**Juan Sebastian Velasquez Diaz  
Métodos de análisis cuantitativos II  
Octubre de 2020**

**EXAMEN 1**

En la siguiente tabla aparecen algunos de los datos, y las variables que serán objeto de estudio, siendo “luz” la variable dependiente y x0, x1, x2, x3, x4, x5 y mes las variables independientes.

Se llama la base de datos con el siguiente código = ***Base1<-read.csv("examen1.csv")***

Y se instalan los siguientes paquetes

***install.packages("plm")***

***install.packages("ggplot2")***

***library("plm")***

***library("ggplot2")***

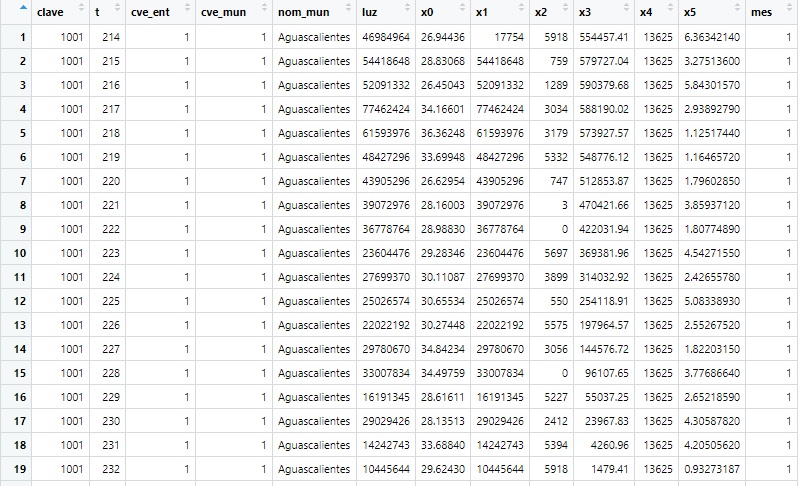
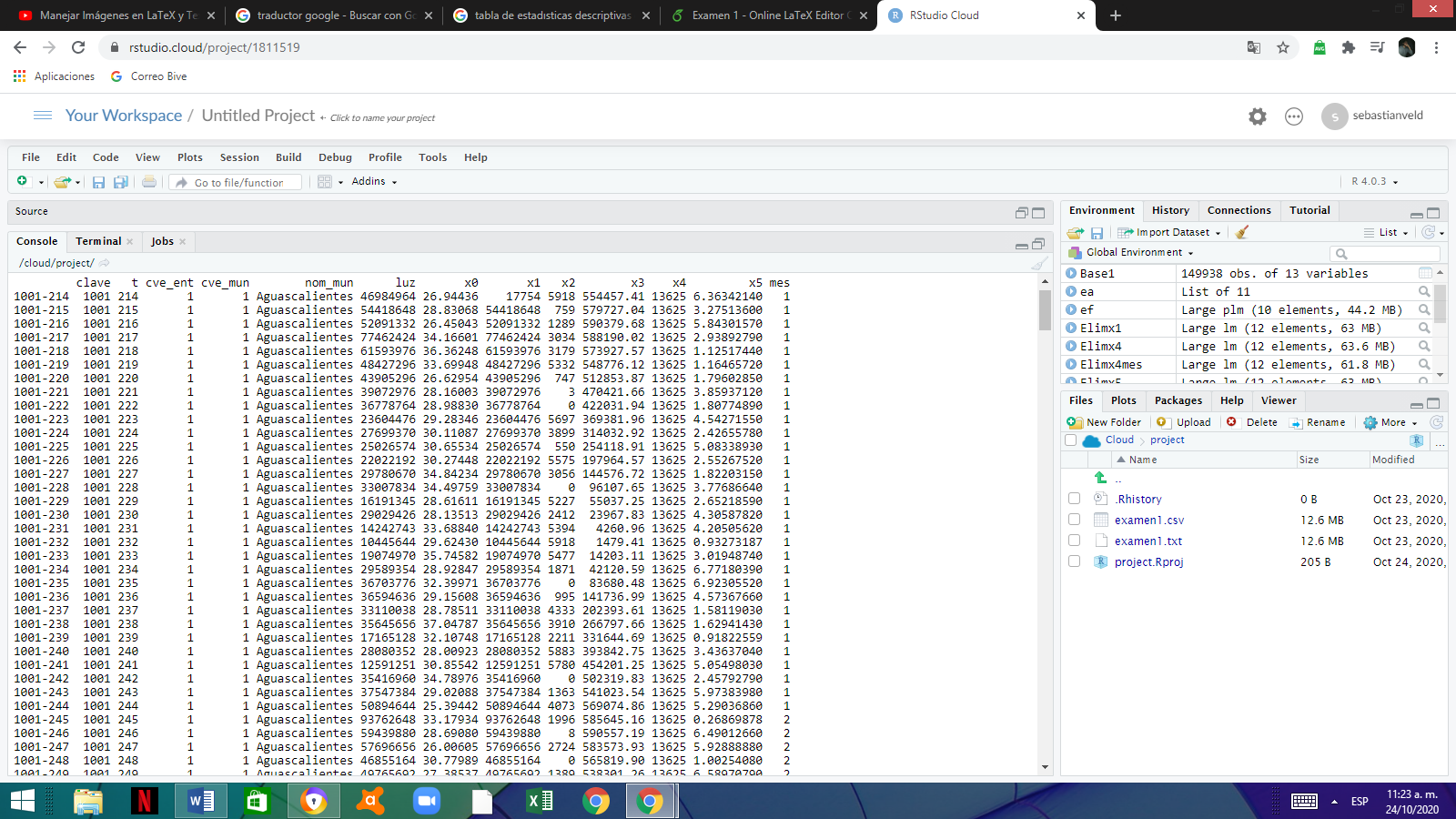


Tabla . Extracto de la base de datos

**1.** Incorporar la base de datos (panel) en R, y declararla como panel

Se utiliza el siguiente código para declarar la base de datos previamente incorporada a R, como panel =

***Panel1 <- pdata.frame(Base1, index = c("clave", "t"))***  , y se utiliza el cogido ***Panel1***, para extraer una parte de él.



2. Realizar una tabla de estadísticas descriptivas de las variables

Se realiza la siguiente regresión para conocer y presentar en una tabla de estadística descriptiva de las variables

***reg1<-lm(luz~x0+x1+x2+x3+x4+x5+mes,data=Panel1)***

***summary(reg1)***

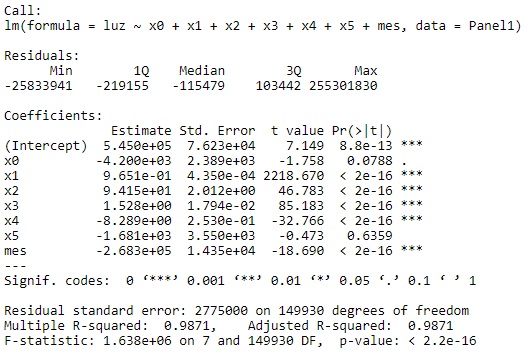
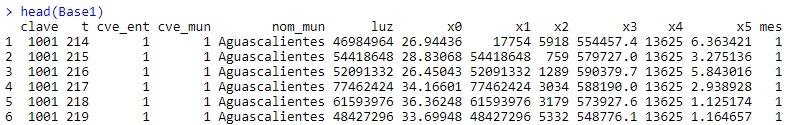


Tabla . Estadísticos descriptivos

Esta tabla nos muestra que el intercepto, al igual que las variables x1, x2, x3, x4 y mes son significativas a más del 99%. Y las variables explican en un 0,9871 el modelo.

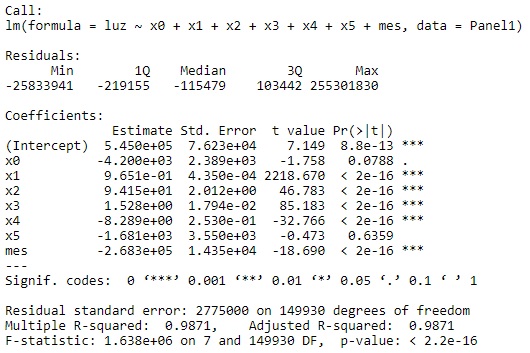
***head(Base1)***



3. Realizar regresión pooled-OLS

***OLS<- lm(luz~x0+x1+x2+x3+x4+x5+mes,data=Panel1)***

***summary(OLS)***

  
Tabla . Regresion pooled OLS

Se observa que los coeficientes del intercepto, x0, x2, x3, x4 y mes son significativamente distintos de cero. Puesto que las estimaciones de los coeficientes x0 y x4 son negativas, cuando estas variables aumentan disminuye los niveles de luz. Por otro lado, puesto que las estimaciónes de x1, x2, x3 y x5 son positivas, cuando éstas variables aumentan también lo hacen los niveles de luz. Finalmente, puesto que el p-valor es menor que 0.05, se tiene que el modelo es válido conjuntamente.

4. Determinar si la regresión es de efectos fijos o aleatorios

Efectos fijos

***ef<-plm(luz~x0+x1+x2+x3+x4+x5+mes,data=Panel1,index=c("clave", "t"),model="within")***

***summary(ef)***

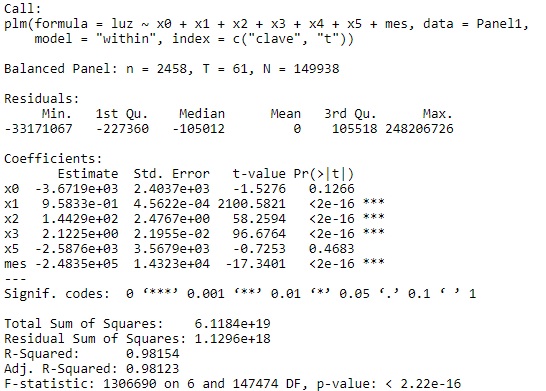


Tabla 4. Efectos fijos

La variable independiente no aparece ya que el modelo estimado es el de desviaciones con respecto a la media. De hecho, elimina todas las variables constantes en cada individuo, y en este caso elimina x4 porque presenta dicha característica.

Se presentan cambios además en las estimaciones de los coeficientes que son significativamente cercanos a 0, ya que son x1, x2, x3 y mes los que presentan esta condiciones. Además, las variables x0, x5 y mes al ser negativas cuando presentan un aumento hacen disminuir los niveles de luz, mientras que el incremento en las otras variables tiene el mismo efecto en los niveles de luz.

Efectos aleatorios

***ea <- plm(luz~x0+x1+x2+x3+x4+x5+mes,data=Panel1, index=c("clave", "t"), model="random")***

***summary(ea)***

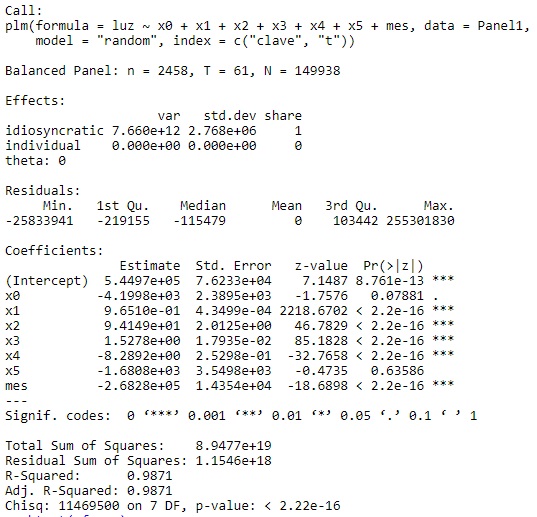


Tabla . Efectos aleatorios

Al observar el valor estimado de Theta se puede concluir que los efectos aleatorios se asemejan más a los resultados obtenidos por el modelo agrupado, ya que este es equivalente a 0.

Prueba para elegir efectos fijos o efectos aleatorios

***phtest(ef, ea)***

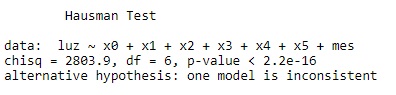


Tabla . Prueba Hausman Test

Al utilizar la prueba de Hausman en donde la hipótesis nula es que “el modelo de efectos aleatorios es mejor que el de efectos fijos”, y en este caso se rechaza dicha hipótesis ya que el P value es menor a 0,05, por lo tanto se prefieren los efectos fijos.

5. Probar si la base de datos tiene heteroscedasticidad, si la tiene, corregirla

***library(lmtest)***

***bptest(ef)***

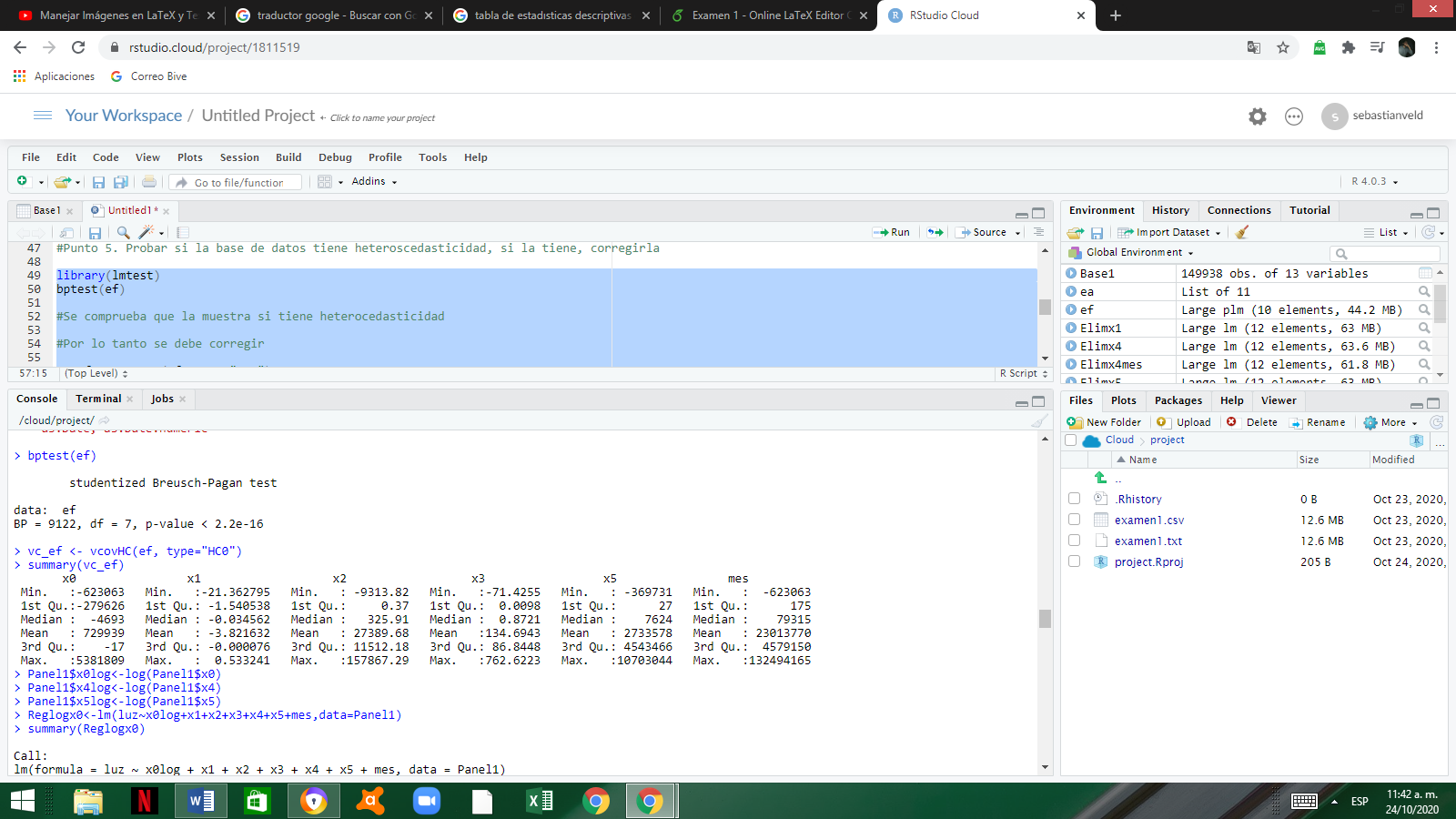


Tabla . Prueba de heterocedasticidad

Se comprueba que la muestra si tiene heterocedasticidad , por lo tanto se debe corregir

***vc\_ef <- vcovHC(ef, type="HC0")***

***summary(vc\_ef)***

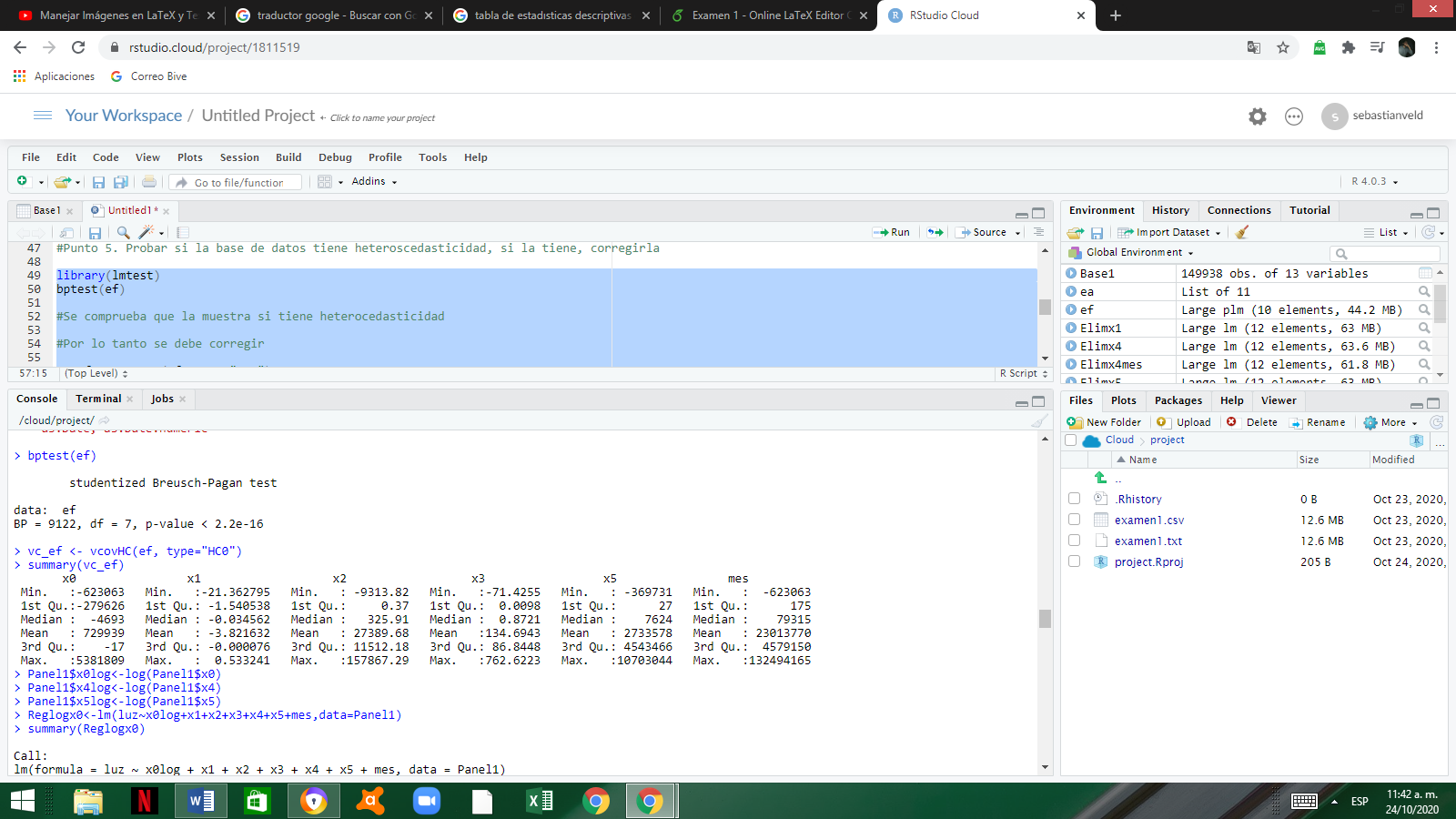


Tabla . Corrección de heterocedasticidad

6. ¿Qué otras maneras de presentar las variables se te ocurren para que la Adj.R2 aumente de valor? ¿Es valida esa especificación?

Opción 1. Haciendo logarítmicas algunas variables

***Panel1$x0log<-log(Panel1$x0)***

***Panel1$x4log<-log(Panel1$x4)***

***Panel1$x5log<-log(Panel1$x5)***

***Reglogx0<-lm(luz~x0log+x1+x2+x3+x4+x5+mes,data=Panel1)***

***summary(Reglogx0)***

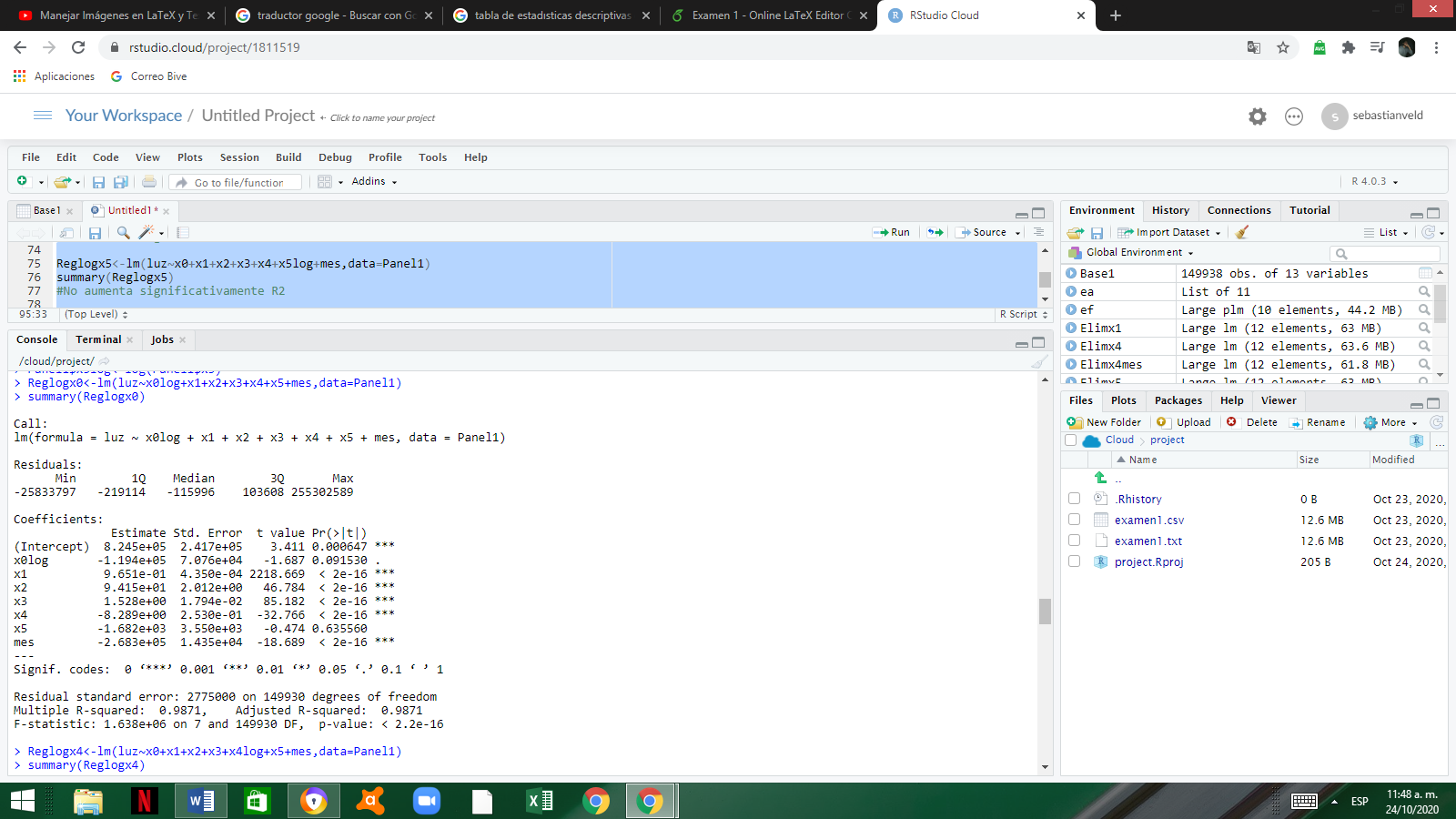


Tabla . Prueba 1

No aumento R2

***Reglogx4<-lm(luz~x0+x1+x2+x3+x4log+x5+mes,data=Panel1)***

***summary(Reglogx4)***

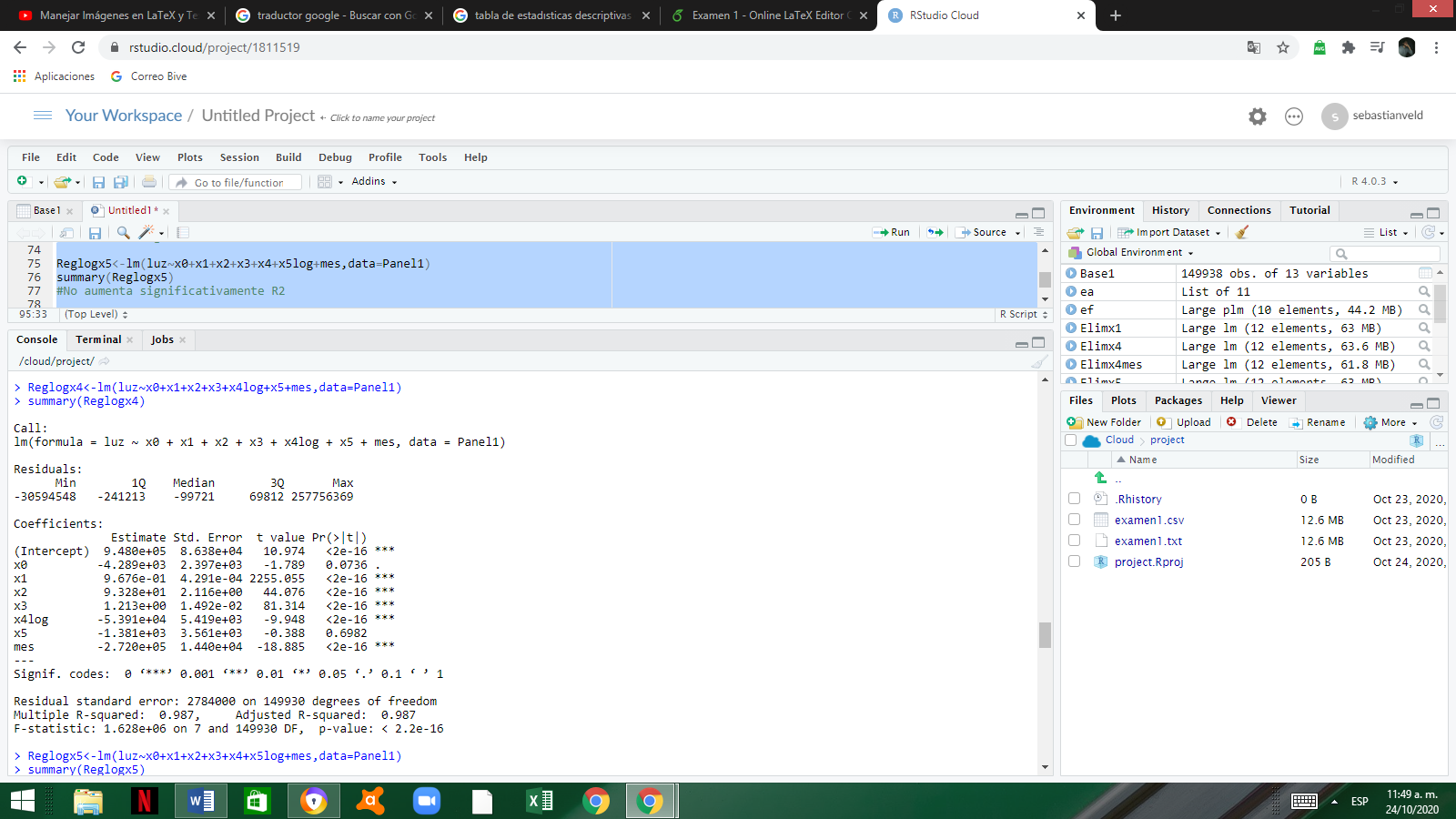


Tabla . Prueba 2

No aumenta signifcativamente R2

***Reglogx5<-lm(luz~x0+x1+x2+x3+x4+x5log+mes,data=Panel1)***

***summary(Reglogx5)***

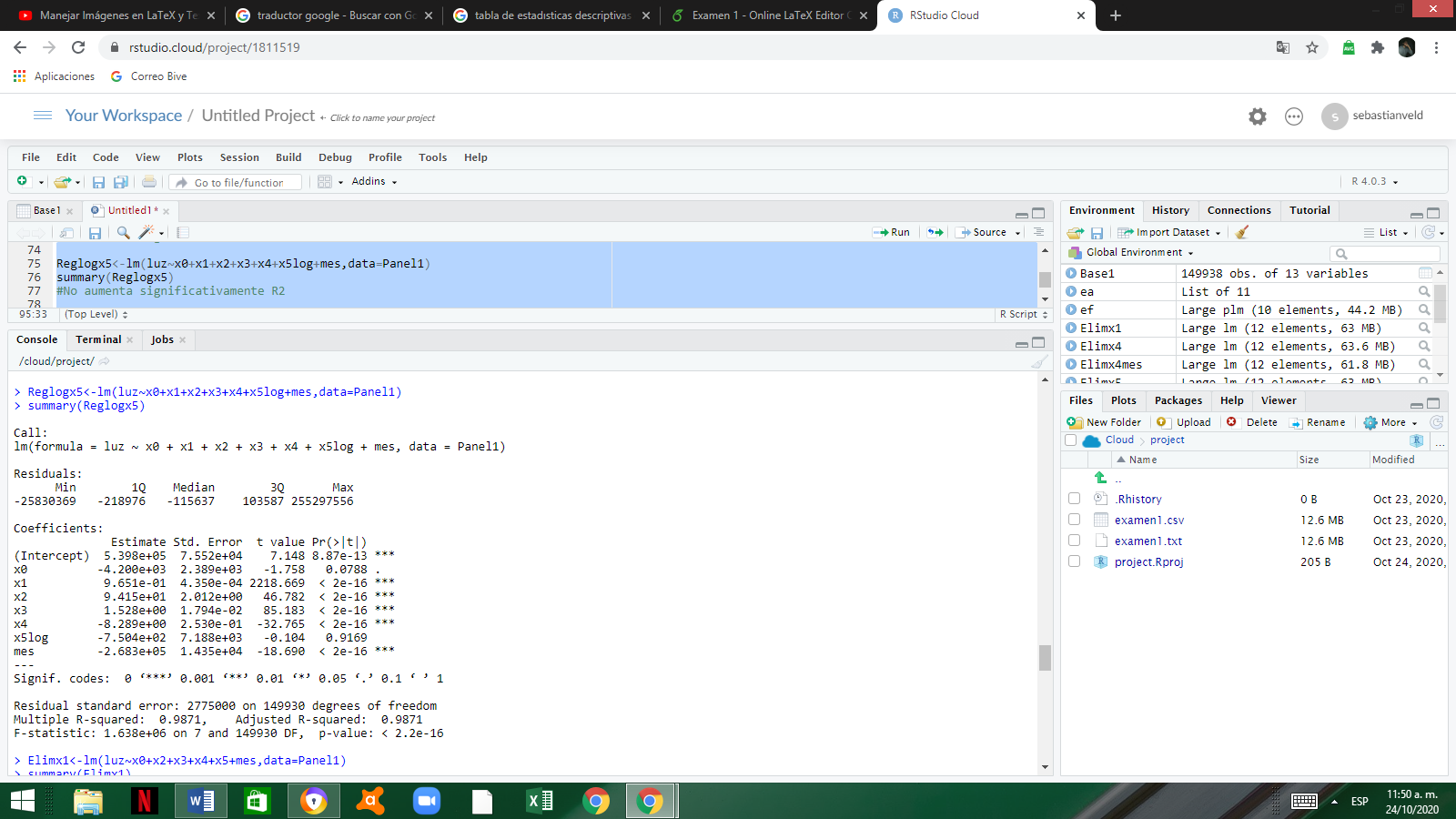


Tabla . Prueba 3

No aumenta significativamente R2

Opción 2. Eliminando variables

***Elimx1<-lm(luz~x0+x2+x3+x4+x5+mes,data=Panel1)***

***summary(Elimx1)***

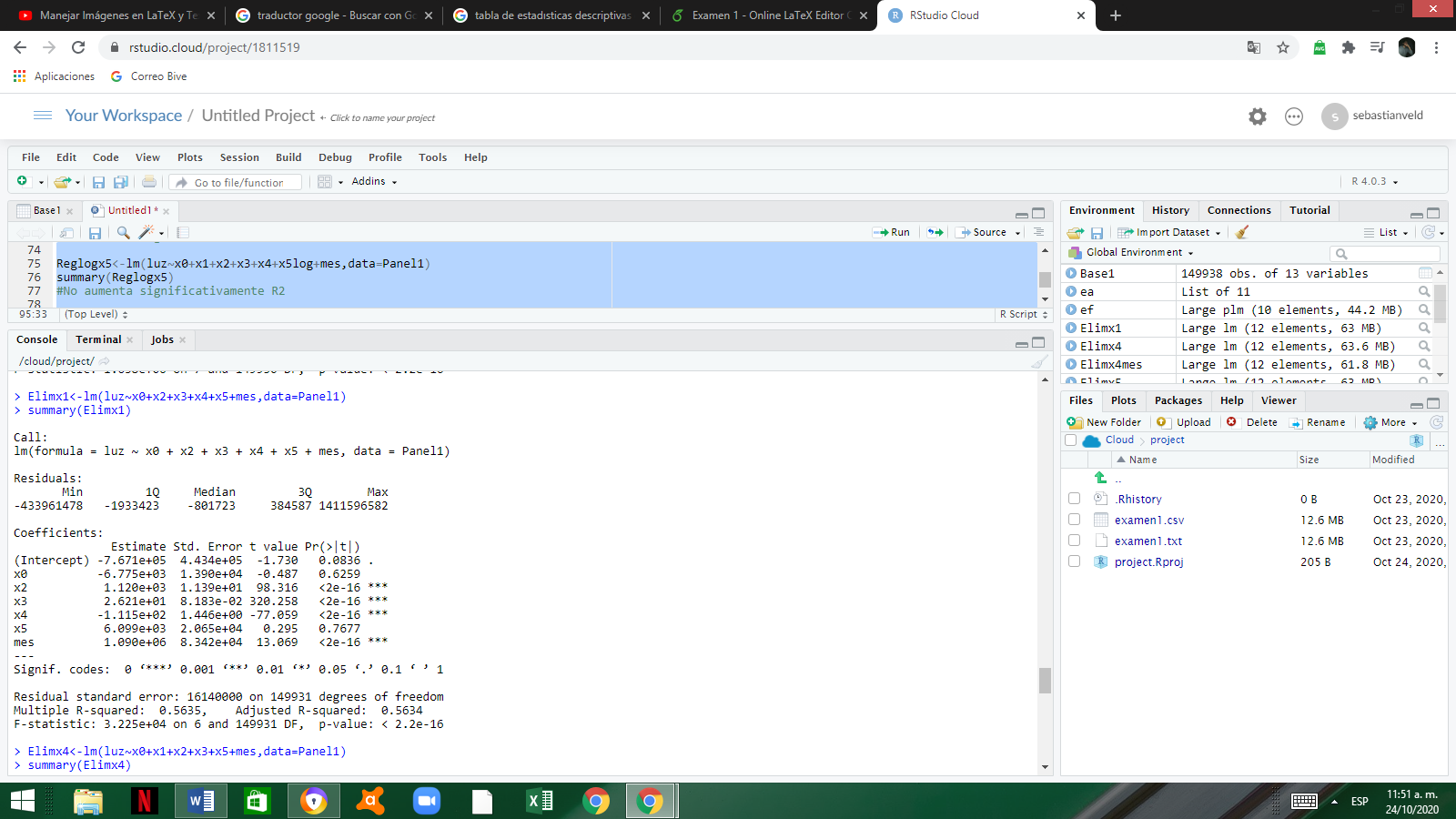


Tabla . Prueba 4

Disminuye considerablemente R2

***Elimx4<-lm(luz~x0+x1+x2+x3+x5+mes,data=Panel1)***

***summary(Elimx4)***

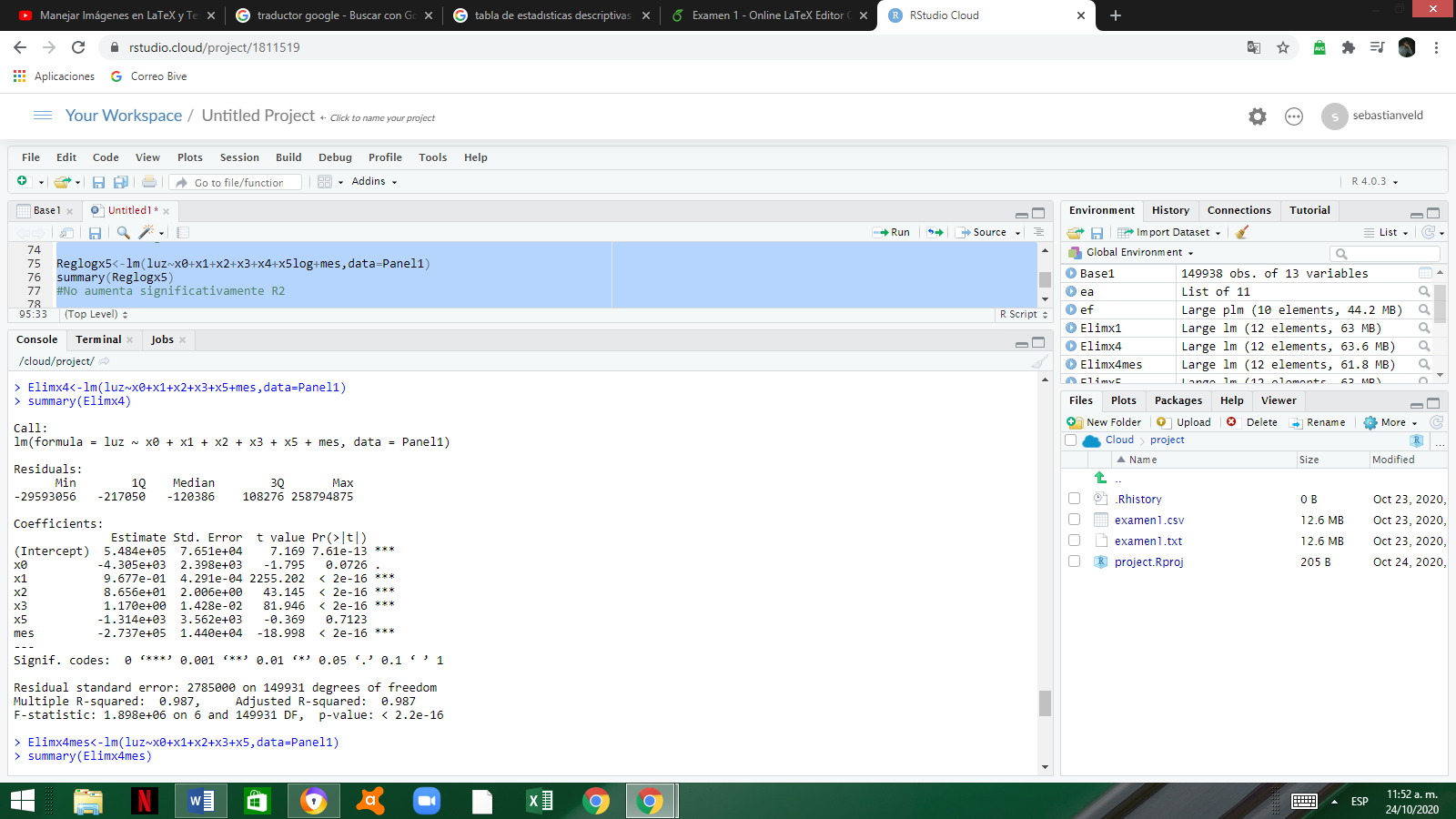


Tabla . Prueba 5

No aumenta R2

***Elimx4mes<-lm(luz~x0+x1+x2+x3+x5,data=Panel1)***

***summary(Elimx4mes)***

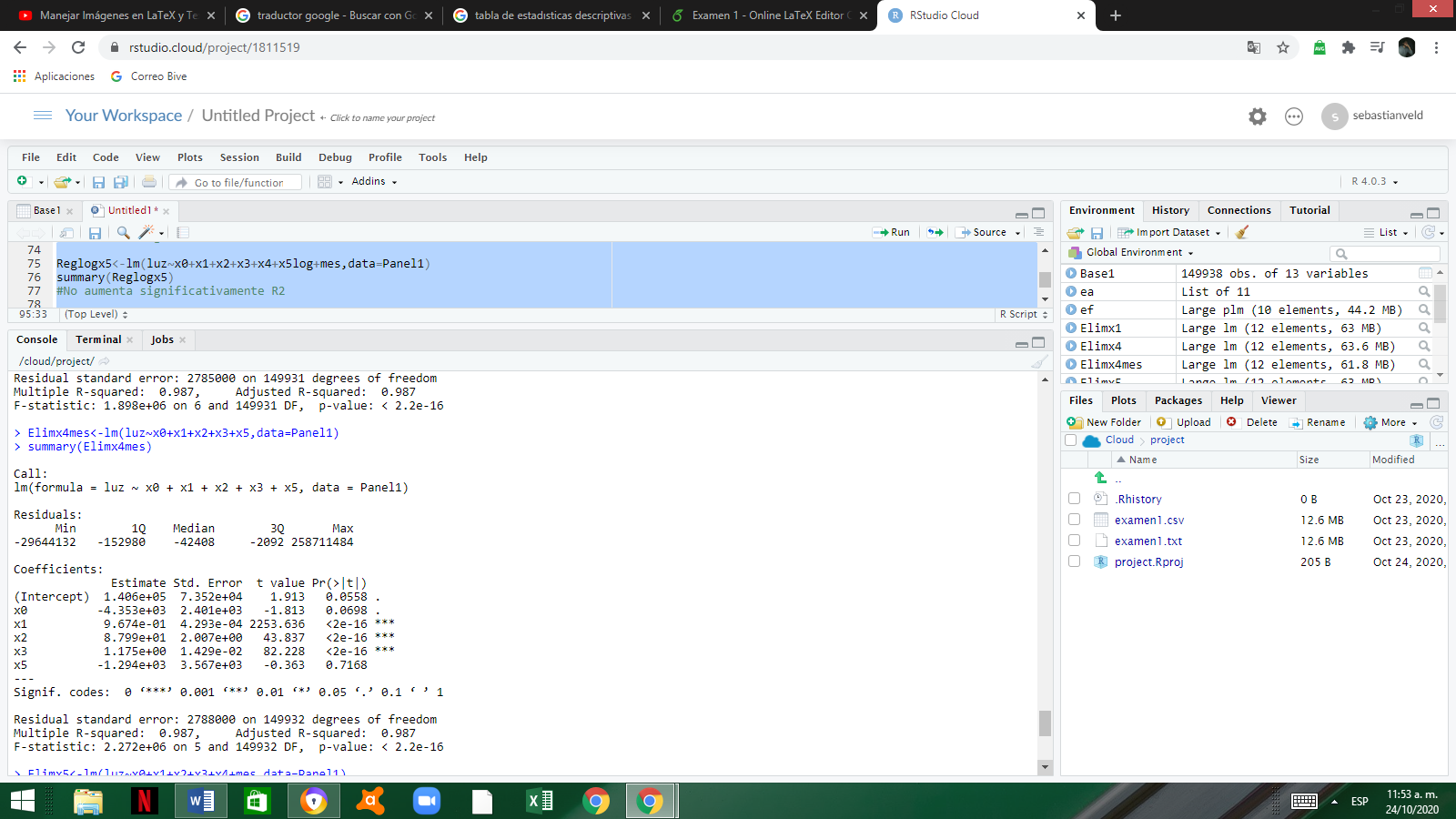


Tabla . Prueba 6

No aumenta R2

***Elimx5<-lm(luz~x0+x1+x2+x3+x4+mes,data=Panel1)***

***summary(Elimx5)***

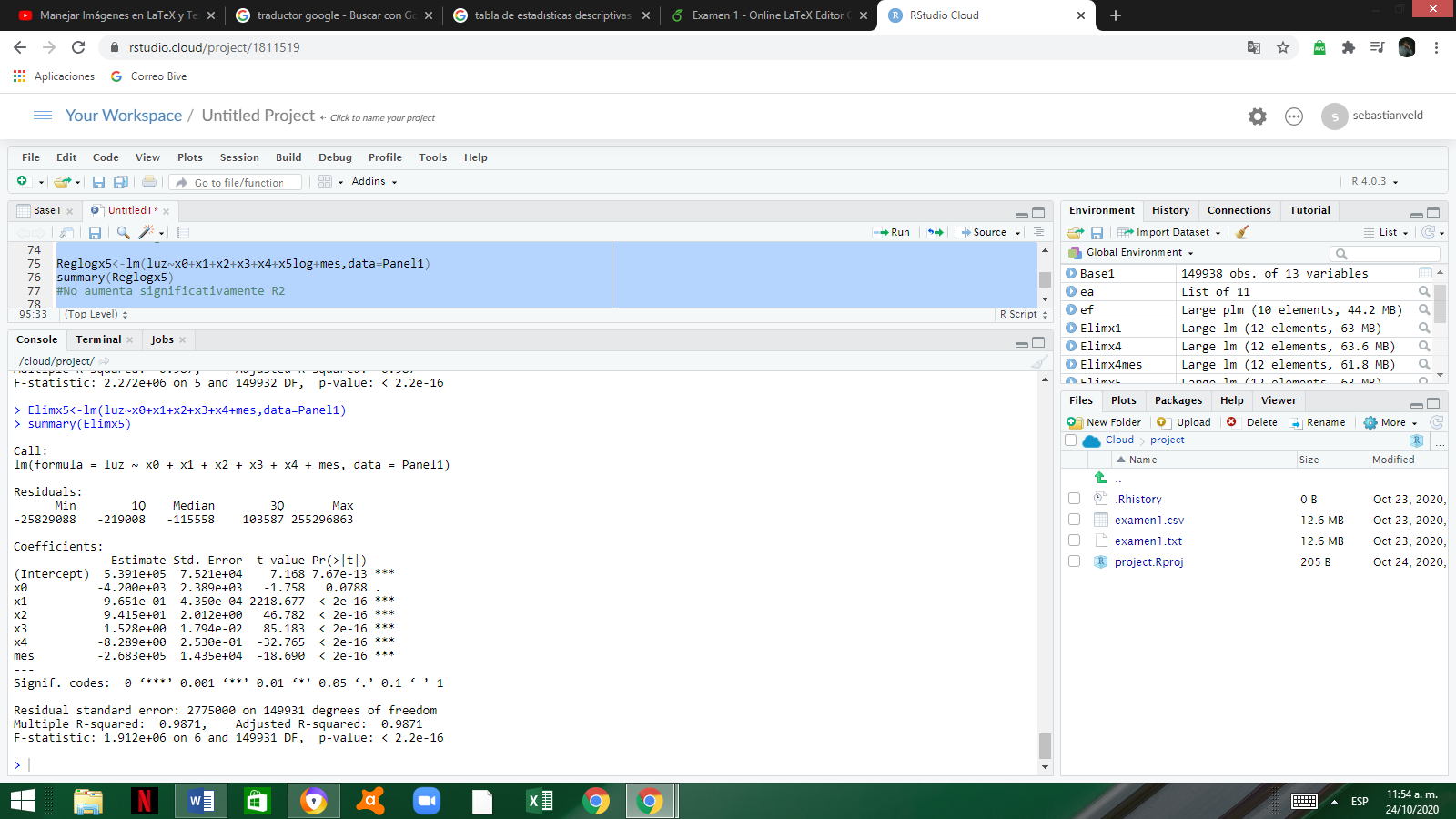


Tabla . Prueba 7

No aumenta significativamente R

Con ninguna de las dos alternativas se logra incrementar la R cuadrada, sin embargo se observa que eliminando algunas variables específicas (x1 por ejemplo) dicho valor disminuye significativamente. Sin embargo, considero que la eliminación de variables no es una opción aceptable, mientras que utilizar algunas variables en logaritmo para eliminar el efecto de las unidades de las variables sobre los coeficientes, si lo es, pero no representan estas ningún cambio considerable en el resultado de R cuadrada.

**EXAMEN, CODIGO**

