**Taller 1.**

**Predicción de Ingresos.**

**Integrantes:** Isabella Mendez Pedraza. Cód.: 201814239

Manuela Ojeda Ojeda. Cód.:

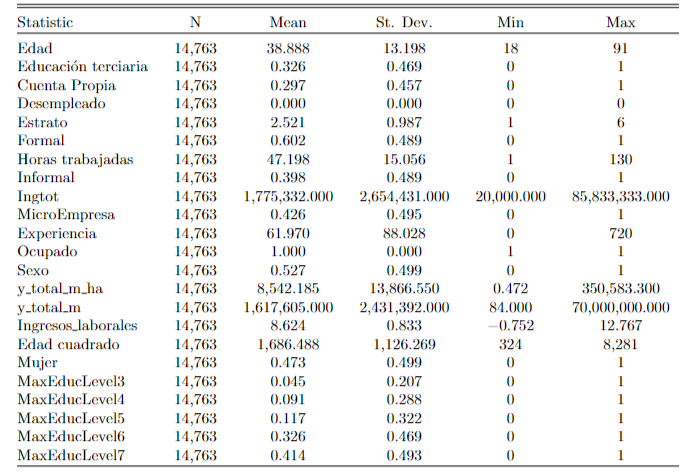
Juan Sebastian Tellez Melo. Cód.: 201513710

Andres Mauricio Palacio Lugo. Cód.: 201618843

1. **Introducción.**
2. **Datos.**

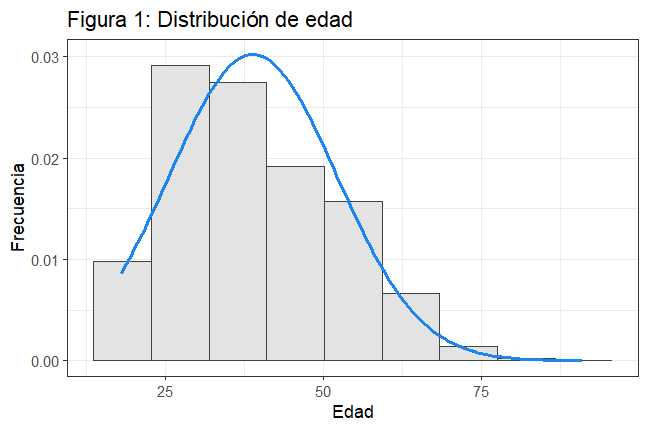
**Estadísticas descriptivas**

Tabla 1. Estadísticas descriptivas

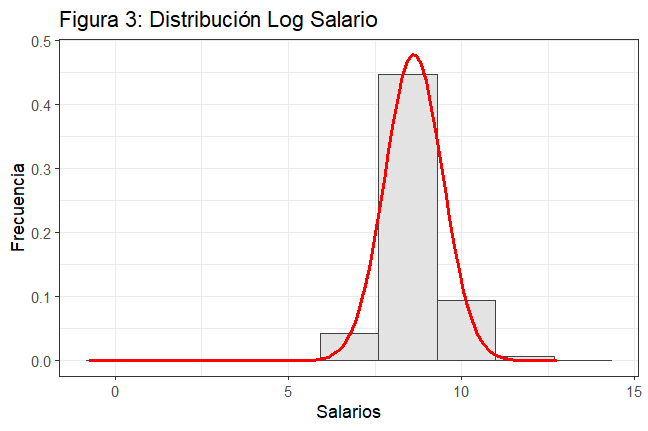
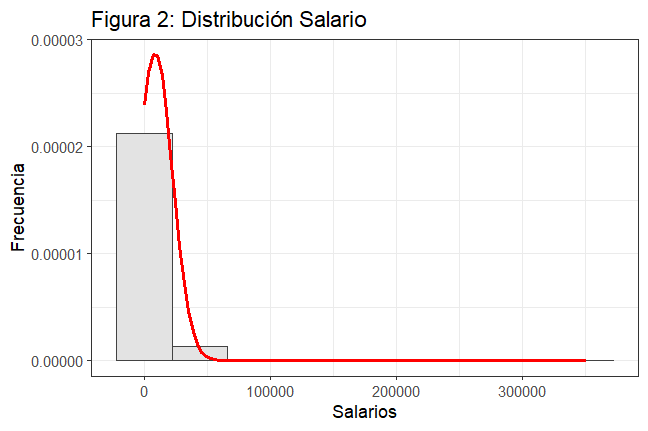


Como parte de la limpieza de datos, se restringió la información a sujetos con edades a partir de los 18 años y que estuviesen ocupados (utilizando la variable “ocu =1”). Así, llegamos a una muestra de 14,763 observaciones. De esta muestra, el 47.3% corresponde a mujeres mientras que el 52.7% son hombres.

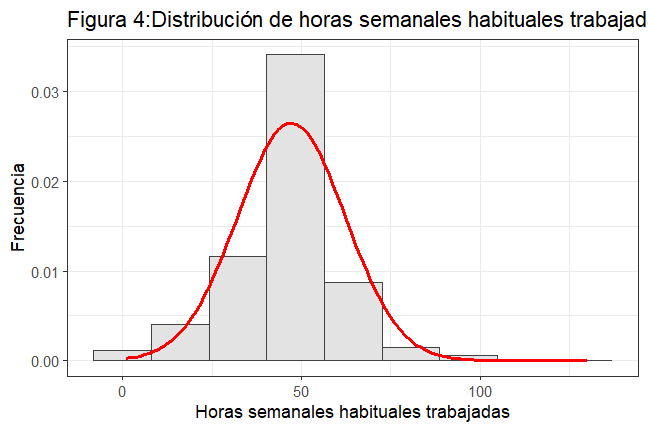
La edad promedio de las personas es 39 años con una desviación estándar de 13 años y la persona ocupada con mayor edad tiene 91 años.



En la figura 1 observamos la distribución por edad, allí se evidencia que solo tenemos personas mayores de edad y hay muy pocas personas que tienen más de 75 años. También se observa que las edades con mayor frecuencia están entre los 20 años y 30 años.



En la figura 2 se observa la distribución de los salarios sin ningún tratamiento, sin embargo, en la figura 3 se evidencia cómo es la distribución del logaritmo del salario de los individuos, que es la variable que se utilizará en las estimaciones de este documento. Se observa que la media del logaritmo del salario está entre $8 y $9 mil pesos por hora.



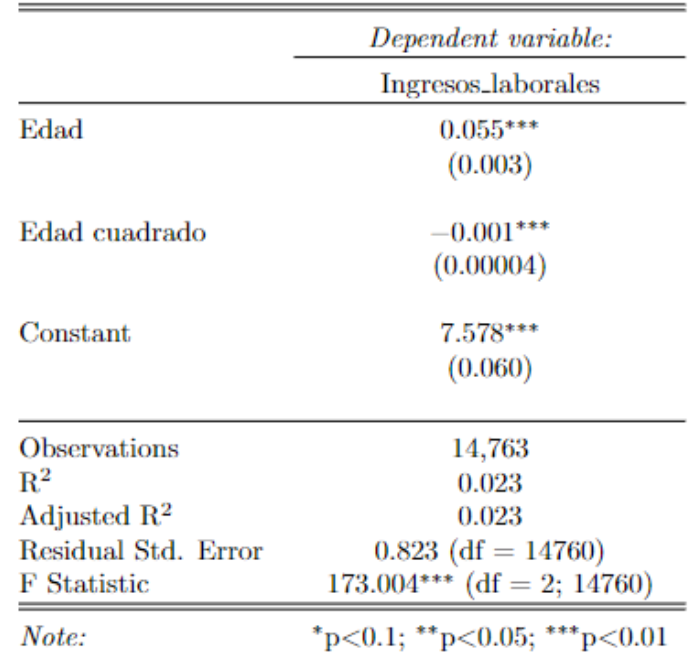
A partir de la tabla 1 y la figura 4 se observa que en promedio los individuos trabajan 47 horas habitualmente en la semana, la persona que menos trabaja es 1 hora y la que más hora trabaja llega a 130 horas. Los individuos trabajan con mayor frecuencia entre 40 y 50 horas habitualmente en la semana.

1. **Perfil Edad-Salario**

Para estimar el perfil edad-salario de los individuos vamos a estimar:

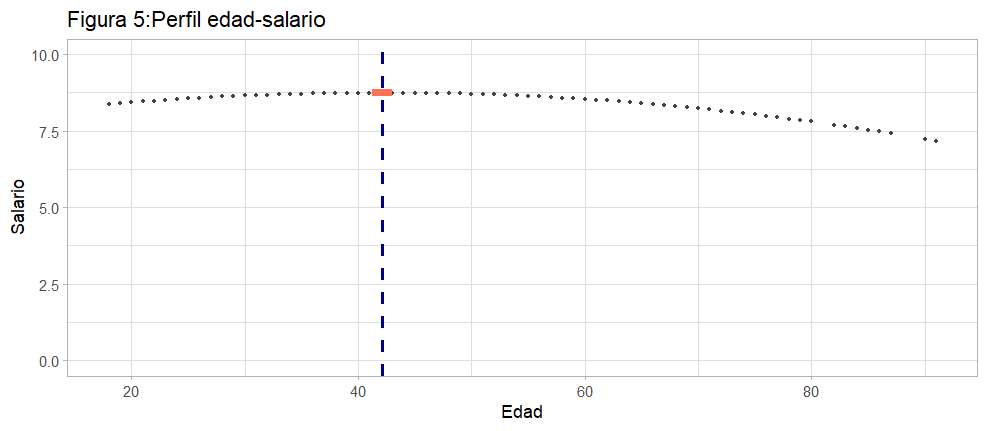
Tomamos nuestra variable Ingresos\_totales que es el logaritmo de y\_total\_m\_ha (ingresos asalariados + independientes total - nominal por hora). Al estimar el modelo encontramos que el incremento de 1 año de edad es asociado a un cambio en los ingresos de 5.5%. Observamos que nuestro modelo tiene un R2 de 0.02, es decir, que solamente el 20% de la varianza total de nuestro resultado objetivo está siendo explicado por nuestro modelo, por lo que podemos pensar que no es el mejor modelo.

Tabla 2, Perfil edad-salario



En la figura 5 observamos que como era de esperarse y según la evidencia de la economía laboral los salarios tienden a ser bajos cuando el trabajador es joven y aumentan a medida que el trabajador crece llegando a un máximo, en donde el salario empieza a disminuir.

Continuando con el modelo anterior y para la construcción del bootstrap utilizamos intervalos de confianza en los cuales encontramos que la edad mínima es 41.24 años y la máxima es 42.95 años. Adicionalmente, se observa que la edad máxima es 42.09 años. Es decir, que en promedio aproximadamente a los 42 años los individuos obtienen su logaritmo de salario máximo de aproximadamente $8.750 por hora con una confianza del 95%.



1. **Brecha salarial por género**
2. **Predicción de ingresos**