a continuación en el caso de la madre:

Algunas veces	list [1375 x
Casi siempre	list [4192 x
Nunca o casi nunca	list [504 x
Siempre	list [18665

y del padre:

Algunas veces	list [2603 x
Casi siempre	list [4981 x
Nunca o casi nunca	list [1507 x
Siempre	list [15645

Por lo que, al homogeneizar el tamaño de cada lista, realizaremos un muestreo aleatorio sin reemplazo de cada una, extrayendo 500 observaciones.

Posteriormente uniremos las cuatro listas de cada una de las dos variables (apoyo de la madre y del padre), creando dos D.B. que emplearemos para continuar con el análisis.

- **Homocedasticidad:** comprobaremos si las varianzas de cada población, cuyas medias queremos comprobar, son similares. Para ello haremos dos test de Levene, uno para evaluar la homocedasticidad de la variable S.C.P. en función del valor que tome la variable “ánimo de la madre”, y otro para

“ánimo del padre”. Tenemos presente que la hipótesis nula del test de Levene es que hay homocedasticidad.

```
> leveneTest(datosmuestra31$`SUMA PRUEBAS`,datosmuestra31$`Relación con tu madre-Me anima a estudiar`)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  2.816 0.03788 *
      1996
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el caso de la variable “ánimos de la madre”, vemos cómo el P-valor es mayor que el nivel de significación de 0,01, por lo que se acepta la hipótesis nula, y se puede afirmar que hay homocedasticidad. Continuaremos con el ANOVA tras analizar el test de Levene para el caso de los ánimos por parte del padre.

```
> leveneTest(datosmuestra32$`SUMA PRUEBAS`,datosmuestra32$`Relación con tu padre-Me anima a estudiar`)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  1.0178 0.3837
      1996
```

Para la variable “ánimos del padre”, el P-valor es mucho mayor que nuestro nivel de significación, de forma que se cumple la hipótesis nula, que indica la homogeneidad de las varianzas poblacionales. Es decir, en el caso de esta variable referente al padre, también se cumple la hipótesis básica de homocedasticidad.

Una vez comprobadas las hipótesis de investigación, pasaremos a mostrar y comentar los resultados del ANOVA.

```
> anovaanimomadre<-aov(datosmuestra31$`SUMA PRUEBAS`~datosmuestra31$`Relación con tu madre-Me anima a estudiar`)
> summary(anovaanimomadre)
              Df    Sum Sq Mean Sq F value    Pr(>F)
datosmuestra31$`Relación con tu madre-Me anima a estudiar`  3 1.500e+08 49989090  19.1 3.29e-12 ***
Residuals              1996 5.225e+09  2617622
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el caso de esta primera variable, referente a la madre, vemos que se niega la hipótesis nula, afirmando que hay diferencias significativas entre las medias. Realizando un test de Tukey, veremos qué diferencias son las que previamente nos ha indicado el ANOVA.



> TukeyHSD(anovaanimomadre)

Tukey multiple comparisons of means  
95% family-wise confidence level

Fit: aov(formula = datosmuestra31\$`SUMA PRUEBAS` ~ datosmuestra31\$`Relación con tu madre-Me anima a estudiar`)

```
$`datosmuestra31$`Relación con tu madre-Me anima a estudiar``
              diff      lwr      upr      p adj
Casi siempre-Algunas veces    505.24839  242.1496  768.3472 0.0000051
Nunca o casi nunca-Algunas veces -77.53487 -340.6337  185.5639 0.8733683
Siempre-Algunas veces          507.03711  243.9383  770.1359 0.0000047
Nunca o casi nunca-Casi siempre -582.78326 -845.8821 -319.6845 0.0000001
Siempre-Casi siempre           1.78872  -261.3101  264.8875 0.9999981
Siempre-Nunca o casi nunca     584.57199  321.4732  847.6708 0.0000001
```

Como es comprensible, encontraríamos que la mayor diferencia de R.A. (evaluado según el resultado de la S.C.P.), se daría entre los alumnos cuyas madres les animan a estudiar siempre o casi siempre y los que sus madres no les animan nunca a ello. También es relevante el hecho de que no se encuentran diferencias significativas entre el R.A. de los alumnos que sus madres animan a estudiar siempre con los que cuyas madres animan casi siempre. Lo mismo ocurre entre los alumnos que les animan nunca o casi nunca y los que les anima algunas veces.

La mayor diferencia que nos encontramos es de 582,78 puntos de media, en base a un intervalo marcado por un nivel de confianza del 95%. En términos relativos, debemos tener presentes las medias de la S.C.P. para cada valor de la variable “ánimo de la madre al estudio”:

Valor	Media S.C.P.
1: Nunca o casi nunca	9617,23
2: Algunas veces	9694,77
3: Casi siempre	10201,80
4: Siempre	10200,01

Es decir, utilizando términos relativos, podemos extraer la siguiente conclusión:

***Queda probado que los alumnos cuyas madres animan al estudio siempre o casi siempre, obtienen un rendimiento académico aproximadamente del 5% mayor que aquellos otros cuyas madres animan al estudio nunca, casi nunca o algunas veces.***

Por otra parte, realizaremos el ANOVA de la variable S.C.P., que haciendo analogía se referirá al R.A.; y la variable “ánimos al estudio por parte del padre”.

```

> anovaanimopadre<-aov(datosmuestra32$`SUMA PRUEBAS`~datosmuestra32$`Relación con tu padre-Me anima a estudiar`)
> summary(anovaanimopadre)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
datosmuestra32\$`Relación con tu padre-Me anima a estudiar`	3	2.021e+07	6736587	2.855	0.0359 *
Residuals	1996	4.710e+09	2359730		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

El estudio nos indica que, al nivel de significación de 0,01 con el que hemos trabajado en esta hipótesis, no se encuentran diferencias significativas entre el R.A. de los alumnos cuyos padres incitan a estudiar en distintos niveles. Llevando a cabo el test de Tukey veremos de qué forma varían las medias y por qué esta diferencia no es significativa.

```

> TukeyHSD(anovaanimopadre)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = datosmuestra32$`SUMA PRUEBAS` ~ datosmuestra32$`Relación con tu padre-Me anima a estudiar`)

$`datosmuestra32$`Relación con tu padre-Me anima a estudiar``
      diff      lwr      upr      p adj
Casi siempre-Algunas veces    206.06348  -43.73887  455.86583  0.1467347
Nunca o casi nunca-Algunas veces -64.55420  -314.35655  185.24815  0.9104352
Siempre-Algunas veces          72.07163  -177.73072  321.87398  0.8801542
Nunca o casi nunca-Casi siempre -270.61768  -520.42003  -20.81533  0.0276400
Siempre-Casi siempre          -133.99185  -383.79420  115.81050  0.5125907
Siempre-Nunca o casi nunca    136.62583  -113.17652  386.42818  0.4955048

```

Como observamos, en este caso la diferencia media mayor (y la única significativa en el caso de que bajemos el nivel de significación a 0,05) que encontramos entre las medias, es de 270,62 puntos, que relacionando con las siguientes medias:

Valor	Media S.C.P.
1: Nunca o casi nunca	9954,43
2: Algunas veces	10018,98
3: Casi siempre	10225,04
4: Siempre	10091,05

Nos indica que la mayor diferencia, dada entre los alumnos cuyos padres no animan a estudiar nunca o casi nunca y aquellos estudiantes a los que sus padres les animan casi siempre a ello; es del 2,7%, por lo que es perfectamente comprensible que el ANOVA nos indicase que no hay diferencias significativas.

En consecuencia, podemos extraer la siguiente conclusión:

***Queda probado que la frecuencia en la que los padres (no madres) animan a estudiar a los alumnos no repercute en un mayor rendimiento académico de los estudiantes.***

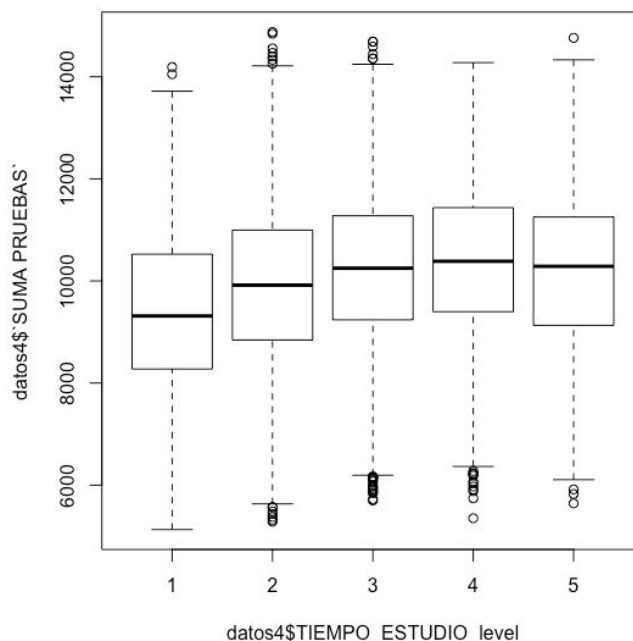
**4. A mayor tiempo dedicado a los deberes, los alumnos muestran un mejor desempeño académico.**

En esta cuarta hipótesis de investigación del trabajo, utilizaremos la variable “tiempo dedicado a estudiar después de clase”, que nos indica el tiempo dedicado a los estudios fuera de clase, así como el resultado de la S.C.P.. La primera variable presenta los valores que mostraremos a continuación, y que trataremos como cinco diferentes niveles:

- 1: Generalmente, no tengo deberes
- 2: 1 hora o menos
- 3: De 1 a 2 horas
- 4: De 2 a 3 horas
- 5: Más de 3 horas

Que de forma gráfica, se comportan de la siguiente manera, como nos muestra el gráfico 18, referente al desempeño académico (representado por la variable S.C.P.) en función del tiempo dedicado a estudiar después de clase.

Gráfico 18: Resultado S.C.P. en función del tiempo dedicado al estudio.



Fuente: elaboración propia.

Queda plasmado que parece haber una mejora en el resultado de la suma de las cuatro pruebas, a medida que se dedica más tiempo al estudio; sin embargo, esta mejoría deja del momento en que se dedica más de 1 a 2 horas medias diarias.

Por ello, para comprobar de forma empírica y estadística este fenómeno, haremos uso del ANOVA. En primer lugar comprobaremos el cumplimiento de las hipótesis básicas, para posteriormente comprobar si hay diferencias significativas entre las medias y en caso de que las haya, cuáles son:

- Independencia: al estar tomadas las observaciones de forma independiente y cada estudiante comportarse de forma independiente en su domicilio, supondremos que se cumple la hipótesis de independencia entre observaciones.
- Normalidad: comprobaremos su cumplimiento mediante el test de Anderson-Darling, sabiendo que su hipótesis nula es la normalidad de las poblaciones.

```
> ad.test(datos4split[[1]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos4split[[1]]$`SUMA PRUEBAS`  
A = 2.5068, p-value = 2.492e-06
```

```
> ad.test(datos4split[[2]]$`SUMA PRUEBAS`)
```

```
Anderson-Darling normality test
```

```
data: datos4split[[2]]$`SUMA PRUEBAS`  
A = 2.238, p-value = 1.13e-05
```

Observamos que no se cumple la hipótesis de normalidad, dado que en los dos casos mostrados, el P-valor toma valores menores que el nivel de significación de 0,01. Los casos son referentes a la normalidad de los resultados de las S.C.P. para los estudiantes que estudien 1 hora o menos y de 1 a 2 horas, respectivamente.

Por ello tomaremos muestras de 500 observaciones sin reemplazamiento de cada una de las poblaciones mostradas a continuación:

1 hora o menos	list [7008 x
De 1 a 2 horas	list [10655
De 2 a 3 horas	list [4661 x
Generalmente, no te...	list [1673 x
Más de 3 horas	list [1741 x

De esta forma podremos continuar con el ANOVA, ya que es robusto frente ausencias de normalidad, cuando los tamaños poblacionales son iguales.

- **Homocedasticidad:** la evaluaremos una vez más con el test de Levene, cuya hipótesis nula hace referencia a que se da homogeneidad entre las varianzas poblacionales.

```
> leveneTest(datosmuestra4$`SUMA PRUEBAS`,datosmuestra4$`TIEMPO DE ESTUDIO DESPUÉS DE CLASE`)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  1.8012 0.1258
2495
```

Asumimos la hipótesis nula, ya que el P-valor es mucho mayor que nuestro nivel de satisfacción de 0,01.

En consecuencia, una vez comprobadas las hipótesis nulas, podremos realizar el análisis de la varianza o ANOVA, que nos indicará si hay diferencias significativas entre las medias poblacionales.

```
> anovaestudio<-aov(datosmuestra4$`SUMA PRUEBAS`~datosmuestra4$`TIEMPO DE ESTUDIO DESPUÉS DE CLASE`)
> summary(anovaestudio)
              Df    Sum Sq Mean Sq F value Pr(>F)
datosmuestra4$`TIEMPO DE ESTUDIO DESPUÉS DE CLASE`  4 2.158e+08 53950741  22.09 <2e-16 ***
Residuals                2495 6.093e+09 2442226
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El análisis nos muestra que en efecto las hay, por lo que sabemos que los alumnos no muestran el mismo R.A. independientemente del tiempo que dediquen al estudio después de clase. El test de Tukey servirá para saber en qué medida y forma se presentan estas diferencias.

> TukeyHSD(anovaestudio)

Tukey multiple comparisons of means  
95% family-wise confidence level

Fit: aov(formula = datosmuestra4\$`SUMA PRUEBAS` ~ datosmuestra4\$`TIEMPO DE ESTUDIO DESPUÉS DE CLASE`)

```
$`datosmuestra4$`TIEMPO DE ESTUDIO DESPUÉS DE CLASE``
              diff            lwr            upr            p adj
De 1 a 2 horas-1 hora o menos      365.21836      95.41383      635.02289 0.0020914
De 2 a 3 horas-1 hora o menos      422.45689     152.65237      692.26142 0.0001931
Generalmente, no tengo deberes-1 hora o menos -365.88458     -635.68910     -96.08005 0.0020383
Más de 3 horas-1 hora o menos       300.24557       30.44104      570.05010 0.0203436
De 2 a 3 horas-De 1 a 2 horas         57.23853     -212.56599      327.04306 0.9781546
Generalmente, no tengo deberes-De 1 a 2 horas -731.10293    -1000.90746     -461.29841 0.0000000
Más de 3 horas-De 1 a 2 horas        -64.97279     -334.77732      204.83174 0.9652906
Generalmente, no tengo deberes-De 2 a 3 horas -788.34147    -1058.14600     -518.53694 0.0000000
Más de 3 horas-De 2 a 3 horas       -122.21133     -392.01585      147.59320 0.7299661
Más de 3 horas-Generalmente, no tengo deberes 666.13014      396.32561      935.93467 0.0000000
```

Los resultados exponen que no hay diferencias significativas entre las medias de todos los alumnos sin que importe el tiempo dedicado a estudiar, sino que se agrupan en los tres siguientes “niveles”:

- generalmente, no tengo deberes,
- 1 hora o menos, y
- de 1 a 2 horas, de 2 a 3 horas, y más de 3 horas.

Hecho que sumado a las medias de cada valor, nos permiten formular la conclusión que se muestra a continuación:

Valor	Media S.C.P.
1: Generalmente, no tengo deberes	9546,63
2: 1 hora o menos	9912,51
3: De 1 a 2 horas'	10277,73
4: De 2 a 3 horas	10334,97
5: Más de 3 horas	10212,75

***Si establecemos los siguientes tres niveles en función del tiempo dedicado a estudiar después de clase:***

- 1: generalmente, no tengo deberes,***
- 2: 1 hora o menos, y***
- 3: de 1 a 2 horas, de 2 a 3 horas, y más de 3 horas.***

***Queda probado que, a medida que se aumenta de nivel, el rendimiento académico medio del alumno medio aumenta de forma significativa, llegando a haber una diferencia de aproximadamente más de 7 puntos porcentuales entre el primer y el último nivel.***

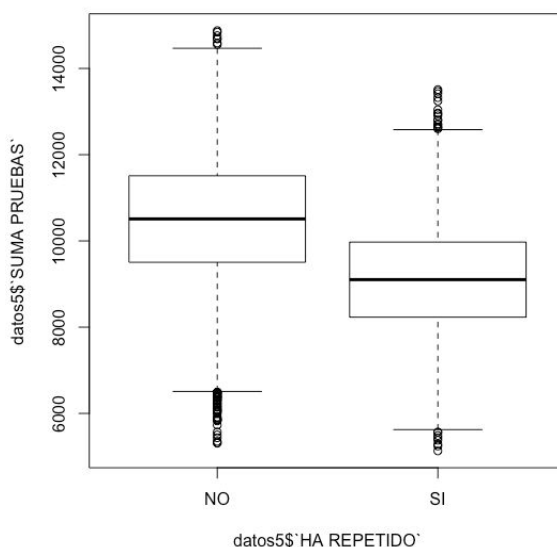
**5. Los alumnos que históricamente presentaban un peor R.A. tienen un peor desempeño académico.**

En este caso el rendimiento histórico únicamente lo podemos evaluar en función de si el estudiante ha **repetido algún curso** o no, mientras que el R.A. lo evaluaremos en función de la suma de los resultados de las cuatro pruebas (la ya mencionada variable SUMA\_PRUEBAS). Nuestro objetivo en esta última hipótesis es comprobar si los alumnos que históricamente presentaban un peor R.A., viéndose materializado este hecho en que han repetido algún curso (2º, 4º ó 6º de Educación Primaria; ó 1º ó 2º de Educación Secundaria Obligatoria). Para ello realizaremos un ANOVA (Analysis on Variance), de forma que comprobaremos de forma veraz si la media de los resultados de los alumnos que sí han repetido curso y los que no es la misma.

En primer lugar, comprobaremos de forma gráfica cómo difieren los resultados para las observaciones que en la variable HA REPETIDO muestran el valor SI y NO, dado que se han retirado previamente todas las observaciones que tomen el valor “no contesta”.

A continuación, como se puede observar en el gráfico 19, se muestra de qué forma se distribuyen los valores tomados en la variable S.C.P, que representa el R.A., en función de si el estudiante ha repetido algún curso o no:

Gráfico 19: Resultado S.C.P. en función de si el alumno ha repetido.



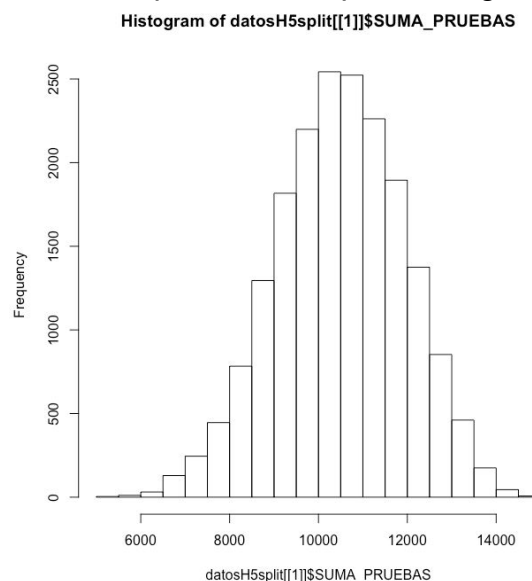
Fuente: elaboración propia.

Observamos que hay cierta diferencia entre las medianas de la suma de resultados de las dos tipologías de observaciones. De hecho, de forma gráfica e intuitiva se podría llegar a afirmar que los alumnos que no han repetido, tienen un mejor desempeño académico. Sin embargo, las intuiciones que percibimos por los gráficos se deben demostrar, por lo que por ello realizaremos el ANOVA.

En primer lugar, para poder realizarlo y que los resultados que nos conceda sean veraces y el análisis correcto, se deben cumplir las hipótesis de partida de dicho análisis. Las condiciones o hipótesis de partida son:

- Independencia de las observaciones: la que se cumple, ya que en este caso se trata de estudiantes independientes y separados.
- Normalidad de las poblaciones: a pesar de que los gráficos 20 y 21 muestren lo que parece una distribución normal en cuanto a la forma en que se comporta la variable S.C.P., ésta será probada mediante el test de Anderson-Darling.

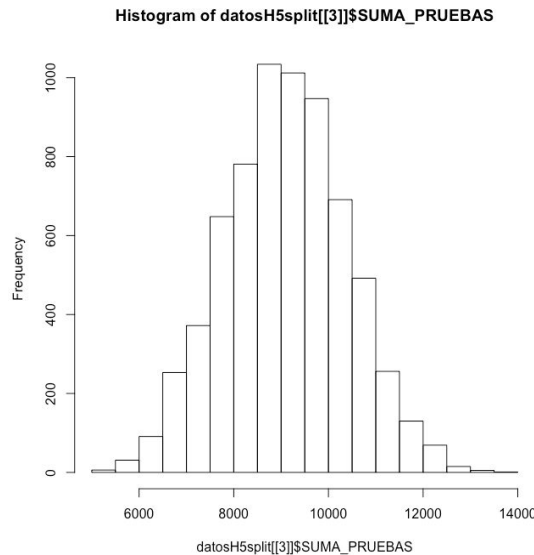
Gráfico 20: Histograma de frecuencia para la suma de las pruebas en observaciones que no han repetido ningún curso.



Fuente: elaboración propia.



Gráfico 21: Histograma de frecuencia para la suma de las pruebas en observaciones que sí han repetido algún curso.



Fuente: elaboración propia.

El previamente mencionado test, nos indicará si efectivamente se cumple la hipótesis nula referente a que se sigue una distribución normal, o por el contrario se niega.

```
> ad.test(datos5split[[1]]$`SUMA PRUEBAS`)
```

Anderson-Darling normality test

```
data: datos5split[[1]]$`SUMA PRUEBAS`  
A = 5.7449, p-value = 3.837e-14
```

```
> ad.test(datos5split[[2]]$`SUMA PRUEBAS`)
```

Anderson-Darling normality test

```
data: datos5split[[2]]$`SUMA PRUEBAS`  
A = 0.5889, p-value = 0.1248
```

Mientras que el primer test se ha realizado para aquellos alumnos que no han repetido ningún curso, y el segundo para los que sí lo han hecho; comprobamos que no se cumple la hipótesis de normalidad, dados los P-valores mayores que 0,01.

Por ello se hará uso de la característica del ANOVA de robustez frente a desviaciones de la normalidad en los casos que las poblaciones sean de tamaños similares. Teniendo en cuenta el número de observaciones que presenta cada valor.

NO	list [19099
SI	list [6834 x

Extraeremos de forma aleatoria 500 observaciones sin reemplazo para así crear una B.D. que contenga dicho número de observaciones con cada uno de los dos valores.

- **Homocedasticidad:** las varianzas de las poblaciones, dos en nuestro caso, deben ser homogéneas. Este hecho lo vemos probado empleando el test de Levene, que nos indica si hay diferencias significativas entre las varianzas de cada población, o por el contrario hay homocedasticidad.

```
> leveneTest(datosmuestra5$`SUMA PRUEBAS`,datosmuestra5$`HA REPETIDO`)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  2.1087 0.1468
      998
```

En nuestro caso, se afirma la hipótesis nula, pudiendo afirmar que las varianzas poblacionales son homogéneas, de forma que se cumplen las hipótesis básicas.

A continuación, llevaremos a cabo el ANOVA, que nos indicará, en caso de que se cumpla la hipótesis nula, que no hay diferencias significativas entre las medias de la S.C.P. de aquellos alumnos que no han repetido ningún curso y los que sí.

```
> anovarepetido<-aov(datosmuestra5$`SUMA PRUEBAS`~datosmuestra5$`HA REPETIDO`)
> summary(anovarepetido)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
datosmuestra5\$`HA REPETIDO`	1	4.984e+08	498404575	263.2	<2e-16 ***
Residuals	998	1.890e+09	1893655		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

En este caso, el P-valor (mostrado como "Pr(>F)") o nivel de significación es menor que  $2 \cdot 10^{-16}$ , valor muy cercano a 0 (como muestran los tres asteriscos); por lo que podemos afirmar que hay diferencias significativas en la suma del resultado de las cuatro pruebas entre los alumnos que no han repetido y los que sí. Es decir, muestra que la hipótesis alternativa es significativa, por lo que podemos concluir que hay diferencias significativas entre las dos medias.

Hasta el momento hemos demostrado que hay diferencias significativas entre ambas medias. No obstante, aún tenemos que comprobar en términos numéricos (aunque nos lo indicase el gráfico 19) de qué forma difieren éstas. Para ello hacemos uso del test de Tukey, que aunque sea propicio para contrastes de más de dos medias, en nuestro caso nos será sumamente útil:

```
> TukeyHSD(anovarepetido)
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = datosmuestra5$`SUMA PRUEBAS` ~ datosmuestra5$`HA REPETIDO`)
```

```
$`datosmuestra5$`HA REPETIDO``
      diff      lwr      upr p adj
SI-NO -1411.955 -1582.743 -1241.168    0
```

El test de Tukey nos muestra que con un nivel de confianza del 95%, el hecho de haber repetido algún curso implica un resultado menor en la suma de los resultados de las cuatro pruebas de entre 1241,17 y 1582,74 puntos menos. Es decir, en términos medios y absolutos, el haber repetido algún curso supone un resultado de 1411,96 puntos de media menos, frente a aquellos alumnos que no han repetido ningún curso.

Esta diferencia queda reflejada en términos absolutos, por lo que es importante mostrarlo también en términos relativos, basándonos en las medias de la S.C.P. en los alumnos que sí han repetido y los que no lo han hecho.

Valor	Media S.C.P.
Sí	9115,25
No	10527,21

En términos relativos, teniendo en cuenta los datos anteriormente mostrados, podemos establecer que el hecho de haber repetido algún curso supone, con un nivel de confianza de 95%, obtener un resultado menor de entre 11,79% y 15,03% que aquellos alumnos que no han repetido. Por lo que para concluir, podemos afirmar lo siguiente:

***Queda probado que los alumnos que han repetido algún curso, tanto de Educación Primaria como de Educación Secundaria Obligatoria, obtienen un rendimiento académico medio del 13,4% menor que aquellos otros alumnos que no han repetido ningún curso.***

*6. A mayor nivel socioeconómico y cultural, los alumnos desempeñan un mejor rendimiento académico.*

En este caso comprobaremos si realmente a mayor nivel socioeconómico y cultural, el desempeño académico mejora. Para ello emplearemos la variable ISEC (Índice Social, Económico y Cultural), que incorpora todas las variables referentes a estos campos que se han evaluado en la B.D. original. A su vez, esta variable muestra datos que incitan a pensar que está normalizada, ya que para la población sobre la que estudiaremos la relación mencionada en esta hipótesis, la variable ISEC tiene media de 0,029 y desviación típica de 1,005. A pesar de ello, cabe destacar que el test Anderson-Darling expone que no se comporta como una distribución normal, como queda patente en el output mostrado a continuación.

```
> ad.test(datos6$ISEC)
```

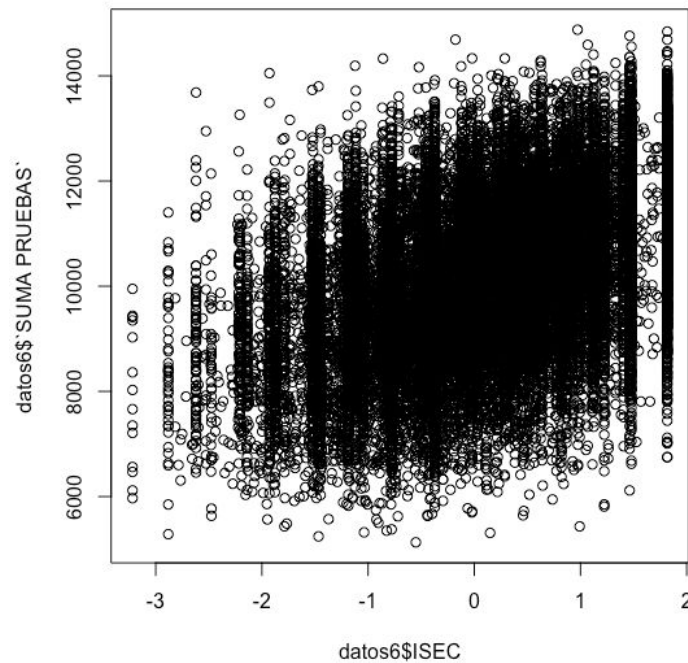
```
Anderson-Darling normality test
```

```
data: datos6$ISEC  
A = 70.505, p-value < 2.2e-16
```

Recordemos que el test de Anderson-Darling evalúa si los valores de la variable se distribuyen como una distribución normal, siendo esta su hipótesis nula; y viéndose negada en este caso, dado el P-valor ínfimo.

En lo referente a la relación que guarda la variable ISEC con el R.A., que una vez más evaluaremos según el resultado de la variable S.C.P., podemos observar en el gráfico 22 que se vislumbra cierta tendencia ascendente o correlación positiva, aunque no sea muy pronunciada.

Gráfico 22: Diagrama de dispersión resultado S.C.P. en función del ISEC.



Fuente: elaboración propia.

De hecho, en el momento que calculamos la correlación entre ambas variables, observamos como efectivamente hay una correlación positiva, aunque débil:

```
> cor(datos6$ISEC,datos6$SUMA PRUEBAS`)  
[1] 0.420037
```

En este caso, al estar trabajando con dos variables cuantitativas continuas (a diferencia de casos anteriores) el coeficiente de correlación de Pearson, que es el calculado anteriormente, es el adecuado para indicarnos la relación entre las variables. Éste nos indica en una escala continua de -1 a 1, en qué grado se da la correlación entre dos variables, siendo -1 una correlación inversa perfecta, 0 ausencia total de correlación, y 1 una correlación directa perfecta. A pesar de que no nos indica la pendiente de la recta de regresión entre variables, es un buen indicador del signo de la correlación, así como para saber si ésta es fuerte o débil.

Para cerciorarnos de que la correlación no se debe al azar, llevamos a cabo el siguiente test de correlación que nos indica, a un intervalo de confianza del 95%, entre qué valores se presentaría la correlación (de Pearson):

```
> cor.test(datos6$ISEC,datos6$`SUMA PRUEBAS`)
```

Pearson's product-moment correlation

```
data: datos6$ISEC and datos6$`SUMA PRUEBAS`  
t = 74.678, df = 26037, p-value < 2.2e-16  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.4099488 0.4299564  
sample estimates:  
      cor  
0.420037
```

Comprobamos que ésta se situaría entre los valores 0,410 y 0,430; a la vez que el P-valor menor que el nivel de significación de 0,01 nos indica que hay una base estadística veraz sobre la que poder afirmar que sí hay correlación entre las variables S.C.P. e ISEC.

Por ello, podemos llegar a la siguiente conclusión:

***Queda probado que hay una correlación positiva entre el rendimiento económico y el nivel sociocultural y económico del alumno, si bien dicha correlación es débil. Indica que en términos medios y generales, a mayor nivel socioeconómico y cultural, mejor rendimiento académico.***

## 7. Conclusiones:

### Estudio de las conclusiones obtenidas.

En el presente trabajo han quedado patentes ciertas informaciones que de cierto modo nos indican otros autores en informes y artículos pasados. Sin embargo, se han procedido a cuantificar, profundizar y probar de forma cualitativa ciertas hipótesis, de las cuales extraemos las siguientes conclusiones:

- *Los alumnos cuyos padres tengan más estudios, desempeñan significativamente un mejor rendimiento académico, viéndose éste más pronunciado en función de los estudios de la madre (hasta una variación del 15%) que del padre (hasta una variación del 10%).* En la misma línea de la afirmación de Santos y Godás (2012) en referencia a que un mayor nivel de estudios de los padres conlleva un mejor desempeño académico del alumno.
- *Los alumnos que acuden a un centro de titularidad privada, desempeñan un mejor rendimiento académico medio del 8% frente a los que acuden a centros de titularidad pública.* También Santos y Godás (2012) señalan que a mayores ingresos familiares, mejor rendimiento académico. Por lo que si hiciésemos una inferencia en que el tener mayores ingresos familiares implica una mayor probabilidad de que el alumno acuda a un centro de titularidad privada, esta conclusión coincide con la bibliografía mencionada.
- *Mientras que el apoyo de la madre en referencia a los estudios puede ayudar a tener un rendimiento académico del 5% mayor, el hecho de que el padre anime al estudio al alumno no produce un mejor rendimiento académico.* Mientras que un gran número de autores han afirmado que el apoyo familiar en referencia a los estudios implica un mejor rendimiento (Lastre, López y Alcázar, 2018; Pérez-Díaz, Rodríguez y Sánchez, 2001); éste hecho sólo se cumple en el caso de la madre.
- *A mayor tiempo dedicado a los deberes, los alumnos muestran un mejor desempeño académico, viéndose éste igualado para los alumnos que estudian a partir de 1 hora diaria en adelante.* La presente conclusión va acorde a los previos estudios que concluyen que a mayor tiempo de estudio, mejor rendimiento académico (Peralbo y Fernández, 2003); Pérez-Díaz, Rodríguez y Sánchez, 2001). Sin embargo, en el trabajo realizado se llega a

la conclusión de que, a partir de una hora diaria de estudio, el rendimiento no mejora, en contra de lo que afirman los mencionados autores.

- *Los alumnos repetidores muestran un desempeño académico significativamente peor (el 13,4% menor) que aquellos alumnos que no han repetido ningún curso.* En este caso la conclusión va completamente acorde a la bibliografía en la que nos hemos basado, que afirma que el rendimiento académico pasado guarda una fuerte relación con el presente (de Miguel, 1988; Goberna y otros, 1987; House, Hurst & Keely 1996; Jiménez, 1987; Wilson & Hardgrave, 1995).
- *En términos generales y medios, a mayor nivel socioeconómico y cultural del alumno, éste muestra un mejor rendimiento académico.* Por último, esta conclusión confirma que la familia tiene un gran peso en la educación del estudiante (Ruiz de Miguel, 2001), y el nivel socioeconómico de la misma influye positivamente (Lorenzo, Santos y Godás, 2012).

En consecuencia, hemos comprobado la veracidad de las hipótesis de investigación, obteniendo como resultado la veracidad de muchas de ellas, pero no de todas. Por ello, se ha aportado un pequeño grano de arena al castillo que supone el estudio del impacto del nivel socioeconómico de los padres en el rendimiento académico de los hijos. Se ha tratado el estudio de forma estadística, objetiva y empírica, para que de esta forma se aporte información veraz sobre el tema a todo aquel que la precise.



## Bibliografía:

alfabetización, Diccionario de la lengua española. Recuperado 1 de junio de 2020, de <https://dle.rae.es/alfabetizaci%C3%B3n>

alfabetizar, Diccionario de la lengua española. Recuperado 1 de junio de 2020, de <https://dle.rae.es/alfabetizar>

Becker, G.S. (1964). *Human Capital*. Nueva York: Columbia University Press for the National Bureau of Economic Research.

Borrás, F., Martínez, C. & Escobar, L.S. (2017). *Estadística Empresarial en 101 ejemplos (volumen I)*. Madrid: EV Services.

California Office of Environmental Health Hazard Assessment (OEHHA). (2017). Nivel Educativo. Recuperado 3 de junio de 2020, de <https://oehha.ca.gov/calenviroscreen/indicator/nivel-educativo#:~:text=El%20nive%20educativo%20es%20el,las%20personas%20con%20menos%20educaci%C3%B3n>

De Miguel, M. (1988). *Preescolarización y rendimiento académico: un estudio longitudinal de las variables psicosociales a lo largo de la E.G.B.* Madrid: Centro de Publicaciones del Ministerio de Educación y Ciencia : C.I.D.E.

De Gabriel, N. (1997). *Alfabetización, semialfabetización y analfabetismo en España (1860-1991)*. Madrid: Revista Complutense de Educación, vol. 8, nº 1, 1997.

Delgado, B. (1993). *"Política educativa" en Historia de la educación en España y América. La educación en la España Moderna*. Madrid: Ediciones Morata.

Edel, R. (2003). El rendimiento académico: concepto, investigación y desarrollo. *REICE (Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación)*, vol. 1, Nº 2, jul-dic, 2003.

educación, Diccionario de la lengua española. Recuperado 1 de junio de 2020, de <https://dle.rae.es/educaci%C3%B3n>

Erazo, O.A. (2012). El Rendimiento Académico, un Fenómeno de Múltiples Relaciones y Complejidades. *Revista Vanguardia Psicológica Clínica Teórica y Práctica*, ISSN2216-0701, Vol. 2, nº 2, octubre-marzo de 2012.

Expansión. (s. f.). *¿Qué es el crecimiento económico y de qué depende?* Recuperado 3 de junio de 2020, de <https://www.expansion.com/economia-para-todos/economia/que-es-el-crecimiento-economico-y-de-que-depende.html>

Feldman, L., Goncalves, L., Chacón-Puignau, G., Zaragoza, J., Bagés, N. & De Pablo, J. (2008). Relaciones entre estrés académico, apoyo social, salud mental y rendimiento académico en estudiantes universitarios venezolanos. *Universitas Psychologica*, V. 7, Nº 3, Sept-Dic 2008, pp. 739-751.

Goberna, M.A., López M.A. & Pastor J.T. (1987). La predicción del rendimiento como criterio para el ingreso en la universidad. *Revista de Educación*, 283, pp. 235-248.

González, R. (2008). *Pobreza Absoluta y Crecimiento Económico, Análisis de Tendencia en México, 1970-2005*. (Tesis de maestría). Universidad Autónoma del Estado de México, Facultad de Economía, Toluca.

House, J. D., Hurst, R.S., Keely, E.J. (1996). Relationship between learner attitudes, prior achievement, and performance in a General Education. Course: A multi-Institutional Study. *International Journal of instructional media*, 23, pp. 257-271.

INEE (Instituto Nacional de Evaluación Educativa). (2017). Análisis de los factores asociados al rendimiento de los alumnos en PISA | Blog de INEE. Recuperado 4 de junio de 2020, de <http://blog.intef.es/inee/2017/02/16/analisis-de-los-factores-asociados-al-rendimiento-de-los-alumnos-en-pisa/>

Jiménez, C. (1987). Rendimiento académico en la universidad a distancia. Un estudio empírico sobre su evolución y predicción (II). *Revista de Educación*, 284, pp. 317-347.

Jiménez, M. (2000). Competencia social: intervención preventiva en la escuela. *Infancia y Sociedad*, 24, pp. 21-48.

Kagan, L. R. (1981). *Universidad y sociedad en la España moderna*. Madrid: Editorial Tecnos.

Lastre, K., López, L. & Alcázar, C. (2018). Relación entre apoyo familiar y el rendimiento académico en estudiantes colombianos de educación primaria. *Psicogente*, 21 (39), pp. 102-115.

Lorenzo, M., Santos, M.A. & Godás, A. (2012). Inmigración y Educación. ¿Influye el Nivel Educativo de los Padres en el Rendimiento Académico de los Hijos? *Teoría Educación*, 24, 2-2012, pp. 129-148.

Martínez, C., Rúa, A., Martín, M.J., Redondo, R., Fabra, M.E. & Núñez, A. (2010). *Influencia del Nivel Educativo de los Padres en el Rendimiento Académico de los Estudiantes de ADE. Un Enfoque de Género*. Recuperado de [https://www.researchgate.net/publication/228506347\\_Influencia\\_del\\_Nivel\\_Educativo\\_de\\_los\\_Padres\\_en\\_el\\_Rendimiento\\_Academico\\_de\\_los\\_Estudiantes\\_de\\_ADE\\_Un\\_Enfoque\\_de\\_Genero](https://www.researchgate.net/publication/228506347_Influencia_del_Nivel_Educativo_de_los_Padres_en_el_Rendimiento_Academico_de_los_Estudiantes_de_ADE_Un_Enfoque_de_Genero)

Datosmacro.com. (s. f.). *Mejora la tasa de alfabetización en España*. Recuperado 1 de junio de 2020, de <https://datosmacro.expansion.com/demografia/tasa-alfabetizacion/espana>

Martínez, C., Álvarez, C., Budría, S., Curto, T., Escobar, L.S. & Borrás, F. (2017). *Modelos Cuantitativos para la Economía y la Empresa en 101 ejemplos*. Madrid: EV Services.

Mella, O. & Ortiz, I. (1999). Rendimiento escolar. Influencias diferenciales de factores externos e internos. *Revista Latinoamericana de Estudios Educativos* (México), vol. XXIX, núm. 1, 1º trimestre, 1999, pp. 69-92.

Neira, I. (2007). Capital Humano y Desarrollo Económico Mundial: Modelos Económicos y Perspectivas. *Estudios Económicos de Desarrollo Internacional*, July-December, año/vol. 7, número 002, pp. 53-80.

OCDE. (s.f.). PISA - PISA. Recuperado 3 de junio de 2020, de <https://www.oecd.org/pisa/pisaenespaol.htm>

Peralbo, M. & Fernández, M.L. (2003). *Estructura Familiar y Rendimiento Escolar en Educación Secundaria Obligatoria*. La Coruña: Revista Galego-Portuguesa de Psicología e Educación, vol. 8, nº 7, 2003.

Pérez, J. & Gardey, A. (2008). Definicion.de: Definición de rendimiento académico. Recuperado 1 de junio de 2020, de <https://definicion.de/rendimiento-academico/>

Pérez-Díaz, V., Rodríguez, J.C. & Sánchez, L. (2001). *La familia española ante la educación de sus hijos*. Barcelona: Fundación "la Caixa".

PISA. (s.f.). Recuperado 3 de junio de 2020, de <https://www.educacionyfp.gob.es/inee/bases-datos/evaluaciones-internacionales/pisa.html>

PISA - PISA. Recuperado 1 de junio de 2020, de <https://www.oecd.org/pisa/pisaenespaol.htm>

Popkewitz, T.S. (2013). PISA: números, estandarización de la conducta y la alquimia de las materias escolares. Profesorado. *Revista de currículum y formación del profesorado*, 17(2).

Programa de las Naciones Unidas para el Desarrollo (PNUD). (2018). Índices e indicadores de desarrollo humano. Actualización estadística de 2018. Recuperado de [http://hdr.undp.org/sites/default/files/2018\\_human\\_development\\_statistical\\_update\\_es.pdf](http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update_es.pdf)

Ruiz de Miguel, C. (2001). Factores familiares vinculados al bajo rendimiento. *Revista Complutense de Educación*, ISSN: 1130-2496, Vol 12, Nº 1(2001), pp. 81-113.

Schultz, T.W. (1961). Investment in Human Capital. *The American Economic Review*, 51.1, 1961, pp. 1-17.

Viñao, A. (1993). *"La educación institucional. Alfabetización y escolarización" en Historia de la educación en España y América. La educación en la España Moderna*. Madrid: Ediciones Morata.

Wilson, R.L., Hardgrave, B.C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational and Psychological Measurement*, 55, 186-195.

## Anexo I: Resto de Variables de la Base de Datos.

### Variables referentes al **idioma**:

- Lengua que habla **en casa**: Castellano, Catalán, Gallego, Valenciano, Vasco, Árabe, Chino, Inglés, Rumano, u Otra lengua.
- Lengua que habla **en el colegio**: Castellano, Catalán, Gallego, Valenciano, Vasco, Árabe, Chino, Inglés, Rumano, u Otra lengua.
- Lengua que habla **con los amigos**: Castellano, Catalán, Gallego, Valenciano, Vasco, Árabe, Chino, Inglés, Rumano, u Otra lengua.

Lengua en la que **escucha la TV**: Castellano, Catalán, Gallego, Valenciano, Vasco, Árabe, Chino, Inglés, Rumano, u Otra lengua.

### **Relación con los profesores, los compañeros y el centro**:

- Los estudiantes se llevan bien con la mayoría de los profesores y profesoras: pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta.
- A la mayoría de los profesores y profesoras les interesa el bienestar de los estudiantes: pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta.
- La mayoría de mis profesores y profesoras realmente escuchan lo que tengo que decir: pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta.
- Recibo ayuda extra de mis profesores y profesoras si la necesito: pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta.
- La mayoría de mis profesores y profesoras me tratan de manera justa: pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta.
- **Compañeros de clase**:
  - Me siento marginada o marginado
  - Hago amigas y amigos fácilmente
  - Me siento integrada o integrado
  - Me siento incómoda o incómodo
  - Caigo bien a otros estudiantes
  - Me siento sola o solo
  - Me siento acosada o acosado por otros compañeros

pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta a cada opción.

- **Razones asistencia** al centro:
  - Este es el centro para los estudiantes que viven en esta zona
  - Este centro tiene fama de ser mejor que otros de la zona
  - Ofrece estudios específicos
  - Tiene unas creencias determinadas
  - Otros miembros de mi familia estudiaron en este centro
  - Otras razones

pudiendo responder Sí, No o No contesta a cada opción.

- En relación con tu **centro**:
  - Mi centro es muy bueno
  - En general, me gusta venir al a mi centro
  - Las clases son entretenidas
  - El trabajo que hago en las clases me gusta mucho
  - Aprendo cosas interesantes y útiles
  - Me gustaría cambiar de centro

pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta a cada opción.

- Situaciones en **clase**:
  - Tenemos tareas asignadas (ordenar la clase, limpieza, avisos, material)
  - Se mantiene un orden adecuado para poder trabajar a gusto
  - El profesor o profesora nos felicita cuando nos portamos bien
  - El profesor o profesora dialoga con nosotros cuando no nos portamos bien
  - El profesor o profesora nos castiga si no respetamos las normas
  - El profesor o profesora fija las normas
  - Nosotros participamos en el establecimiento de normas
  - En general, respetamos las normas
  - Colaboramos para que se cumplan las normas de la clase

pudiendo responder Nunca o casi nunca, Algunas veces, Casi siempre, Siempre, o No contesta a cada opción.

**Desempeño** de las clases:

- **Trabajo** en clase con **profesores**:
  - Explican durante la mayor parte de la clase
  - Los alumnos exponemos temas o trabajos
  - Mientras explican nos preguntan sobre los temas
  - Mientras explican, preguntamos las dudas
  - Hacemos debates en clase
  - Hacemos los ejercicios o actividades que nos proponen
  - Trabajamos individualmente

- Trabajamos en pequeños grupos
- Tomamos apuntes

pudiendo responder Nunca o casi nunca, Algunas veces, Casi siempre, Siempre, o No contesta a cada opción.

- **Utilización materiales en clase:**

- El libro de texto para estudiar y hacer ejercicios
- Otros ejercicios que no son del libro
- Libros y cuentos de la biblioteca
- Materiales elaborados por el profesor
- Prensa o revistas especializadas
- Ordenadores e internet
- Medios audiovisuales, videos, DVD

pudiendo responder Nunca o casi nunca, Algunas veces, Casi siempre, Siempre, o No contesta a cada opción.

- **Evaluaciones:**

- Nos hacen preguntas sobre un tema antes de iniciarlo
- Realizan controles después de cada tema o lección
- Nos ponen controles escritos trimestrales
- Corrigen los deberes y cuadernos
- Revisan las actividades que hacemos en clase
- Nos hacen exámenes orales
- Tienen en cuenta la nota que nosotros mismos nos ponemos
- Valoran el interés y la participación en clase

pudiendo responder Nunca o casi nunca, Algunas veces, Casi siempre, Siempre, o No contesta a cada opción.

### **Creencias** en relación a los estudios y futuro académico:

- **Ir al centro:**

- Me ha enseñado cosas que pueden ser útiles en un trabajo
- Me ha ayudado a tener confianza en mí mismo para tomar decisiones
- Ha hecho poco para prepararme para la vida adulta cuando deje el centro

pudiendo responder Muy en desacuerdo, En desacuerdo, De acuerdo, Muy de acuerdo, o No contesta a cada opción.

- Nivel hasta el que **quiere estudiar:** pudiendo responder Hasta terminar los estudios obligatorios (ESO), Hasta terminar un Ciclo Formativo de grado medio, Hasta terminar Bachillerato, Hasta terminar un Ciclo Formativo de grado superior, Hasta terminar una carrera universitaria, No lo he pensado todavía, o No contesta.

Variables referentes al **estudio en casa**:

- En tu casa hay:
  - Una mesa para estudiar
  - Una habitación para ti solo
  - Un sitio tranquilo para estudiar
  - Un ordenador que puedas utilizar para estudiar
  - Conexión a Internet
  - Tu propio e-book
  - Libros de lectura
  - Libros para ayudarte con tus estudios
  - Un diccionario

pudiendo responder Sí, No o No contesta a cada opción.

- N° de libros en casa: pudiendo responder De 0 a 50, De 51 a 100, De 101 a 200, Más de 200, o No contesta.
- Frecuencia uso en casa:
  - Libros de lectura
  - Prensa o revistas
  - Enciclopedias
  - Ordenador
  - Internet
  - Televisión
  - Vídeo, DVD

pudiendo responder De 0 a 50, De 51 a 100, De 101 a 200, Más de 200, o No contesta a cada opción.



## Anexo II: Código empleado para el análisis en Rstudio.

```
#H1: R.A. en función de estudios de la madre
datos11<-datos[,c(4,44)]
datos11<-datos11[datos11$`ESTUDIOS MADRE`!="No lo sé",]
library(normtest)
library(nortest)
library(rapportools)
levels<-c('No completó estudios obligatorios','Estudios obligatorios
(ESO,EGB)','Técnico FP grado medio','Bachillerato','Técnico superior de
FP','Universitarios medios','Universitarios superiores')
datos11$`ESTUDIOS MADRE level`<-match(datos11$`ESTUDIOS MADRE`,levels)
boxplot(datos11$`SUMA PRUEBAS`~datos11$`ESTUDIOS MADRE level`)
datos11<-na.omit(datos11)
datos11split<-split(datos11,datos11$`ESTUDIOS MADRE`)

ad.test(datos11split[[3]]$`SUMA PRUEBAS`)
hist(datos11split[[3]]$`SUMA PRUEBAS`)
ad.test(datos11split[[2]]$`SUMA PRUEBAS`)
hist(datos11split[[2]]$`SUMA PRUEBAS`)

muestra111<-sample(1:nrow(datos11split$Bachillerato),size=1000,replace=FALSE)
datosmuestra111<- datos11split$Bachillerato[muestra111, ]
muestra112<-sample(1:nrow(datos11split$`Estudios obligatorios
(ESO,EGB)`),size=1000,replace=FALSE)
datosmuestra112<- datos11split$`Estudios obligatorios (ESO,EGB)`[muestra112, ]
muestra113<-sample(1:nrow(datos11split$`No completó estudios
obligatorios`),size=1000,replace=FALSE)
datosmuestra113<- datos11split$`No completó estudios obligatorios`[muestra113, ]
muestra114<-sample(1:nrow(datos11split$`Técnico FP grado
medio`),size=1000,replace=FALSE)
datosmuestra114<- datos11split$`Técnico FP grado medio`[muestra114, ]
muestra115<-sample(1:nrow(datos11split$`Técnico superior de
FP`),size=1000,replace=FALSE)
datosmuestra115<- datos11split$`Técnico superior de FP`[muestra115, ]
muestra116<-sample(1:nrow(datos11split$`Universitarios
medios`),size=1000,replace=FALSE)
datosmuestra116<- datos11split$`Universitarios medios`[muestra116, ]
muestra117<-sample(1:nrow(datos11split$`Universitarios
superiores`),size=1000,replace=FALSE)
```

```

datosmuestra117<- datos11split$`Universitarios superiores`[muestra117, ]

datosmuestratodos<-rbind(datosmuestra111,datosmuestra112,datosmuestra113,da
tosmuestra114,datosmuestra115,datosmuestra116,datosmuestra117)
leveneTest(datosmuestratodos$`SUMA
PRUEBAS`,datosmuestratodos$`ESTUDIOS MADRE level`)
boxplot(datosmuestratodos$`SUMA PRUEBAS`~datosmuestratodos$`ESTUDIOS
MADRE level`)
anovaestmadre<-aov(datosmuestratodos$`SUMA
PRUEBAS`~datosmuestratodos$`ESTUDIOS MADRE`)
summary(anovaestmadre)
TukeyHSD(anovaestmadre)
mean(datosmuestra111$`SUMA PRUEBAS`)
mean(datosmuestra112$`SUMA PRUEBAS`)
mean(datosmuestra113$`SUMA PRUEBAS`)
mean(datosmuestra114$`SUMA PRUEBAS`)
mean(datosmuestra115$`SUMA PRUEBAS`)
mean(datosmuestra116$`SUMA PRUEBAS`)
mean(datosmuestra117$`SUMA PRUEBAS`)

#H1: R.A. en función de estudios del padre
datos12<-datos[,c(5,44)]
datos12<-datos12[datos12$`ESTUDIOS PADRE`!="No lo sé",]
levels<-c('No completó estudios obligatorios','Estudios obligatorios
(ESO,EGB)','Técnico FP grado medio','Bachillerato','Técnico superior de
FP','Universitarios medios','Universitarios superiores')
datos12$`ESTUDIOS PADRE level` <-match(datos12$`ESTUDIOS PADRE`,levels)
boxplot(datos12$`SUMA PRUEBAS`~datos12$`ESTUDIOS PADRE level`)
datos12<-na.omit(datos12)
datos12split<-split(datos12,datos12$`ESTUDIOS PADRE`)

ad.test(datos12split[[1]]$`SUMA PRUEBAS`)
ad.test(datos12split[[2]]$`SUMA PRUEBAS`)
ad.test(datos12split[[3]]$`SUMA PRUEBAS`)
ad.test(datos12split[[4]]$`SUMA PRUEBAS`)
ad.test(datos12split[[5]]$`SUMA PRUEBAS`)
ad.test(datos12split[[6]]$`SUMA PRUEBAS`)
ad.test(datos12split[[7]]$`SUMA PRUEBAS`)

muestra121<-sample(1:nrow(datos12split$Bachillerato),size=1000,replace=FALSE)
datosmuestra121<- datos12split$Bachillerato[muestra121, ]

```

```

muestra122<-sample(1:nrow(datos12split$`Estudios obligatorios
(ESO,EGB)`),size=1000,replace=FALSE)
datosmuestra122<- datos12split$`Estudios obligatorios (ESO,EGB)`[muestra122, ]
muestra123<-sample(1:nrow(datos12split$`No completó estudios
obligatorios`),size=1000,replace=FALSE)
datosmuestra123<- datos12split$`No completó estudios obligatorios`[muestra123, ]
muestra124<-sample(1:nrow(datos12split$`Técnico FP grado
medio`),size=1000,replace=FALSE)
datosmuestra124<- datos12split$`Técnico FP grado medio`[muestra124, ]
muestra125<-sample(1:nrow(datos12split$`Técnico superior de
FP`),size=1000,replace=FALSE)
datosmuestra125<- datos12split$`Técnico superior de FP`[muestra125, ]
muestra126<-sample(1:nrow(datos12split$`Universitarios
medios`),size=1000,replace=FALSE)
datosmuestra126<- datos12split$`Universitarios medios`[muestra126, ]
muestra127<-sample(1:nrow(datos12split$`Universitarios
superiores`),size=1000,replace=FALSE)
datosmuestra127<- datos12split$`Universitarios superiores`[muestra127, ]

```

```

datosmuestra12<-rbind(datosmuestra121,datosmuestra122,datosmuestra123,datos
muestra124,datosmuestra125,datosmuestra126,datosmuestra127)
leveneTest(datosmuestra12$`SUMA PRUEBAS`,datosmuestra12$`ESTUDIOS
PADRE level`)
boxplot(datosmuestra12$`SUMA PRUEBAS`~datosmuestra12$`ESTUDIOS
PADRE level`)
anovaestpadre<-aov(datosmuestra12$`SUMA
PRUEBAS`~datosmuestra12$`ESTUDIOS PADRE`)
summary(anovaestpadre)
TukeyHSD(anovaestpadre)
mean(datosmuestra121$`SUMA PRUEBAS`)
mean(datosmuestra122$`SUMA PRUEBAS`)
mean(datosmuestra123$`SUMA PRUEBAS`)
mean(datosmuestra124$`SUMA PRUEBAS`)
mean(datosmuestra125$`SUMA PRUEBAS`)
mean(datosmuestra126$`SUMA PRUEBAS`)
mean(datosmuestra127$`SUMA PRUEBAS`)

```

```

#H2: R.A. en función de la titularidad del centro
datos2<-datos[,c(2,44)]
boxplot(datos2$`SUMA PRUEBAS`~datos2$TITULAR)
datos2split<-split(datos2,datos2$TITULAR)
hist(datos2split[[1]]$`SUMA PRUEBAS`)

```

```

hist(datos2split[[2]]$`SUMA PRUEBAS`)
ad.test(datos2split[[1]]$`SUMA PRUEBAS`)
ad.test(datos2split[[2]]$`SUMA PRUEBAS`)

library(normtest)
library(nortest)
muestra21<-sample(1:nrow(datos2split$PRIVADO),size=4000,replace=FALSE)
datosmuestra21<- datos2split$PRIVADO[muestra21, ]
muestra22<-sample(1:nrow(datos2split$PUBLICO),size=4000,replace=FALSE)
datosmuestra22<- datos2split$PUBLICO[muestra22, ]

datosmuestra2<-rbind(datosmuestra21,datosmuestra22)
leveneTest(datosmuestra2$`SUMA PRUEBAS`,datosmuestra2$TITULAR)
anovatitularidad<-aov(datosmuestra2$`SUMA
PRUEBAS`~datosmuestra2$TITULAR)
summary(anovatitularidad)
TukeyHSD(anovatitularidad)
mean(datosmuestra21$`SUMA PRUEBAS`)
mean(datosmuestra22$`SUMA PRUEBAS`)

#H3: R.A. en función del apoyo familiar
datos3<-datos[,c(19,30,44)]
levels<-c('Nunca o casi nunca','Algunas veces','Casi siempre','Siempre')
datos3$`madre anima level`<-match(datos3$`Relación con tu madre-Me anima a
estudiar`,levels)
datos3$`padre anima level`<-match(datos3$`Relación con tu padre-Me anima a
estudiar`,levels)
datos3<-na.omit(datos3)
datos31split<-split(datos3,datos3$`Relación con tu madre-Me anima a estudiar`)
datos32split<-split(datos3,datos3$`Relación con tu padre-Me anima a estudiar`)
ad.test(datos31split[[1]]$`SUMA PRUEBAS`)
ad.test(datos31split[[2]]$`SUMA PRUEBAS`)
ad.test(datos32split[[1]]$`SUMA PRUEBAS`)
ad.test(datos32split[[2]]$`SUMA PRUEBAS`)

muestra311<-sample(1:nrow(datos31split$`Algunas
veces`),size=500,replace=FALSE)
datosmuestra311<- datos31split$`Algunas veces`[muestra311, ]
muestra312<-sample(1:nrow(datos31split$`Casi
siempre`),size=500,replace=FALSE)
datosmuestra312<- datos31split$`Casi siempre`[muestra312, ]

```

```

muestra313<-sample(1:nrow(datos31split$`Nunca o casi
nunca`),size=500,replace=FALSE)
datosmuestra313<- datos31split$`Nunca o casi nunca`[muestra313, ]
muestra314<-sample(1:nrow(datos31split$Siempre),size=500,replace=FALSE)
datosmuestra314<- datos31split$Siempre[muestra314, ]
datosmuestra31<-rbind(datosmuestra311,datosmuestra312,datosmuestra313,datos
muestra314)

```

```

muestra321<-sample(1:nrow(datos32split$`Algunas
veces`),size=500,replace=FALSE)
datosmuestra321<- datos32split$`Algunas veces`[muestra321, ]
muestra322<-sample(1:nrow(datos32split$`Casi
siempre`),size=500,replace=FALSE)
datosmuestra322<- datos32split$`Casi siempre`[muestra322, ]
muestra323<-sample(1:nrow(datos32split$`Nunca o casi
nunca`),size=500,replace=FALSE)
datosmuestra323<- datos32split$`Nunca o casi nunca`[muestra323, ]
muestra324<-sample(1:nrow(datos32split$Siempre),size=500,replace=FALSE)
datosmuestra324<- datos32split$Siempre[muestra324, ]
datosmuestra32<-rbind(datosmuestra321,datosmuestra322,datosmuestra323,datos
muestra324)

```

```

leveneTest(datosmuestra31$`SUMA PRUEBAS`,datosmuestra31$`Relación con tu
madre-Me anima a estudiar`)

```

```

leveneTest(datosmuestra32$`SUMA PRUEBAS`,datosmuestra32$`Relación con tu
padre-Me anima a estudiar`)

```

```

anovaanimomadre<-aov(datosmuestra31$`SUMA
PRUEBAS`~datosmuestra31$`Relación con tu madre-Me anima a estudiar`)

```

```

summary(anovaanimomadre)

```

```

anovaanimopadre<-aov(datosmuestra32$`SUMA
PRUEBAS`~datosmuestra32$`Relación con tu padre-Me anima a estudiar`)

```

```

summary(anovaanimopadre)

```

```

TukeyHSD(anovaanimomadre)

```

```

TukeyHSD(anovaanimopadre)

```

```

mean(datosmuestra311$`SUMA PRUEBAS`)

```

```

mean(datosmuestra312$`SUMA PRUEBAS`)

```

```

mean(datosmuestra313$`SUMA PRUEBAS`)

```

```

mean(datosmuestra314$`SUMA PRUEBAS`)

```

```

mean(datosmuestra321$`SUMA PRUEBAS`)

```

```

mean(datosmuestra322$`SUMA PRUEBAS`)

```

```

mean(datosmuestra323$`SUMA PRUEBAS`)

```

```

mean(datosmuestra324$`SUMA PRUEBAS`)

```

```

#H4: R.A. en función del tiempo de estudio
datos4<-datos[,c(16,44)]
levels<-c('Generalmente, no tengo deberes','1 hora o menos','De 1 a 2 horas','De 2
a 3 horas','Más de 3 horas')
datos4$TIEMPO_ESTUDIO_level<-match(datos4$`TIEMPO DE ESTUDIO
DESPUÉS DE CLASE`,levels)
datos4<-na.omit(datos4)
boxplot(datos4$`SUMA PRUEBAS`~datos4$TIEMPO_ESTUDIO_level)
datos4split<-split(datos4,datos4$`TIEMPO DE ESTUDIO DESPUÉS DE CLASE`)
ad.test(datos4split[[1]]$`SUMA PRUEBAS`)
ad.test(datos4split[[2]]$`SUMA PRUEBAS`)

muestra41<-sample(1:nrow(datos4split$`1 hora o
menos`),size=500,replace=FALSE)
datosmuestra41<- datos4split$`1 hora o menos`[muestra41, ]
muestra42<-sample(1:nrow(datos4split$`De 1 a 2
horas`),size=500,replace=FALSE)
datosmuestra42<- datos4split$`De 1 a 2 horas`[muestra42, ]
muestra43<-sample(1:nrow(datos4split$`De 2 a 3
horas`),size=500,replace=FALSE)
datosmuestra43<- datos4split$`De 2 a 3 horas`[muestra43, ]
muestra44<-sample(1:nrow(datos4split$`Generalmente, no tengo
deberes`),size=500,replace=FALSE)
datosmuestra44<- datos4split$`Generalmente, no tengo deberes`[muestra44, ]
muestra45<-sample(1:nrow(datos4split$`Más de 3
horas`),size=500,replace=FALSE)
datosmuestra45<- datos4split$`Más de 3 horas`[muestra45, ]
datosmuestra4<-rbind(datosmuestra41,datosmuestra42,datosmuestra43,datosmues
tra44,datosmuestra45)

leveneTest(datosmuestra4$`SUMA PRUEBAS`,datosmuestra4$`TIEMPO DE
ESTUDIO DESPUÉS DE CLASE`)
anovaestudio<-aov(datosmuestra4$`SUMA PRUEBAS`~datosmuestra4$`TIEMPO
DE ESTUDIO DESPUÉS DE CLASE`)
summary(anovaestudio)
TukeyHSD(anovaestudio)
mean(datosmuestra41$`SUMA PRUEBAS`)
mean(datosmuestra42$`SUMA PRUEBAS`)
mean(datosmuestra43$`SUMA PRUEBAS`)
mean(datosmuestra44$`SUMA PRUEBAS`)
mean(datosmuestra45$`SUMA PRUEBAS`)

```

```

#H5: R.A. en función del desempeño histórico
datos5<-datos[,c(10,44)]
levels<-c('NO','SI')
datos5$HA_REPETIDO_level<-match(datos5$`HA REPETIDO`,levels)
datos5<-na.omit(datos5)
boxplot(datos5$`SUMA PRUEBAS`~datos5$`HA REPETIDO`)
datos5split<-split(datos5,datos5$`HA REPETIDO`)
ad.test(datos5split[[1]]$`SUMA PRUEBAS`)
ad.test(datos5split[[2]]$`SUMA PRUEBAS`)

muestra51<-sample(1:nrow(datos5split$NO),size=500,replace=FALSE)
datosmuestra51<- datos5split$NO[muestra51, ]
muestra52<-sample(1:nrow(datos5split$SI),size=500,replace=FALSE)
datosmuestra52<- datos5split$SI[muestra52, ]
datosmuestra5<-rbind(datosmuestra51,datosmuestra52)

leveneTest(datosmuestra5$`SUMA PRUEBAS`,datosmuestra5$`HA REPETIDO`)
anovarepetido<-aov(datosmuestra5$`SUMA PRUEBAS`~datosmuestra5$`HA
REPETIDO`)
summary(anovarepetido)
TukeyHSD(anovarepetido)
mean(datosmuestra51$`SUMA PRUEBAS`)
mean(datosmuestra52$`SUMA PRUEBAS`)

#H6: R.A. en función del ISEC
datos6<-datos[,c(39,44)]
var(datos6$ISEC)
mean(datos6$ISEC)
median(datos6$ISEC)
sd(datos6$ISEC)
plot(x=datos6$ISEC,y=datos6$`SUMA PRUEBAS`)
cor(datos6$ISEC,datos6$`SUMA PRUEBAS`)
leveneTest(datos6$`SUMA PRUEBAS`,datos6$ISEC)

```