Imagina que trabajas en una agencia de medios y te han encargado que estudies el prestigio conferido por la sociedad respecto a ciertas profesiones. Para ello, vamos a suponer que los datos recogidos en el dataset Duncan (correspondientes a una investigación realizada en 1961 y muy empleados en el ámbito educativo) continúan vigentes en la actualidad.

En la sección «Descargables» dispones de la hoja de cálculo Duncan.csv, con dichos datos separados por comas. Cuando procedas a su descarga comprobarás que el dataset contiene 5 columnas:

• En la primera se relacionan las distintas profesiones a analizar (150 en total).

• En la segunda se indica en qué tipología se encuadra la profesión: o *Blue collar*, que es el término inglés que hace referencia a un trabajo de naturaleza manual.

o *White collar*, que caracteriza labores administrativas y comerciales.

o *Prof*, relativo a trabajos profesionales o gerenciales.

• La tercera columna (*income* o ingresos) recoge el porcentaje de personas, con la profesión correspondiente a la fila en la que se relacionan los datos, que perciben un salario superior a los 32 000 € anuales de la época actual.

• La cuarta columna (*education* o educación) muestra el porcentaje de personas trabajadoras con la profesión que corresponda que cuentan con un grado universitario.

• La quinta columna (p*restige* o prestigio) contiene los porcentajes de respuesta de personas que participaron en una encuesta y que calificaron a la profesión de la que se trate como buena o idónea.

En base a lo expuesto:

1. Descarga los datos de la hoja de cálculo y averigua su estructura, con objeto de identificar las transformaciones que permitan aplicar el método fit. Denota las variables income, education y prestige como ingresos.

2. Analiza la regresión lineal entre income (como variable independiente) y prestige (como variable independiente). En concreto, debes obtener el error cuadrático medio del conjunto de prueba, los coeficientes de regresión para ambos conjuntos y una representación gráfica del modelo sobre el conjunto de prueba. ¿Cómo valoras los resultados obtenidos? Interprétalos.