**Laboratorio 2**

**Explicación del caso:**

Revisar el archivo SPSS “Uso App”. Corresponde a una encuesta sobre el uso de una App (intención de continuar usando la app), donde se ha propuesto que los factores que llevan a una persona a usar esta app son: Usefulness (percepción de utilidad de la app), Enjoyment (percepción de disfrute con la app), Satisfaction (satisfacción con la app), Género, Edad, Actividad Principal.

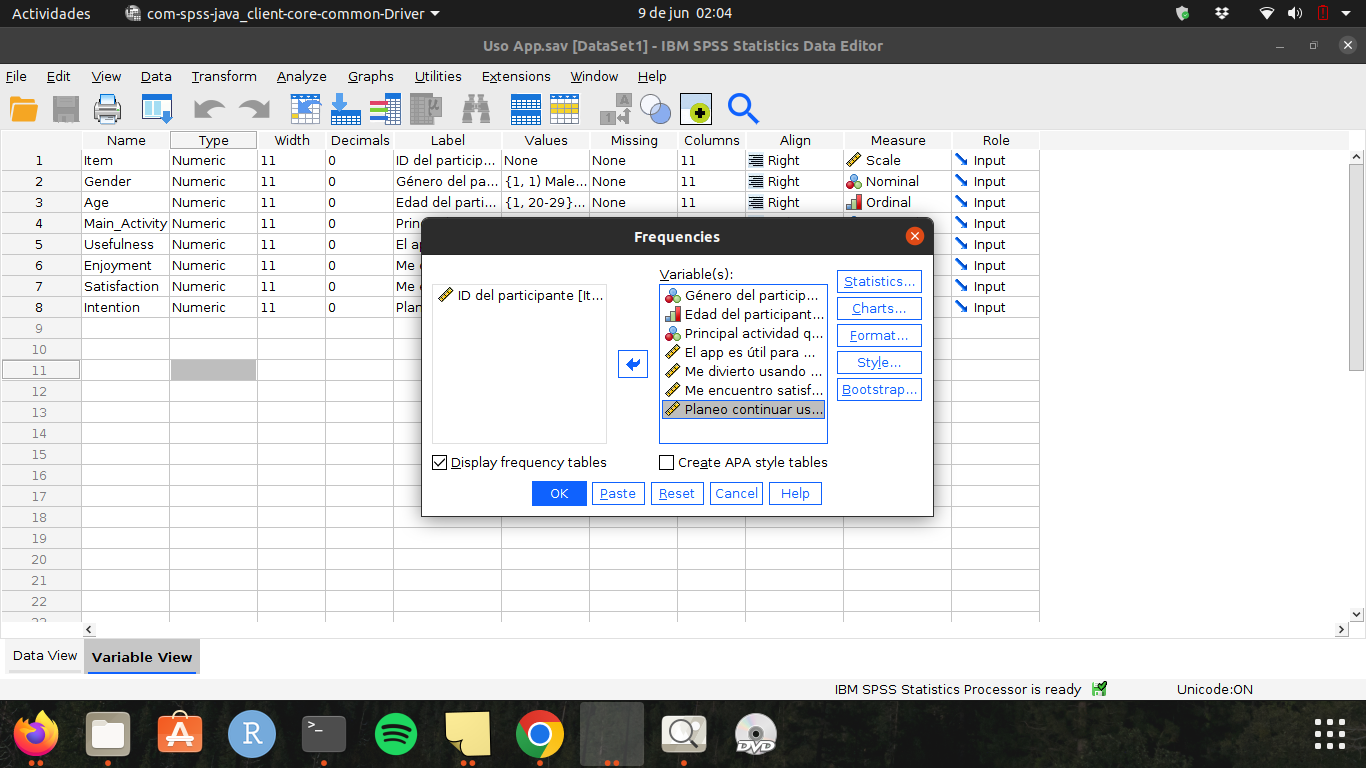
Dado esta propuesta, se nos pide determinar si hay evidencia de que estos factores propuestos tienen un impacto estadísticamente significativo en nuestra intención de continuar usando la app. Para ello, se nos pide que:

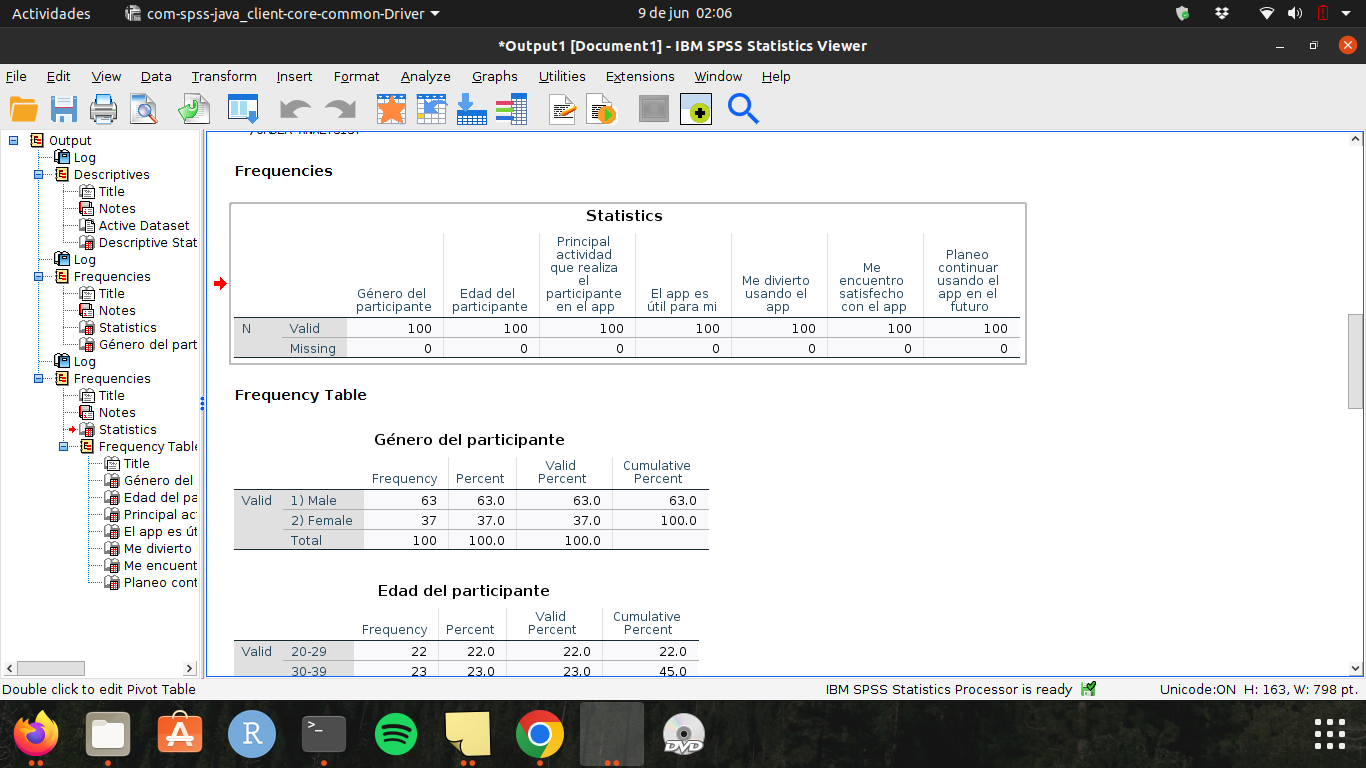
1. Revisar que todas las variables (métricas y no métricas) estén correctamente codificadas en números. De no ser así, se deberá recodificar de acuerdo a lo visto en clase antes de correr la regresión lineal. Por favor, dar detalles del proceso de recodificación de cada una de las variables que haya sido necesario recodificar. Finalmente, establecer el modelo final de regresión lineal à 3 ptos

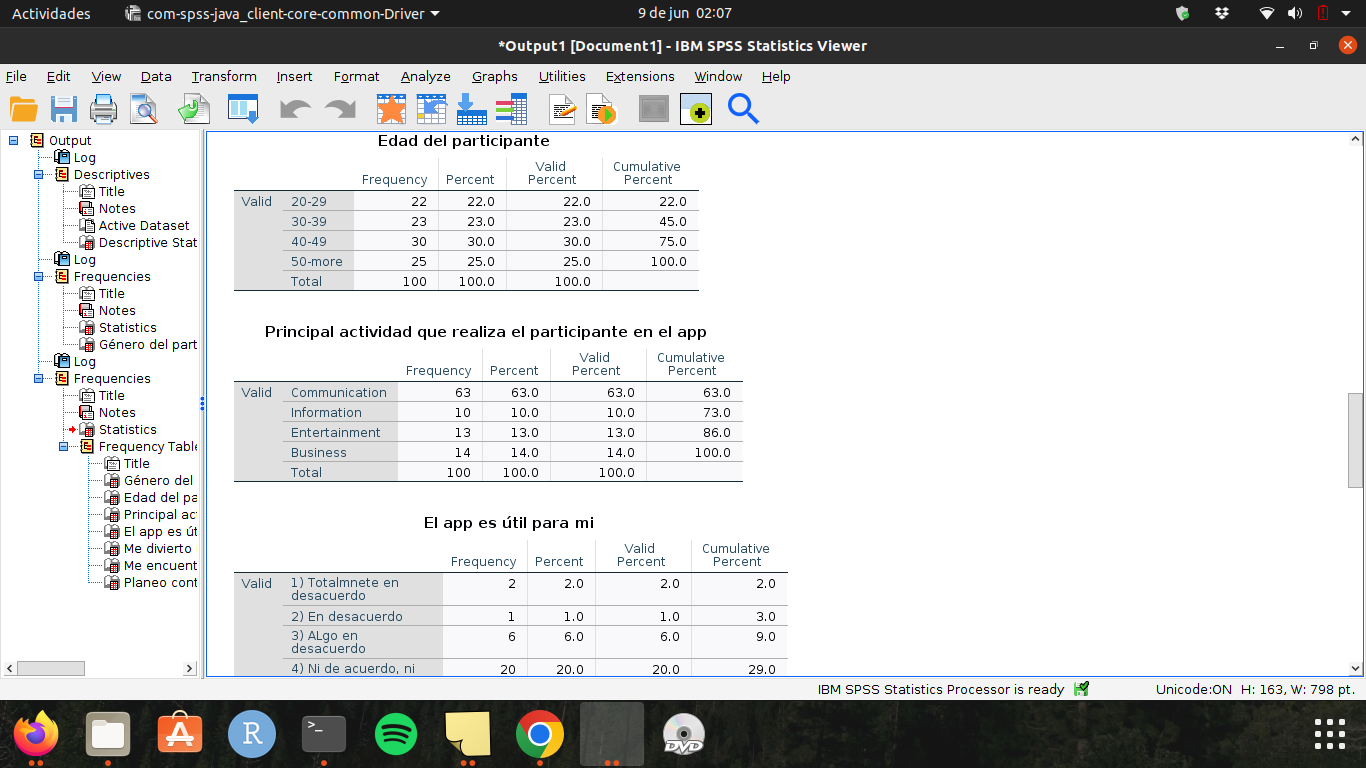
**Nota:** En SPSS se puede recodificar una variable. Ver el siguiente tutorial:

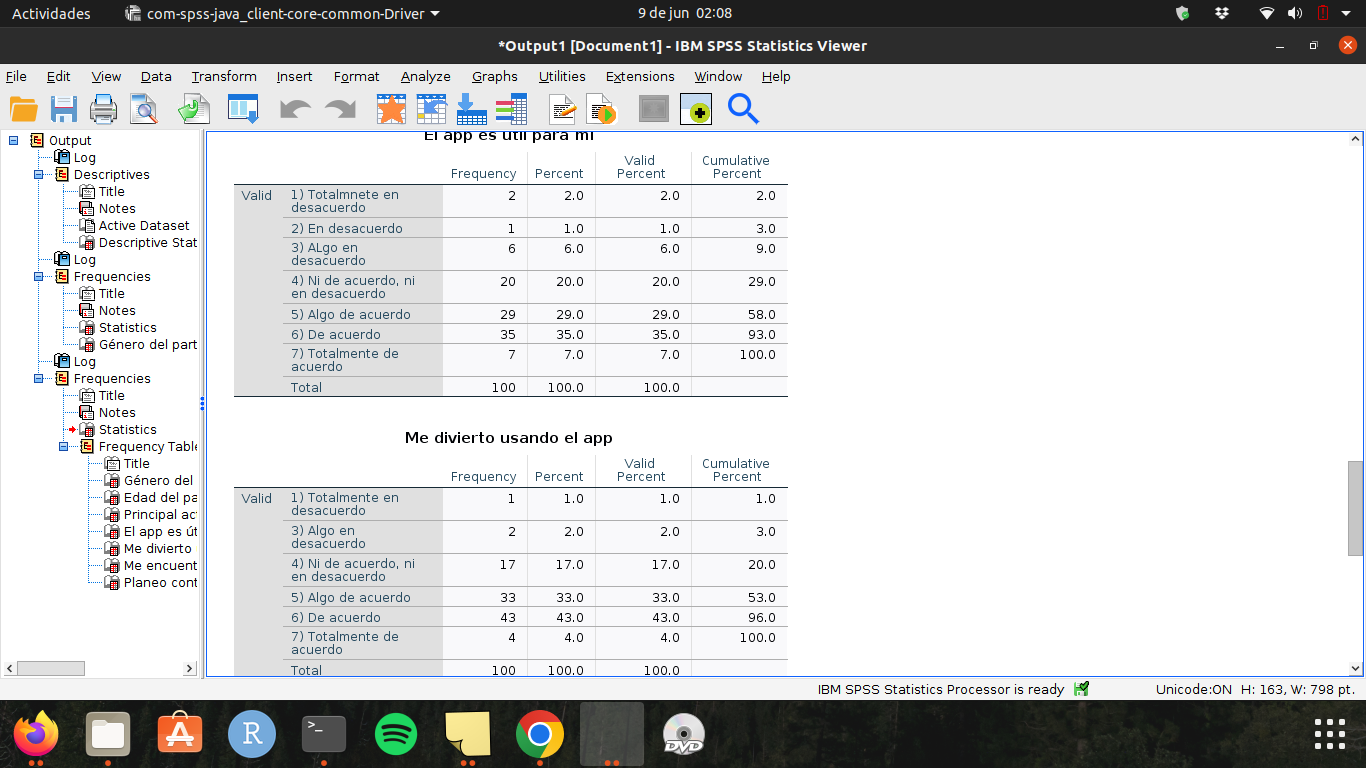
<https://www.youtube.com/watch?v=PBB5tCYxQtc>

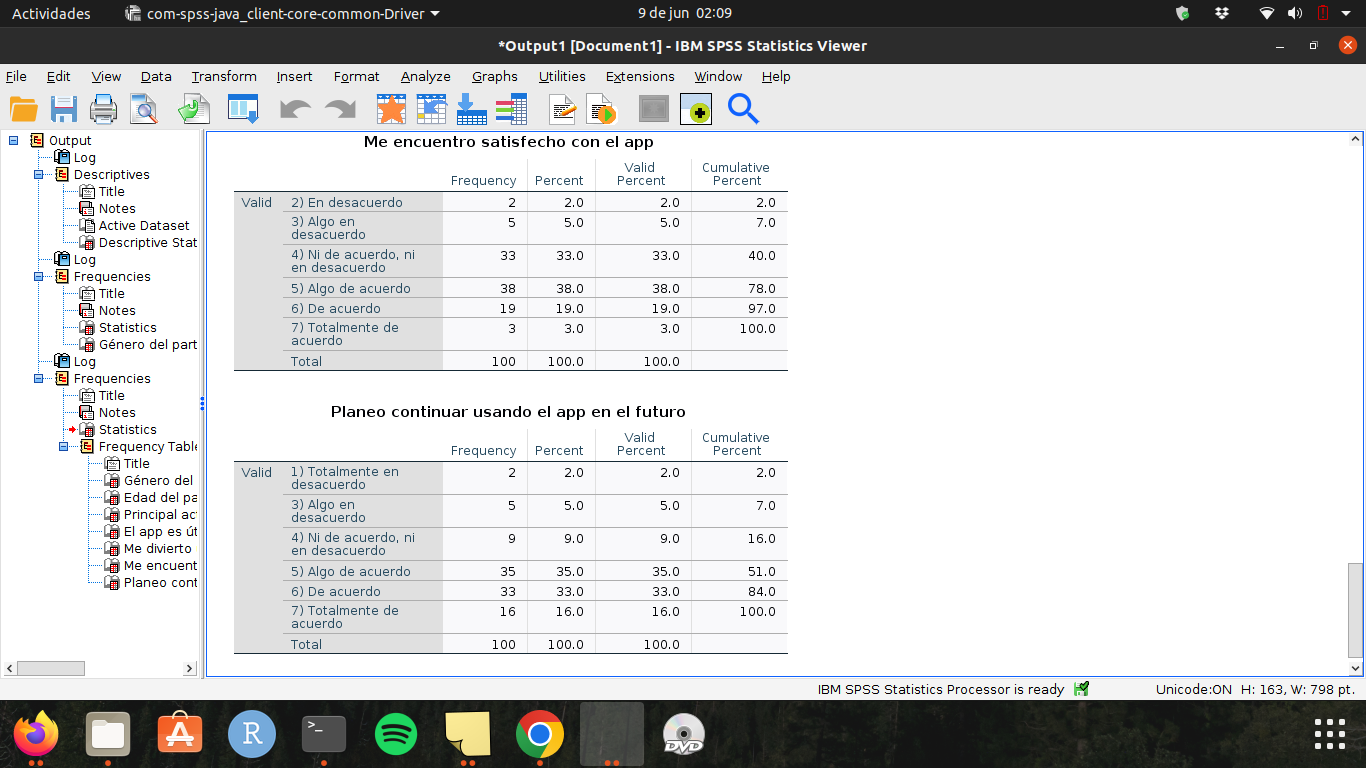
Exploración inicial

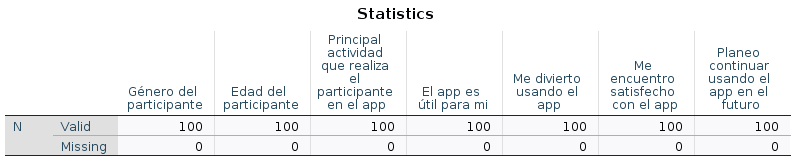




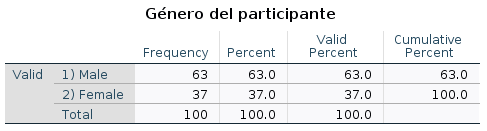






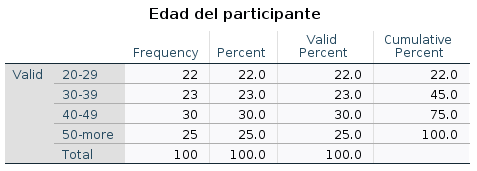


Género

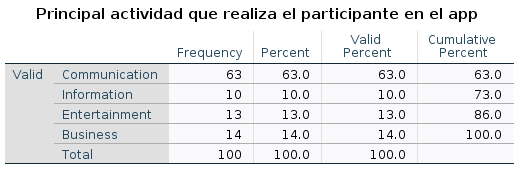


Se observa codificación 1 = Male, 2 = Female

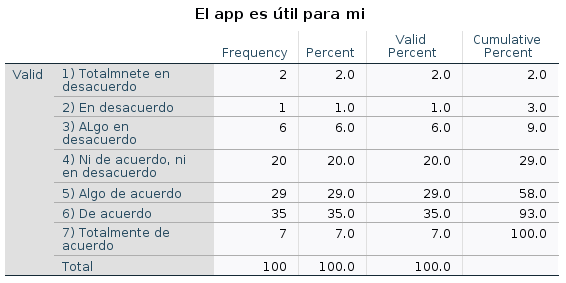
Edad



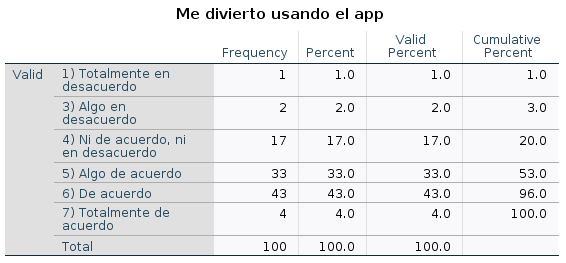
Principal actividad que realiza el participante



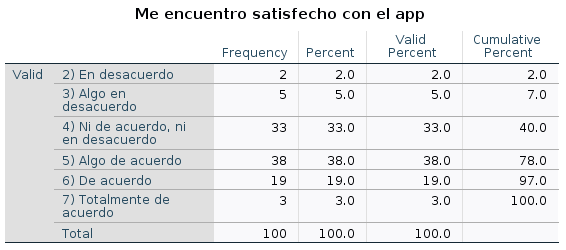
El app es útil



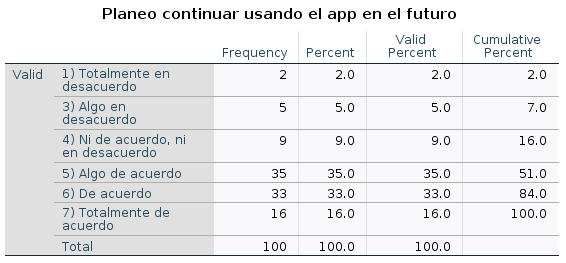
Me divierto usando el app



Me encuentro satisfecho



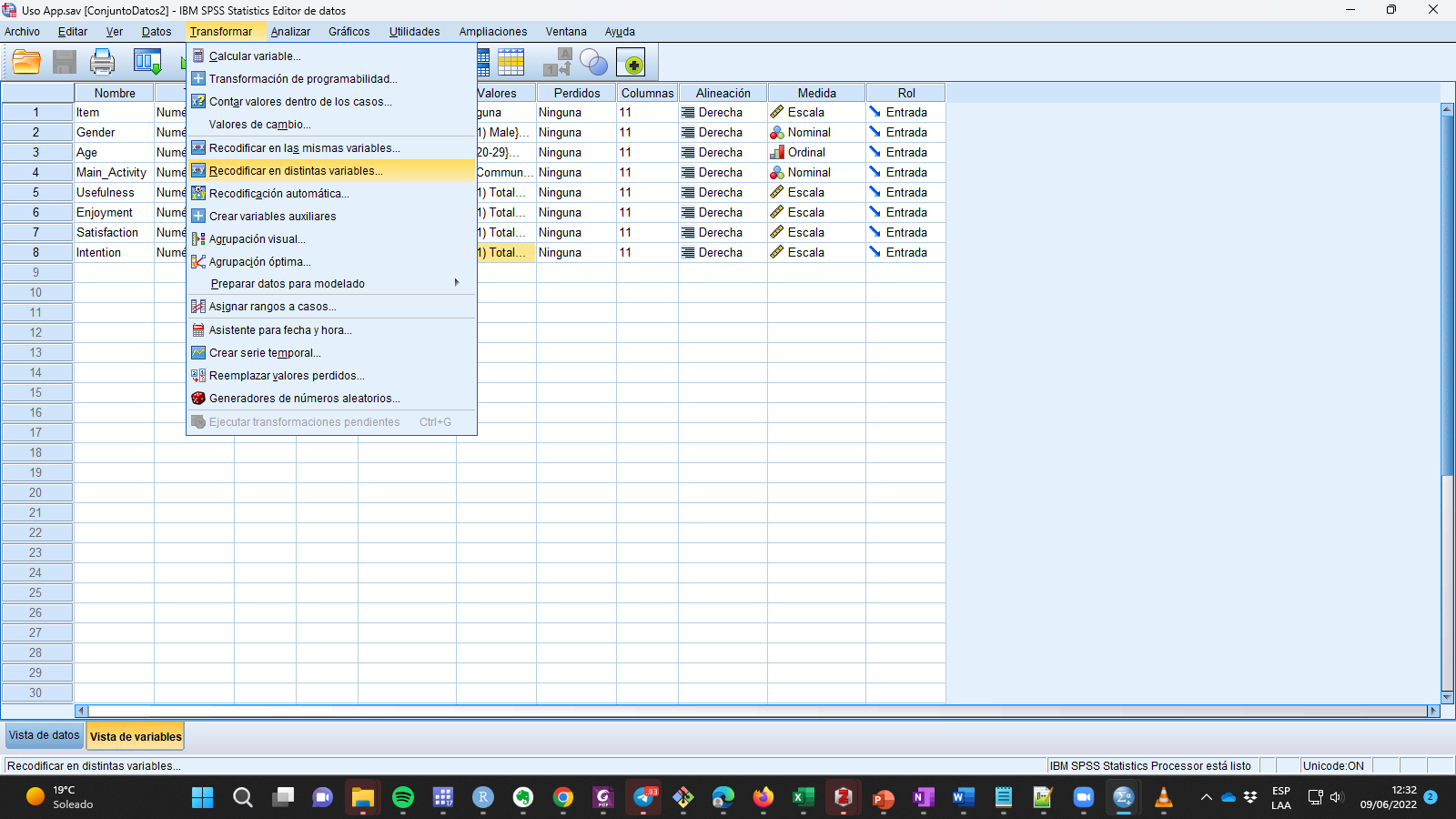
Planeo continuar usando la app



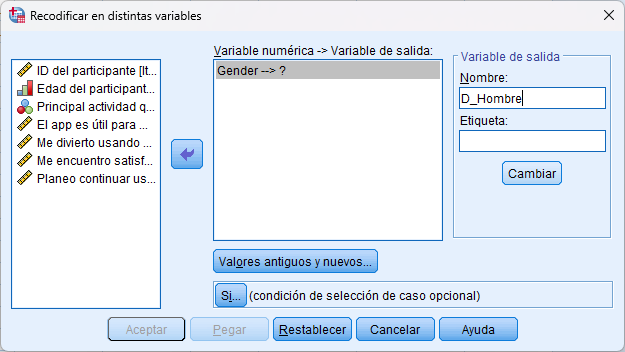
Se observa que hay tres variables categóricas que deberían ser recodificadas como dummy, toda vez que sus valores no se relacionan con alguna magnitud y si se incorporan al modelo como están van a generar distorsiones en la interpretación respecto a su influencia en la variable dependiente. Estas variables son (a) Género, (b) Edad, y (c) Actividad Principal.

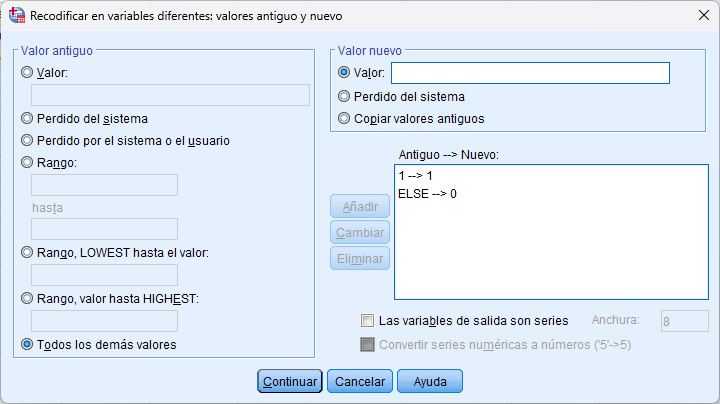
A continuación, el procedimiento para su recodificación como variables dummy:

En primer lugar, se elige la opción transformar/recodificar en distintas variables. Este paso es común para las tres variables categóricas mencionadas:



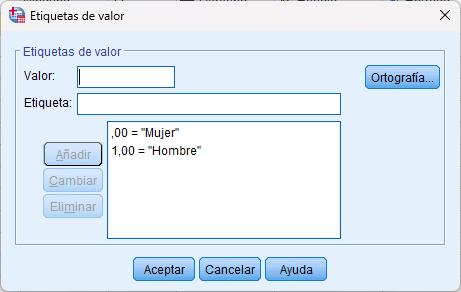
1. En el caso del género, se hizo lo siguiente:

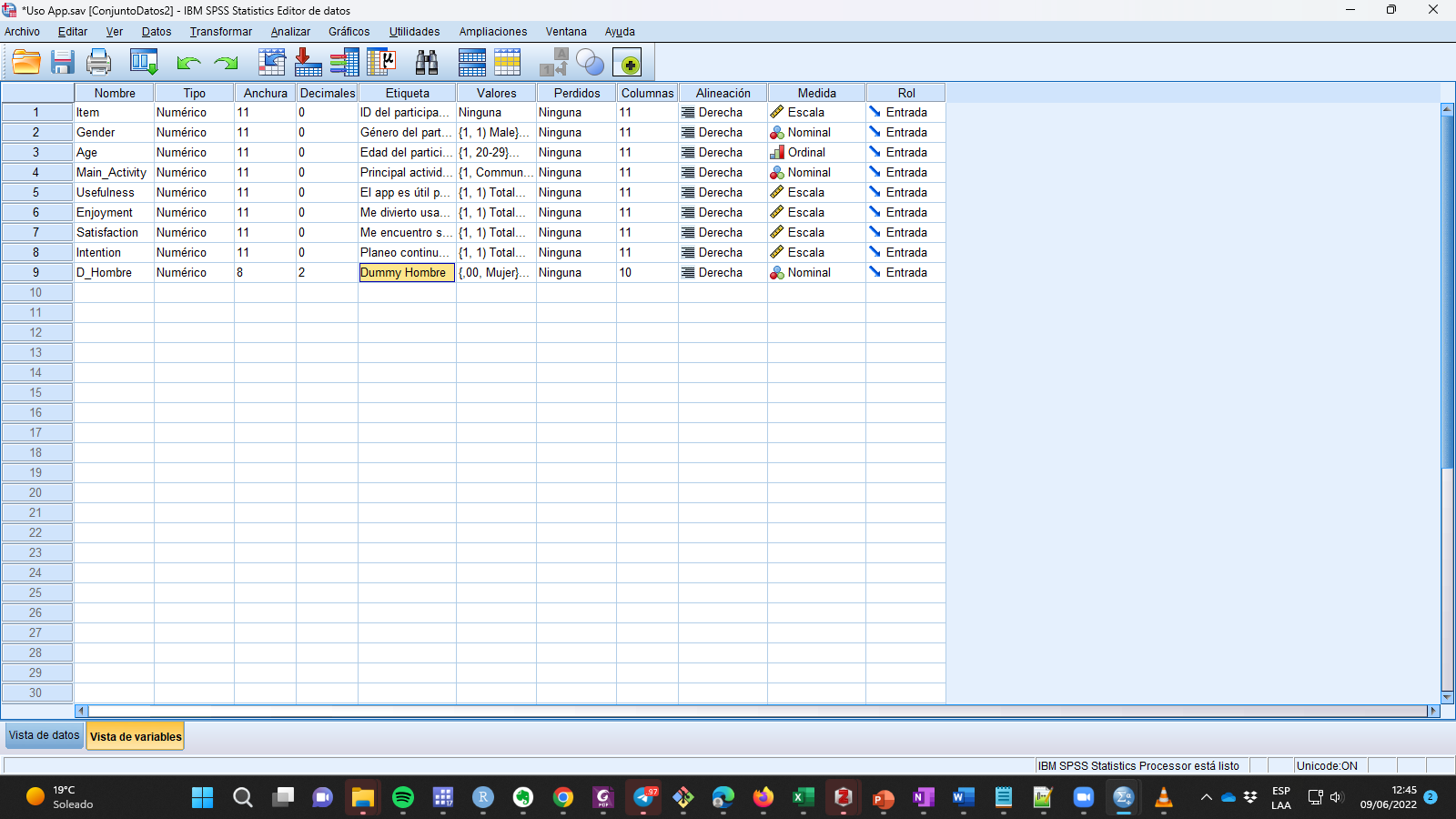




Cuando es hombre toma el valor de 1, y cuando es mujer toma el valor de cero.

Luego se coloca los labels tanto a la variable como a los valores:

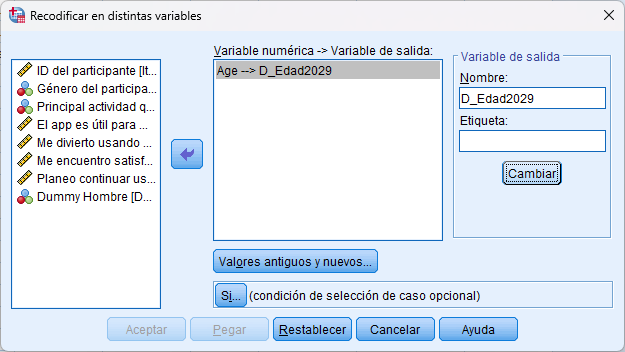




En este caso el género “mujer” es la categoría base.

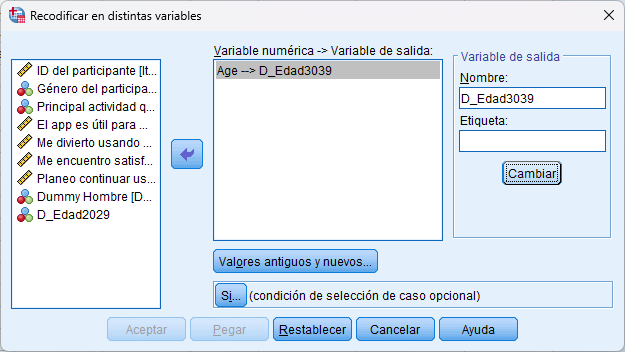
1. Para la edad se generaron 3 dummies, porque se tienen 4 categorías:

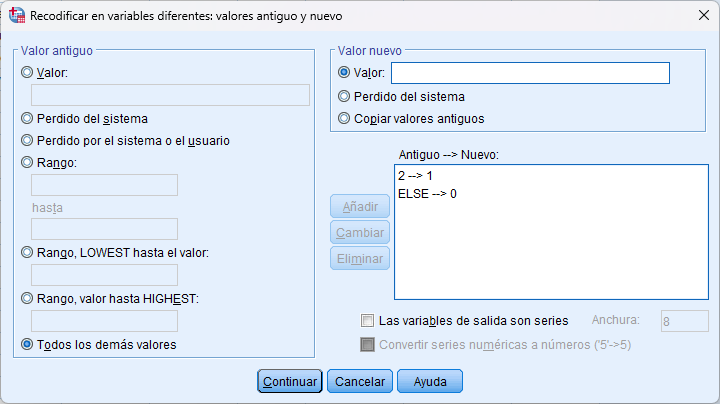
*Edad 20 a 29*



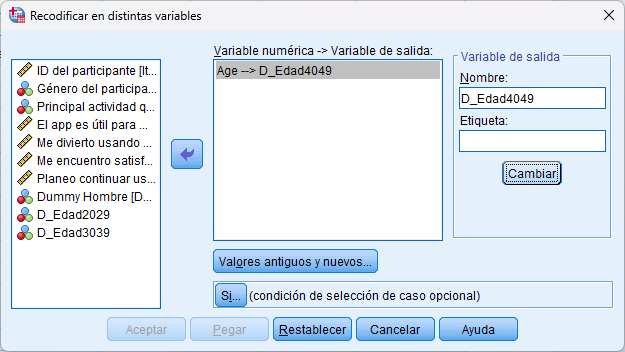


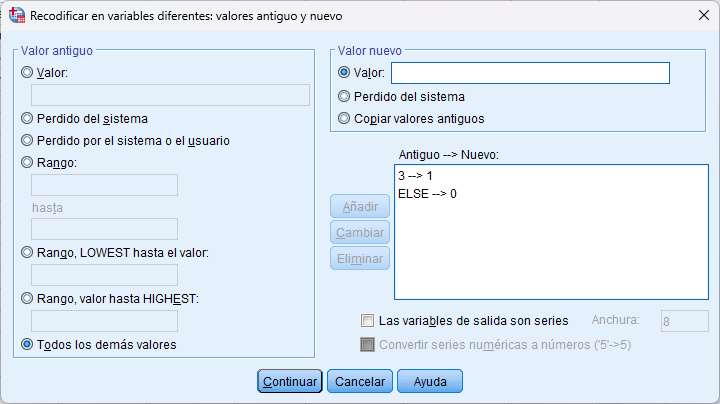
*Edad 30 a 39*



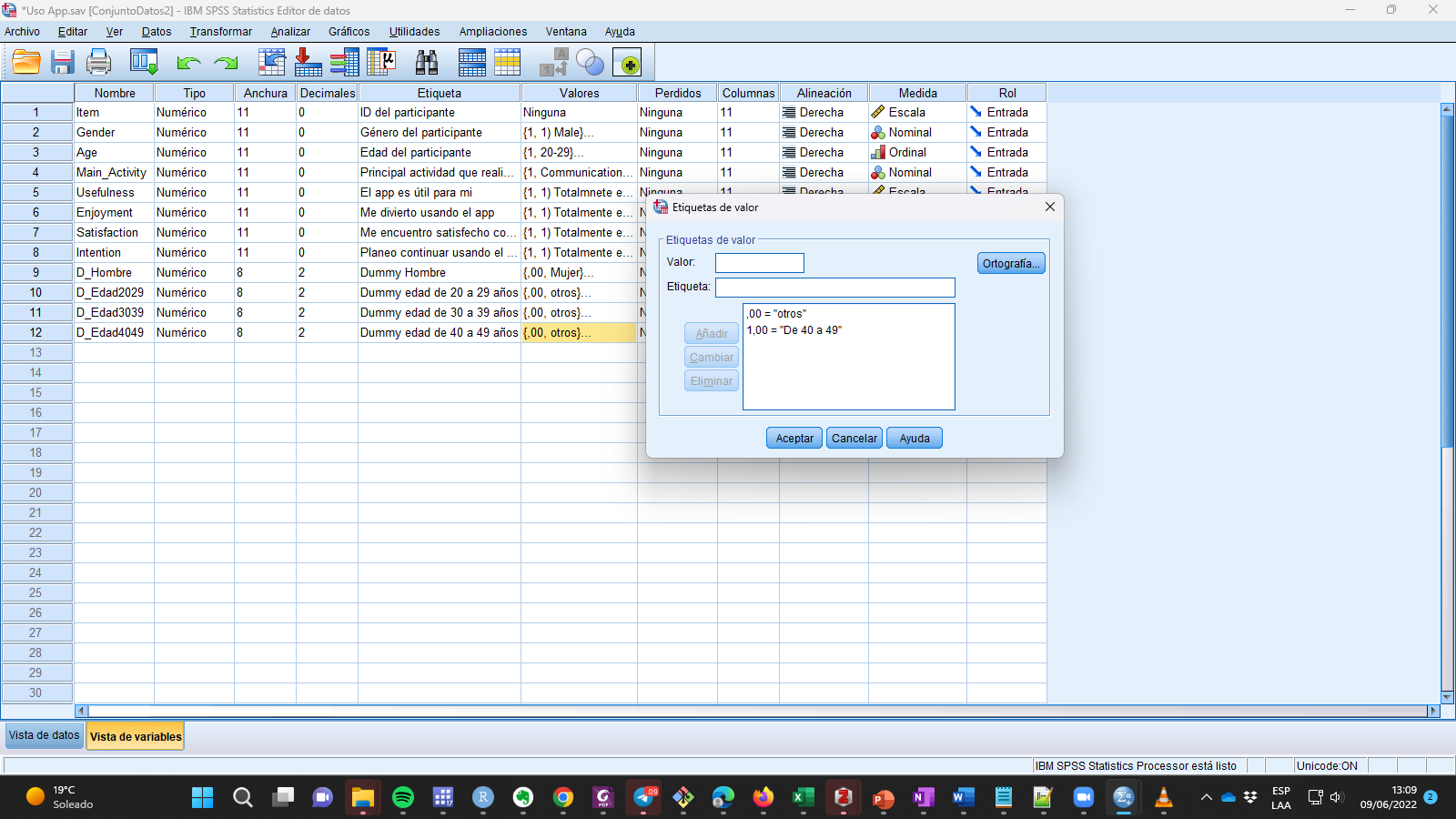


*Edad 40 a 49*



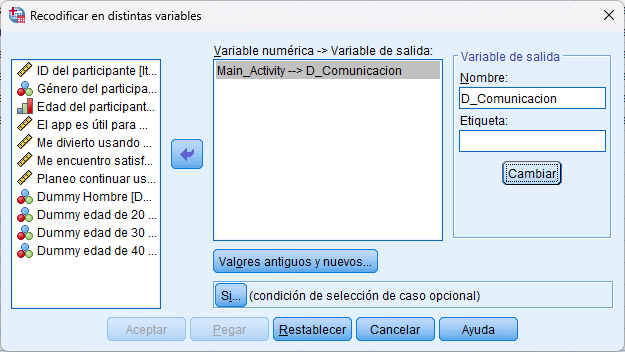


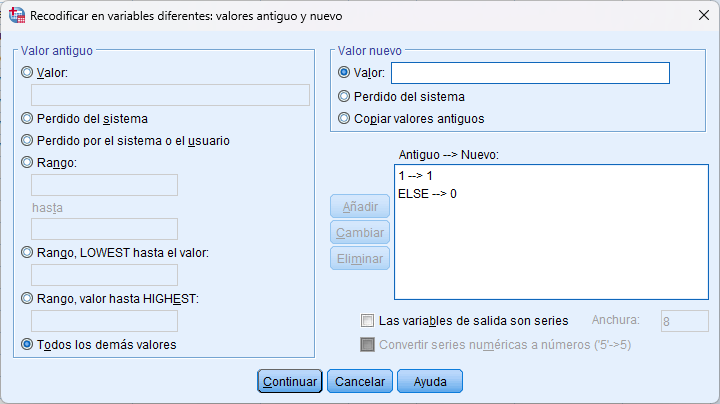
En este caso el rango de edad de 50 a más es la categoría base. Asimismo, se procedió a colocar los labels a nivel de variable y valores:



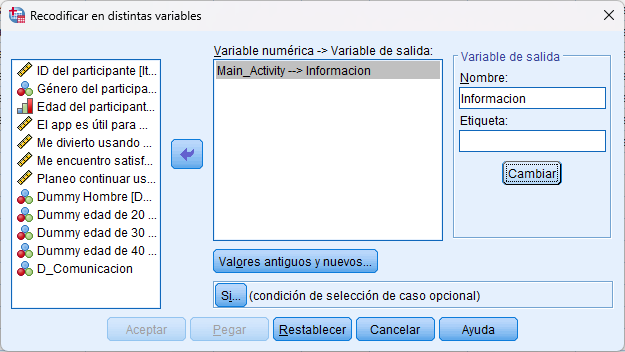
1. Finalmente, se generaron las dummies para la variable categórica “principal actividad”:

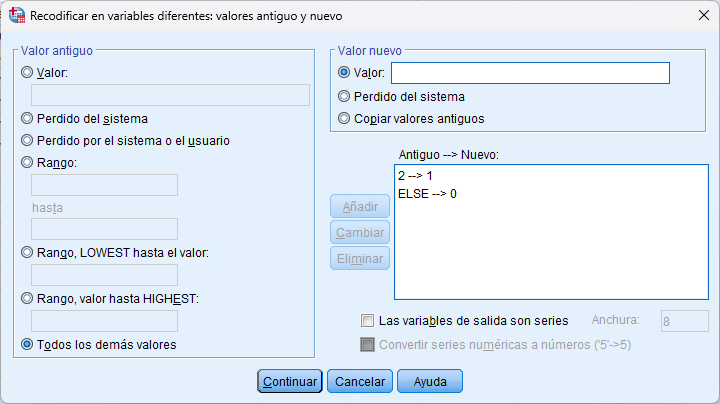
*Comunicación*



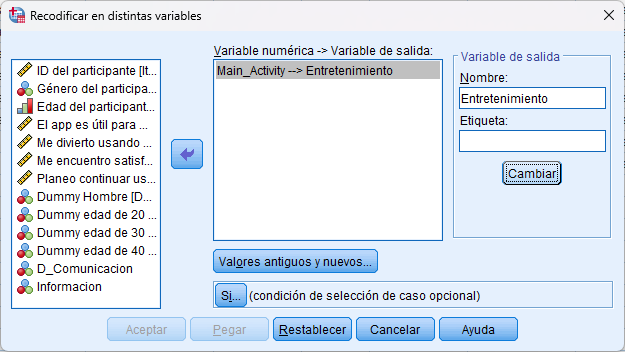


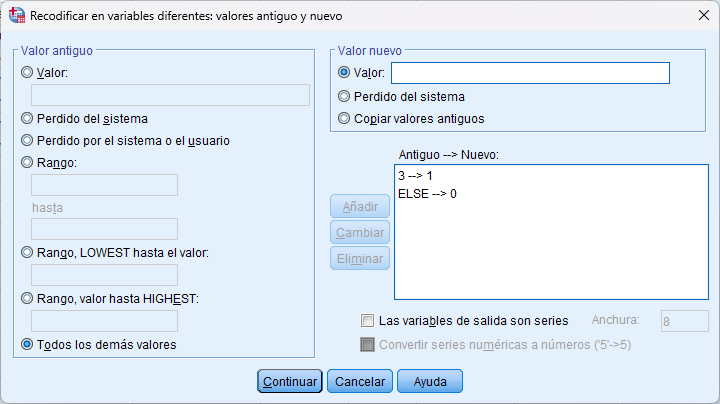
Información



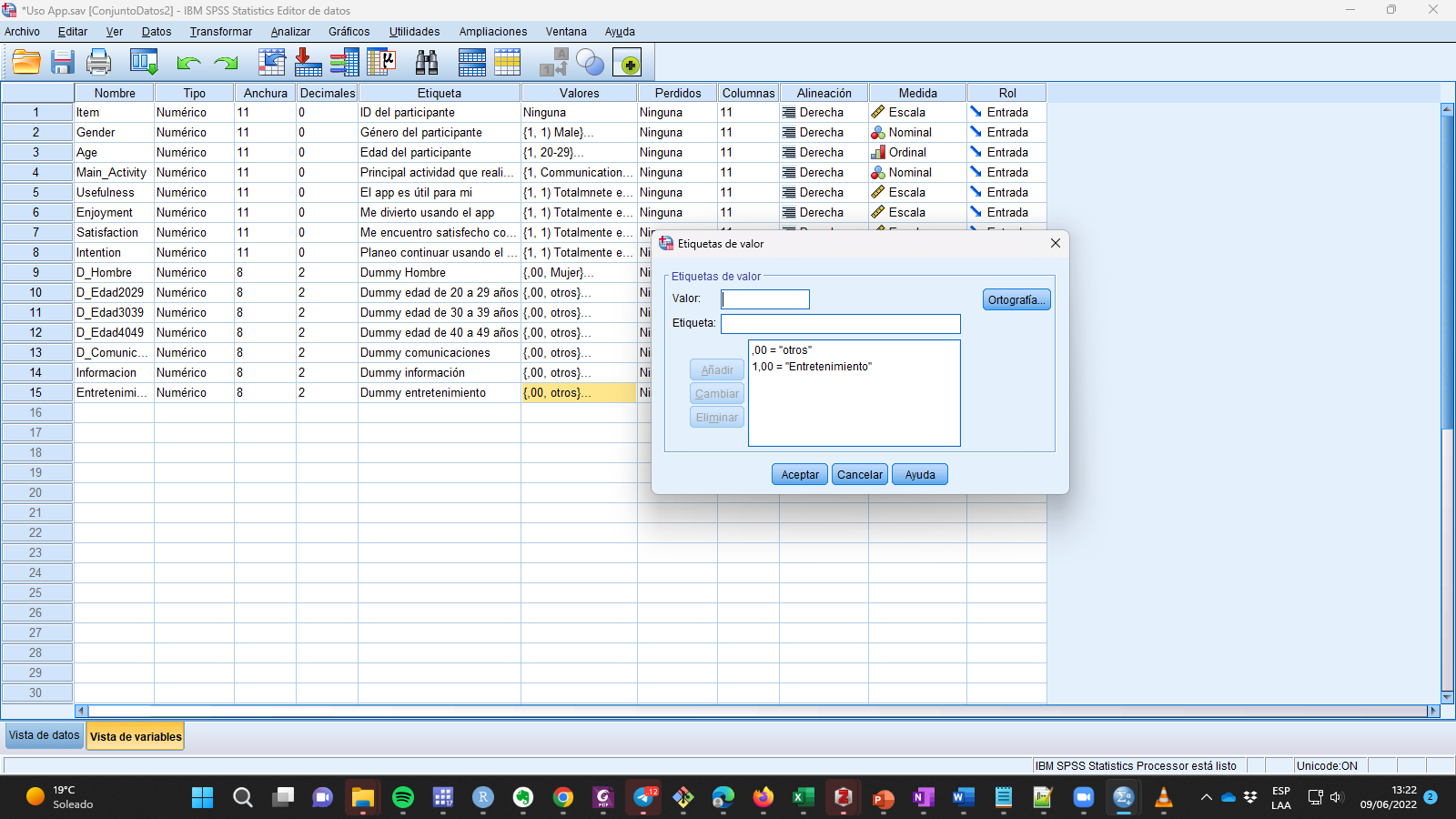


Entretenimiento





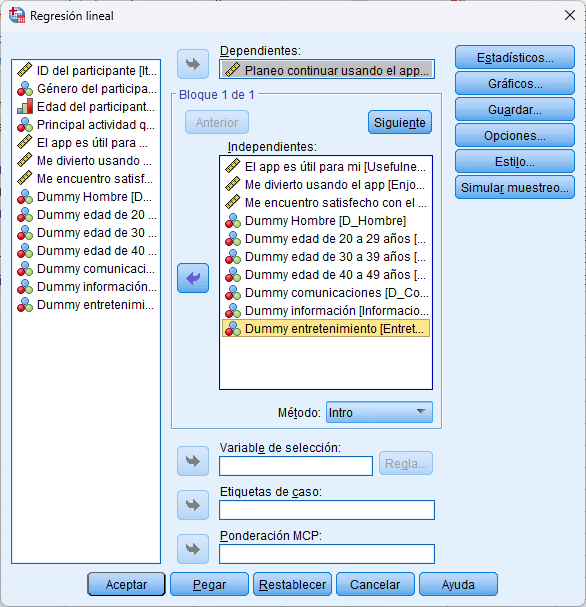
En este caso, la actividad Negocios es la categoría base. Luego, se colocaron las etiquetas a nivel de variable y valores:

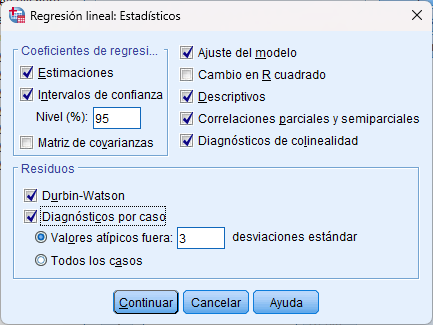


Con esto, el modelo empírico quedaría del siguiente modo:

1. Revisión de supuestos: a) errores no correlacionados, b) multicolinearidad, c) linearidad, d) homocedasticidad, e) ausencia de outliers, f) distribución normal de los errores (2 ptos c/u à 12 ptos en total). Si alguno de los supuestos NO se cumple, por el momento no importa, sólo menciónenlo y continúan con el análisis.

Para realizar los análisis solicitados se ingresa el modelo al SPSS:

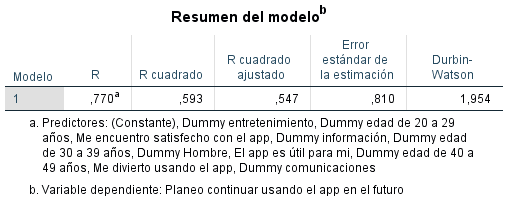




Supuestos:

a) Errores no correlacionados

Para cotejar la independencia de los errores recurrimos al Test de Durbin-Watson. Se observa que el indicador es cercano a 2[[1]](#footnote-3459), por lo cual podemos decir que no se observa errores correlacionados



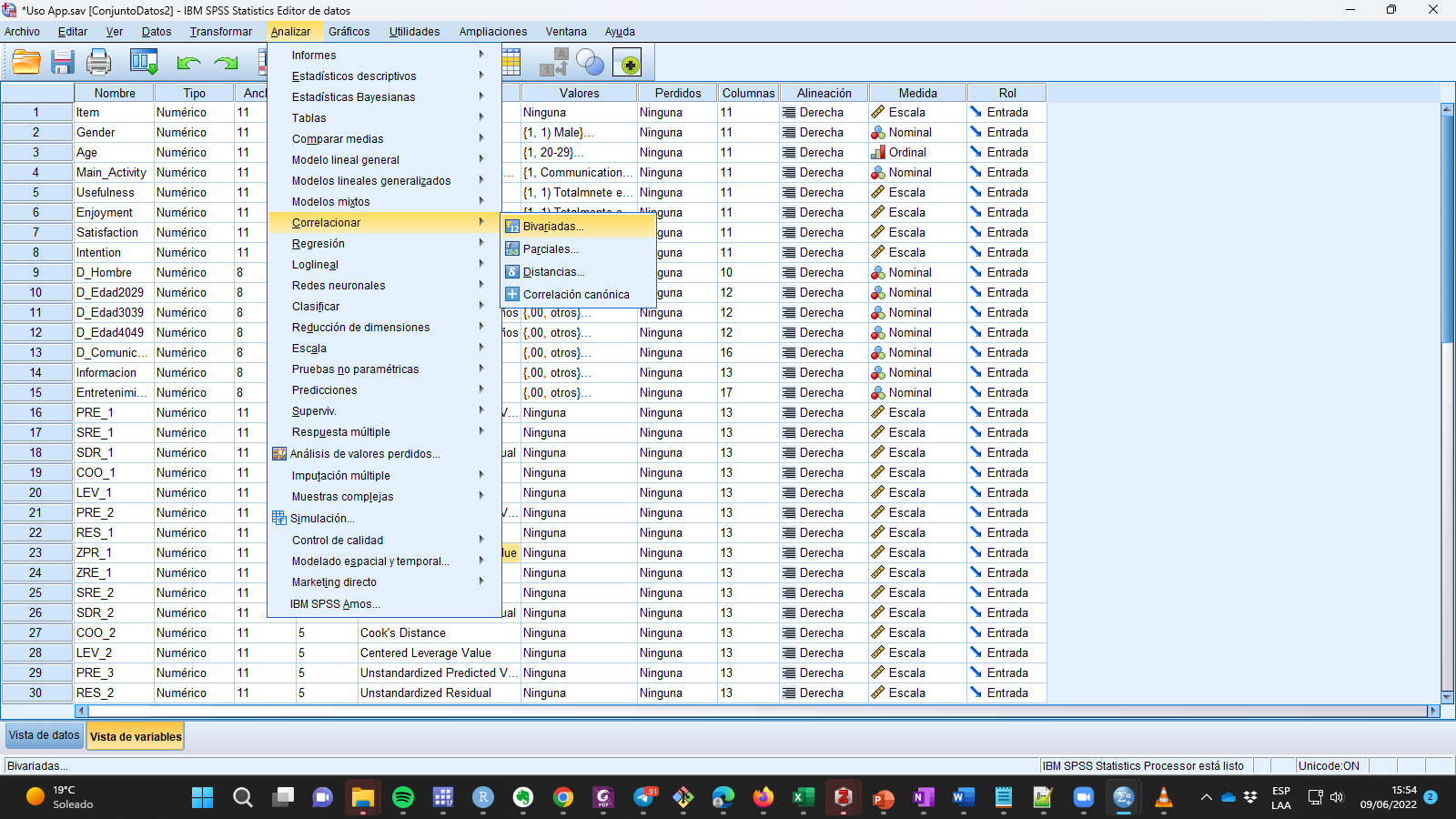
b) Multicolinealidad

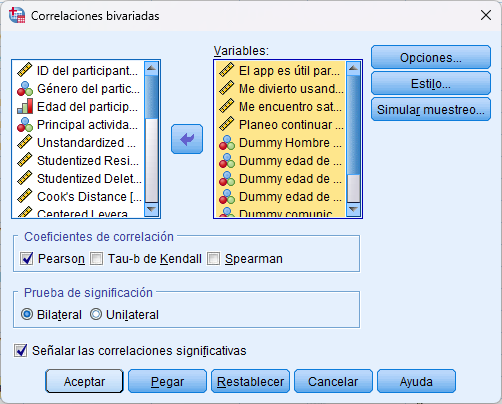


En este caso se comprueba si hay inflación de la varianza (VIF). El criterio es que sea menor o igual a 5. En este caso se observa que los valores VIF relacionado con cada variable predictora son menores a 5. Asimismo, el índice de tolerancia tiene valores mayores a 0.2. Por lo tanto, podemos concluir que no existe multicolinealidad.

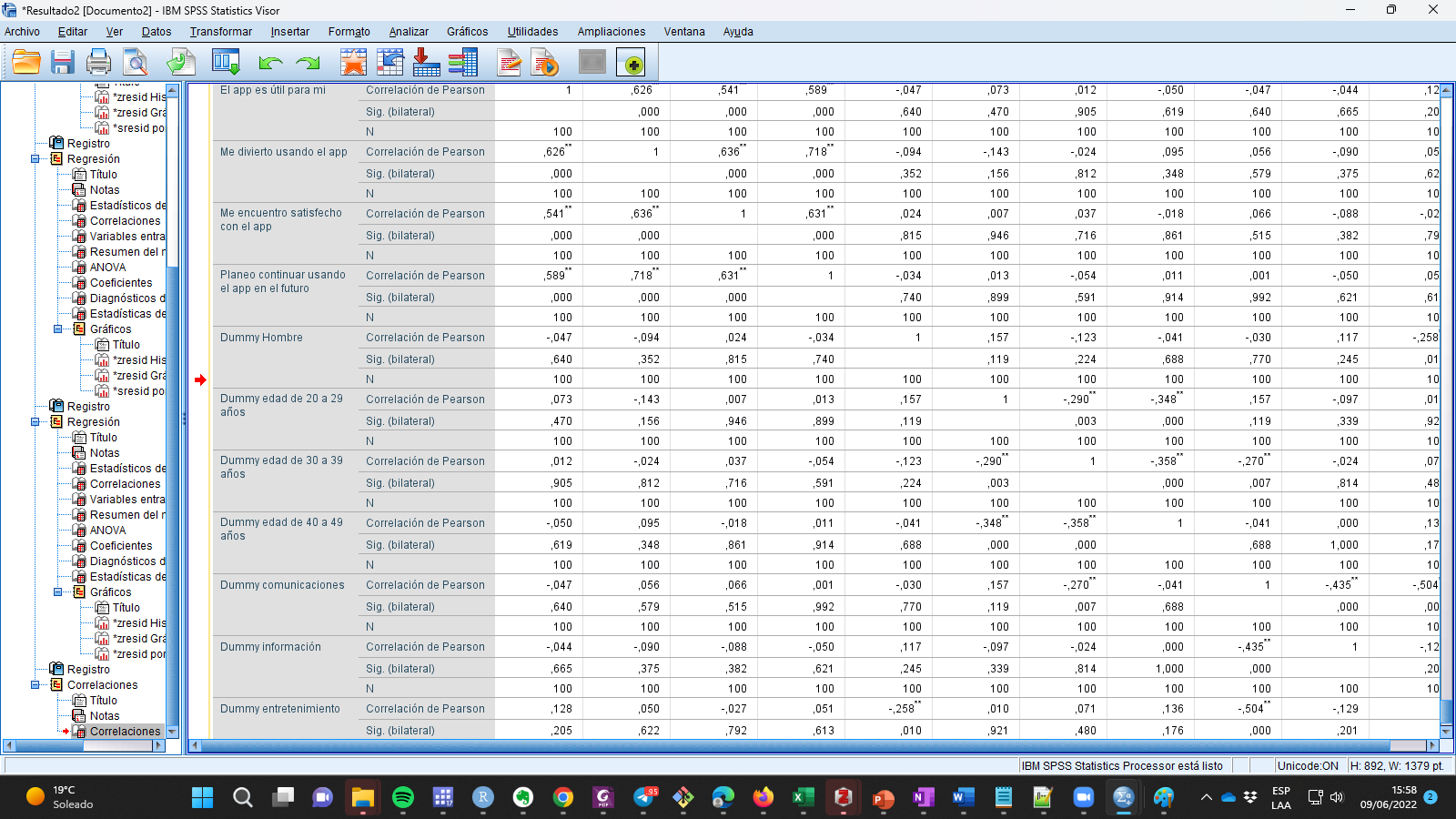
c) Linealidad

Para el supuesto de linealidad verificamos las correlaciones entre las variables:

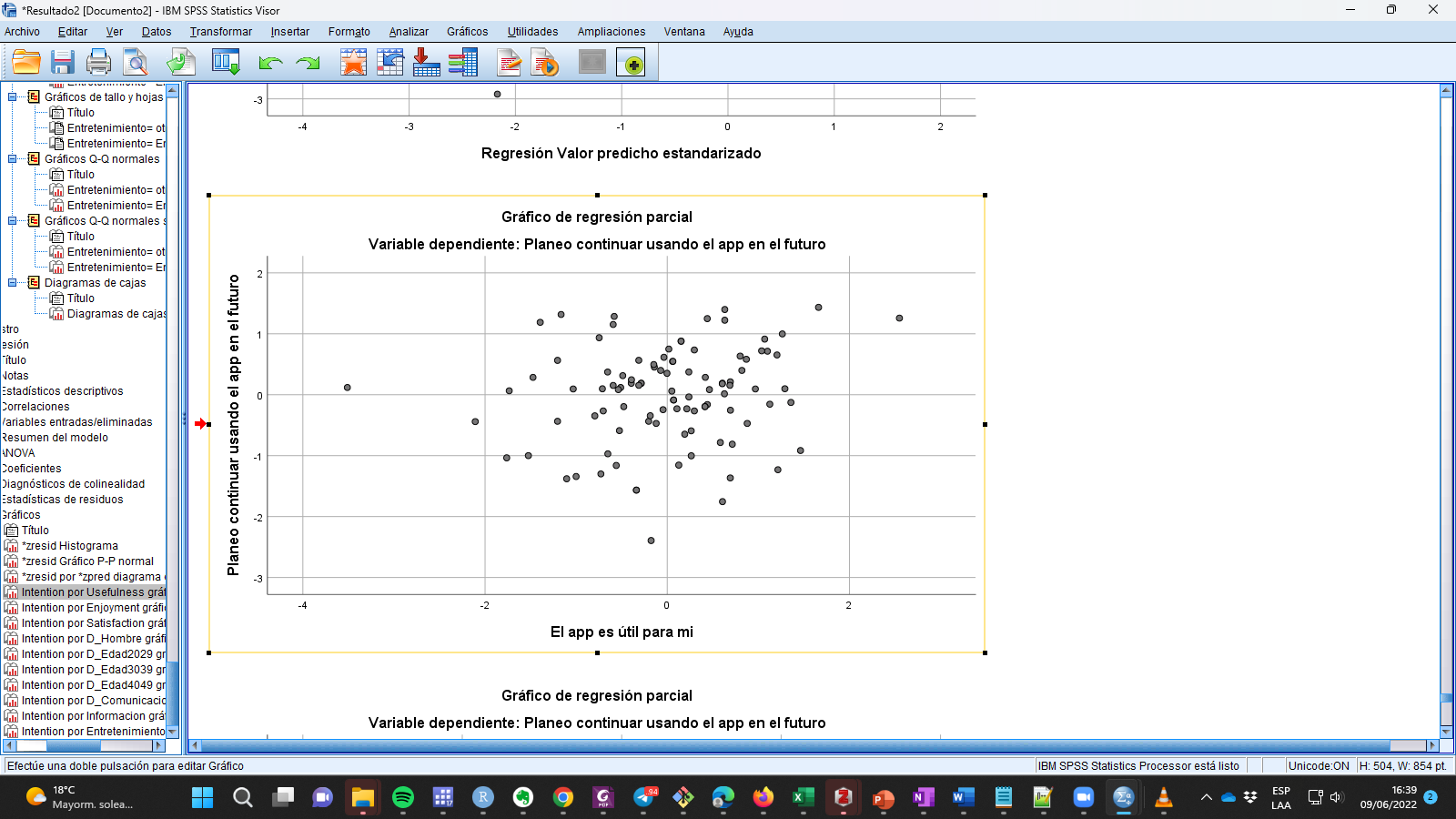


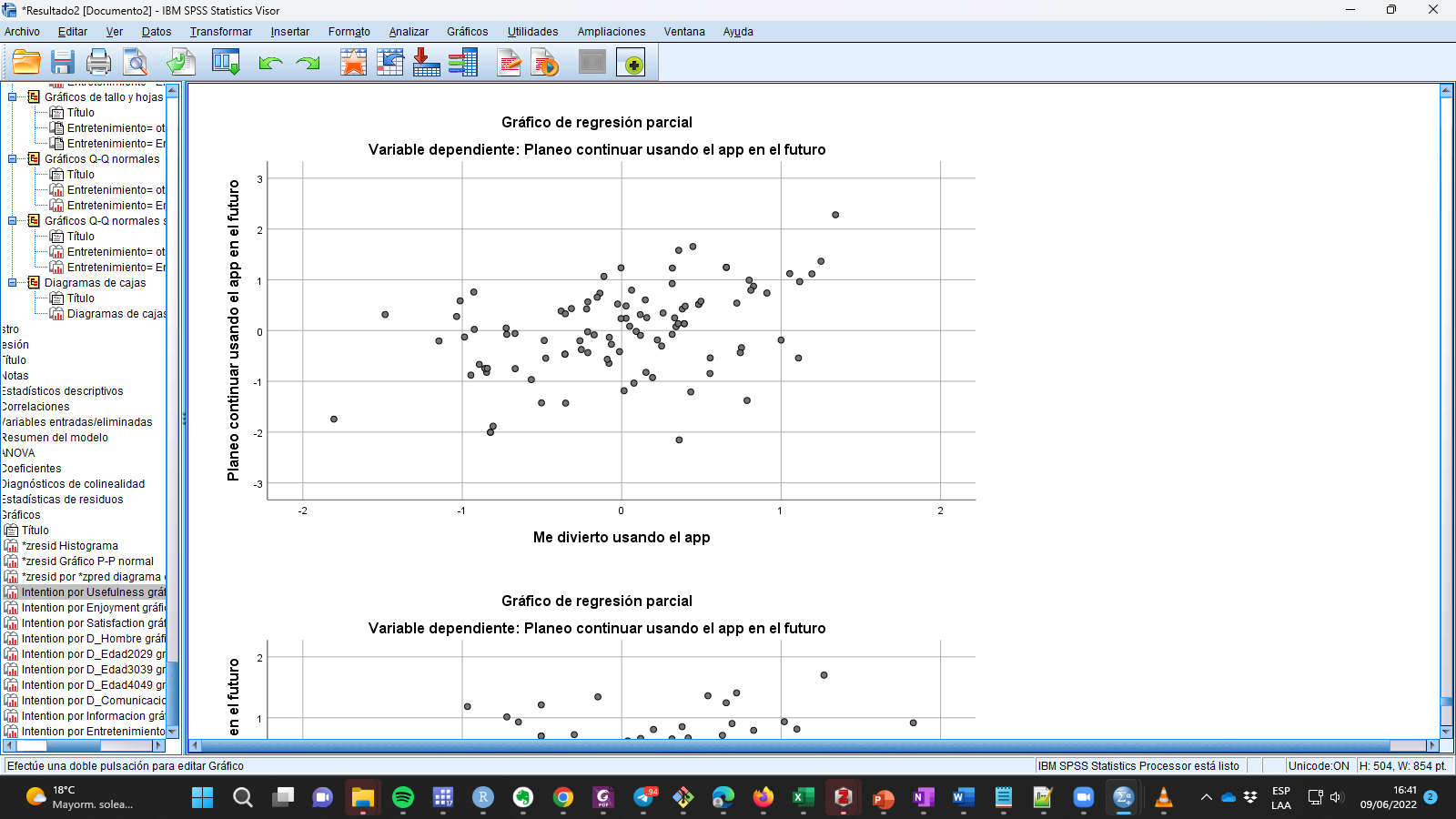


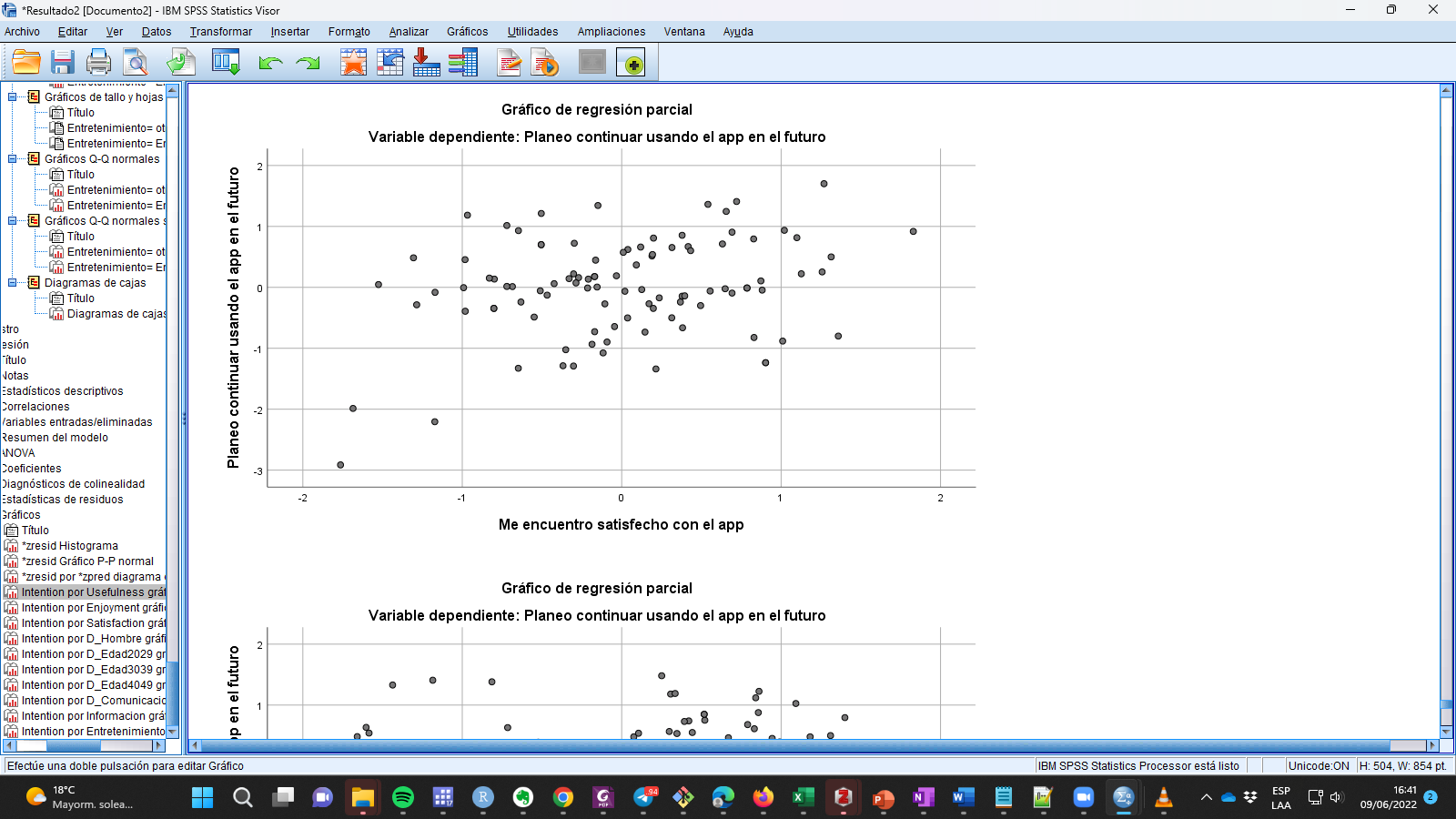
Se obtuvo lo siguiente:

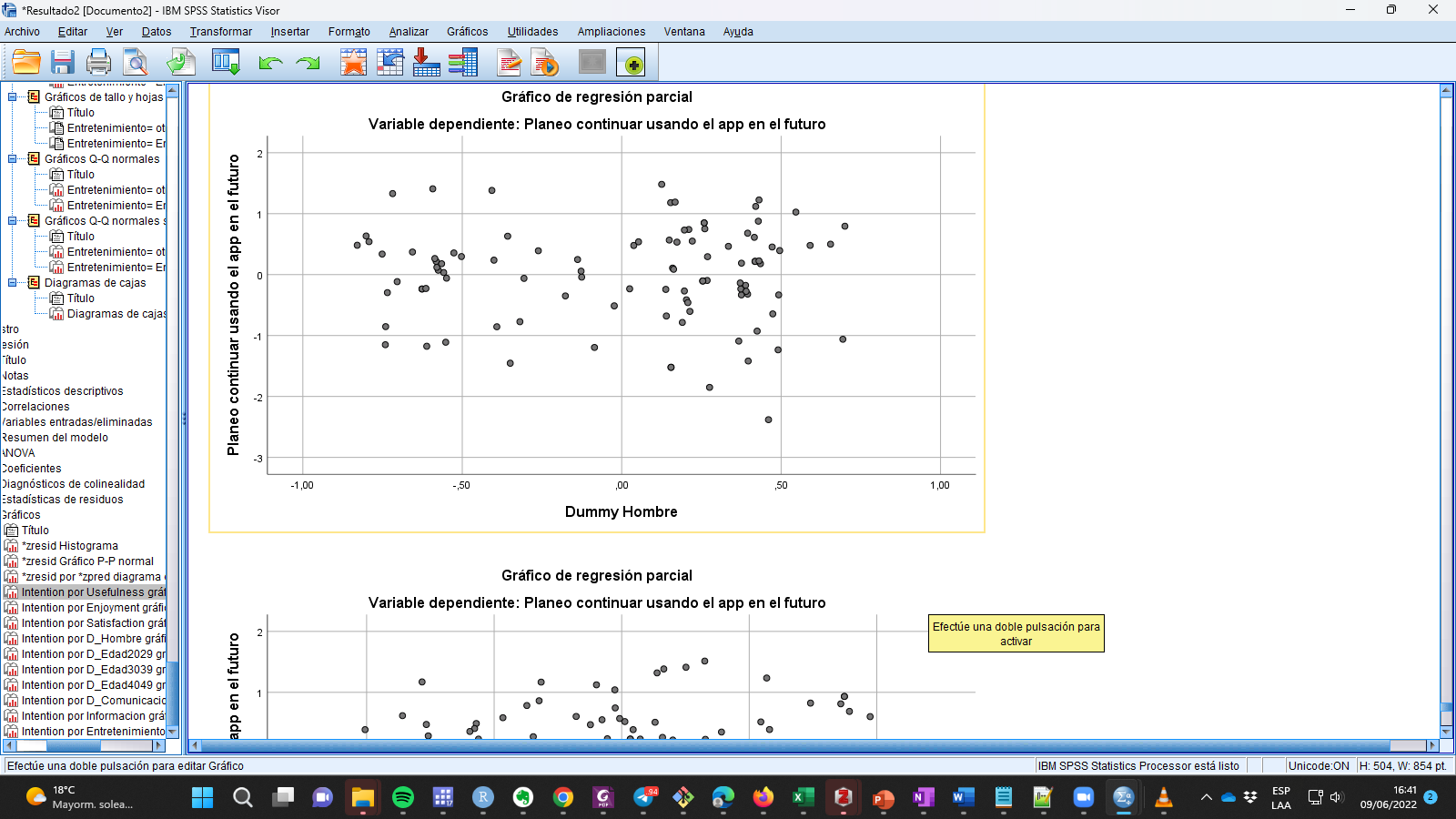


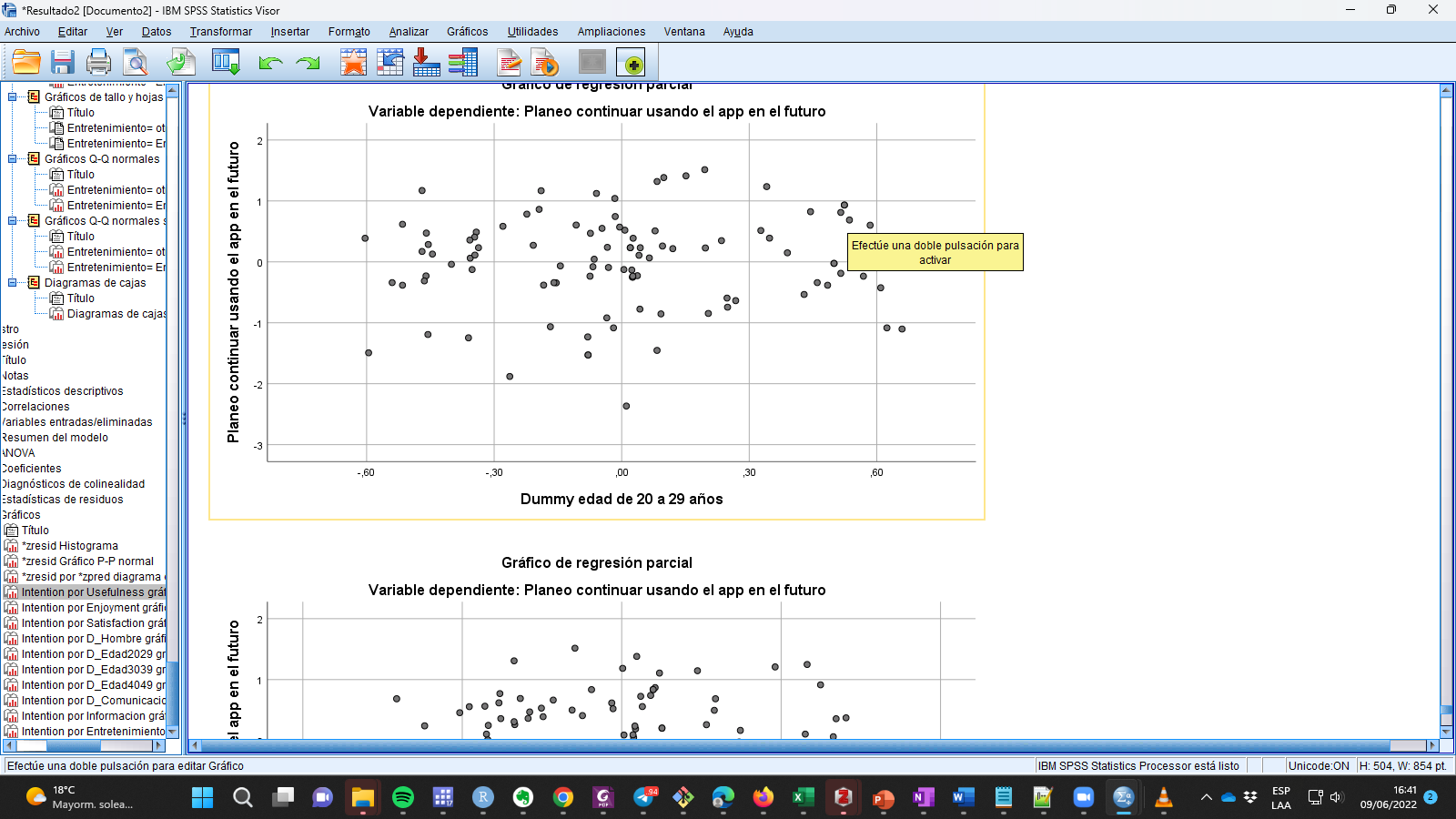
Gráficamente:

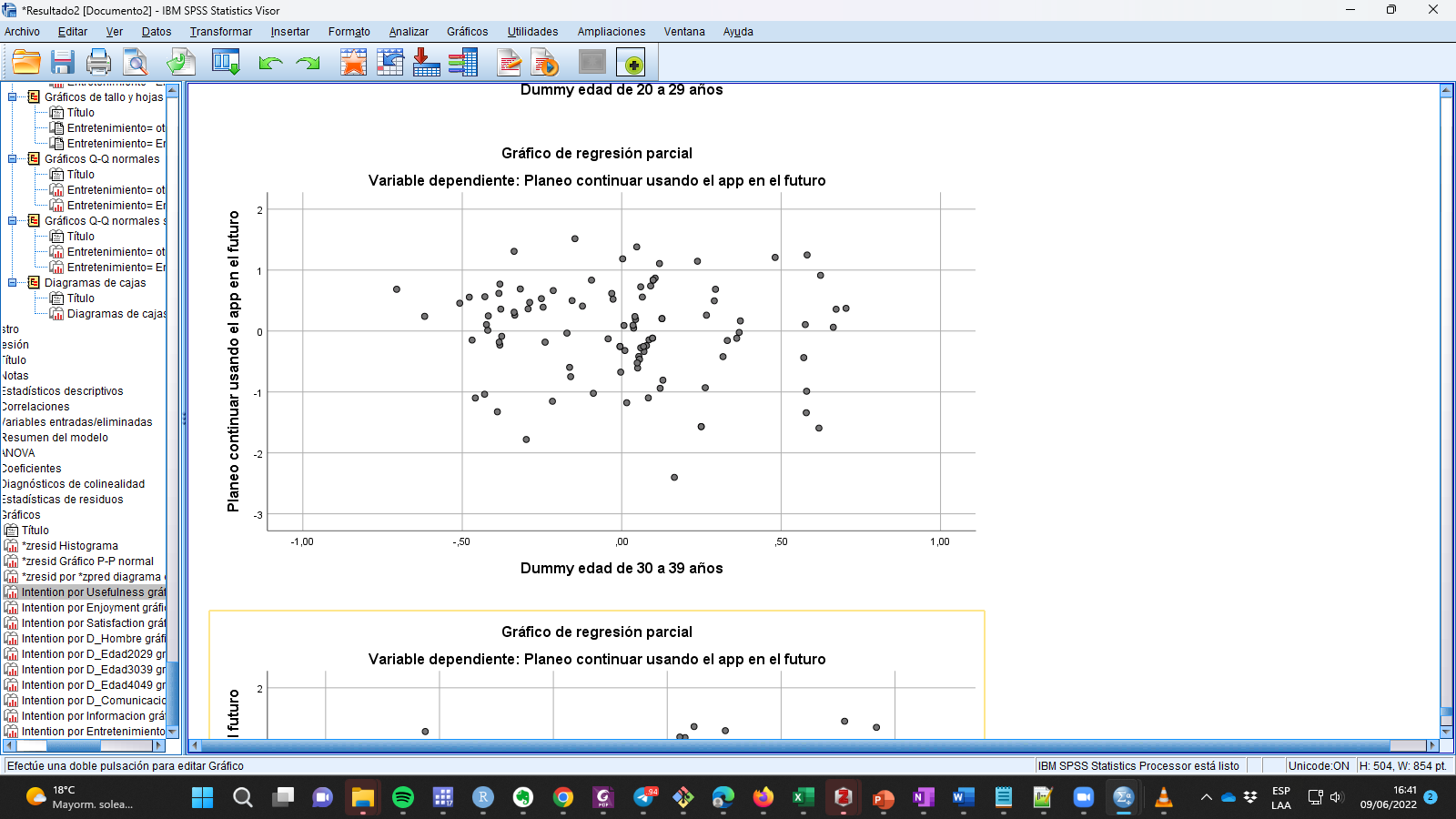


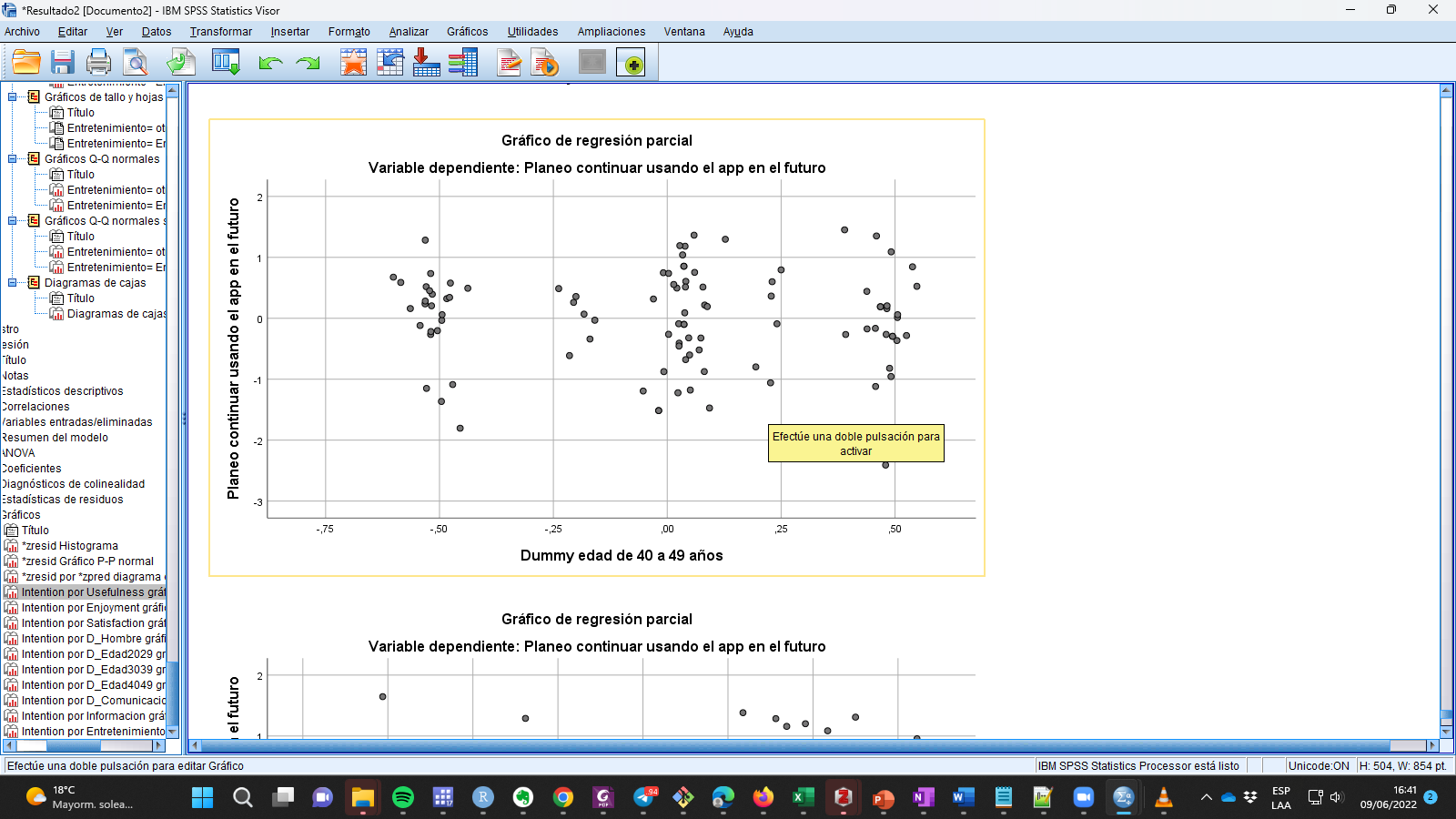


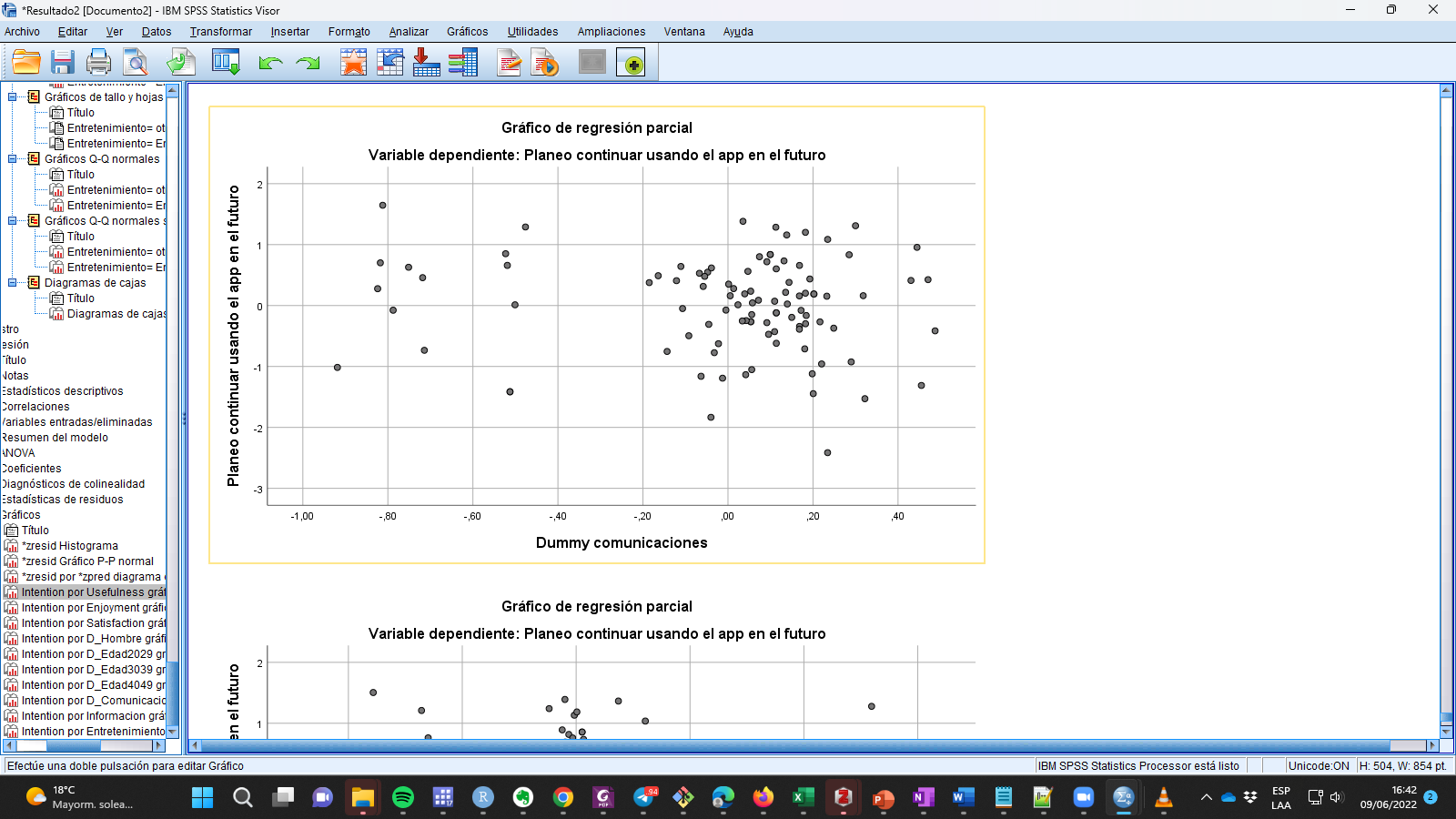


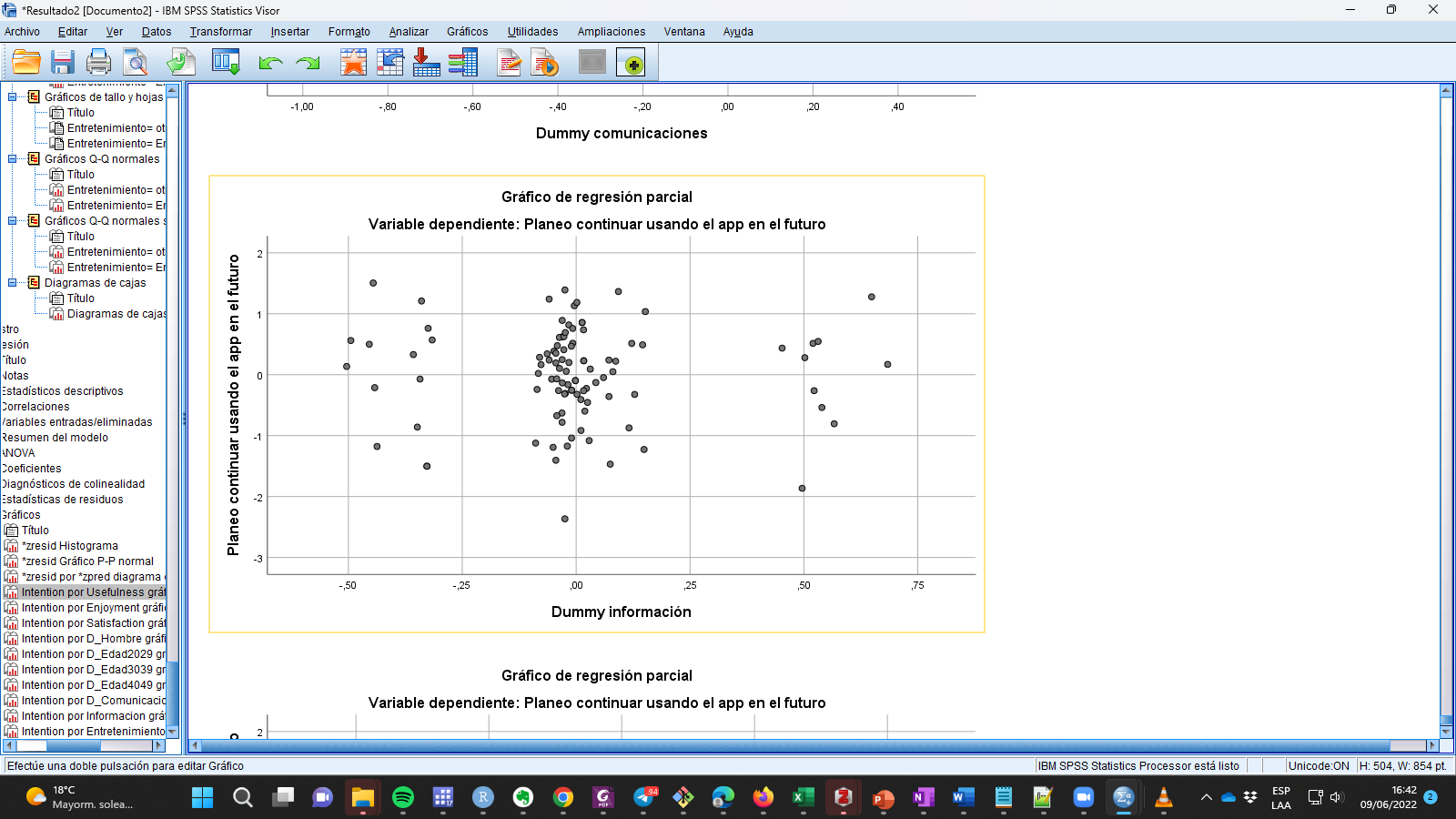


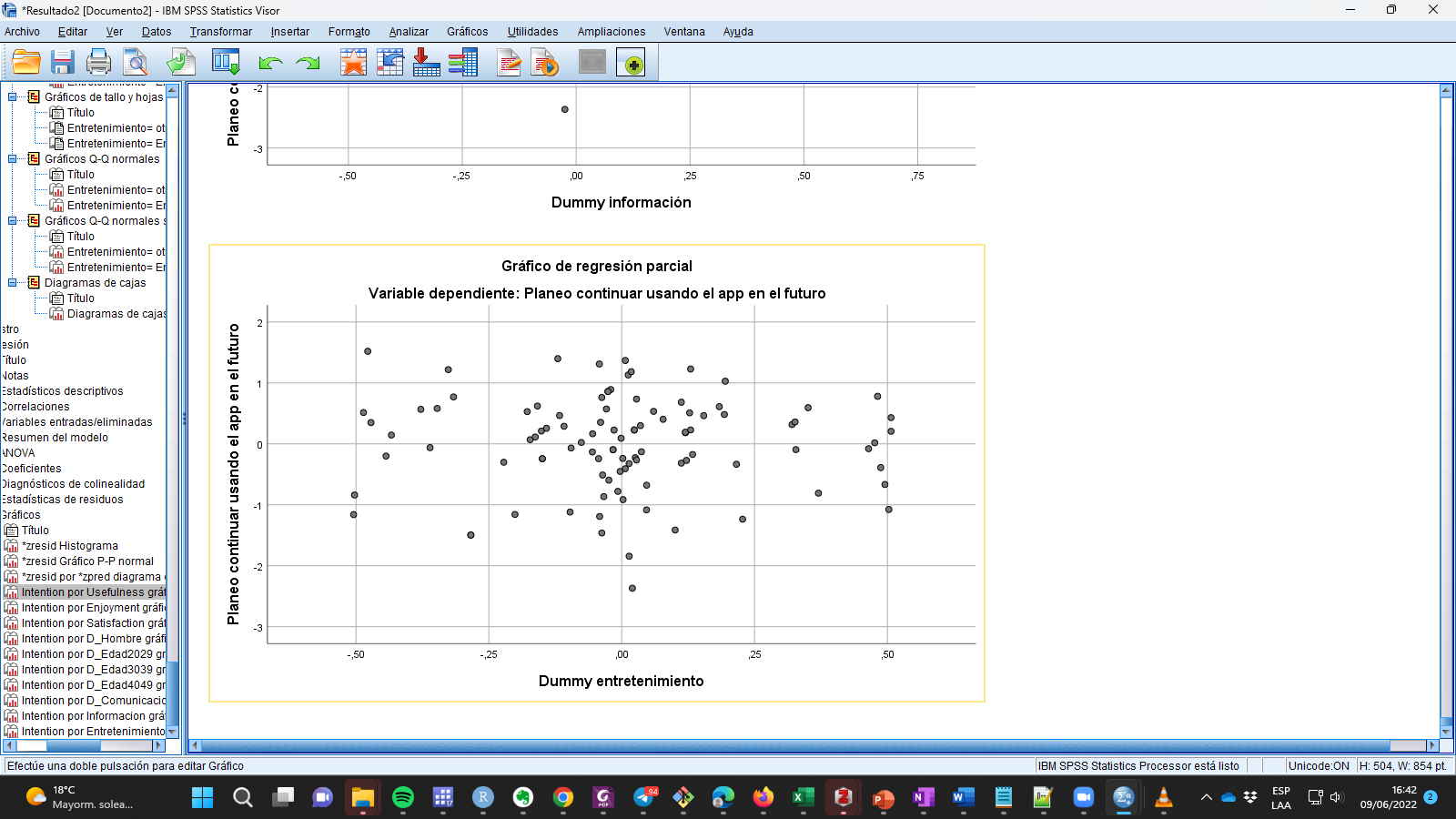






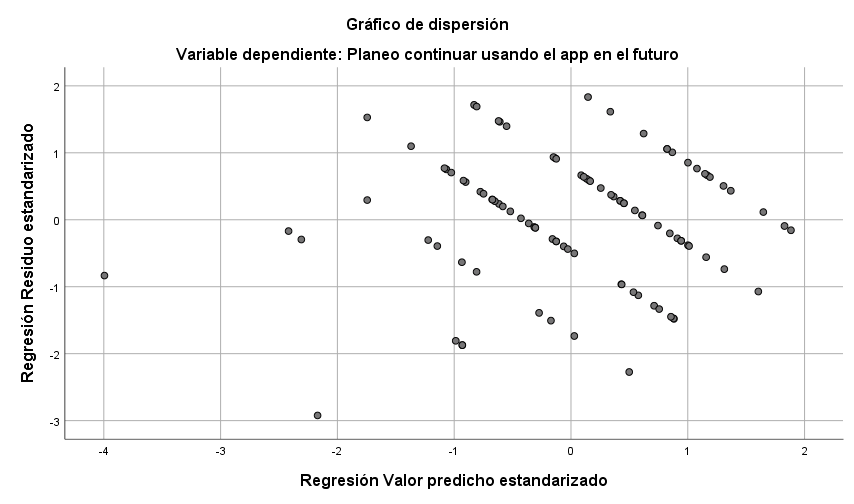






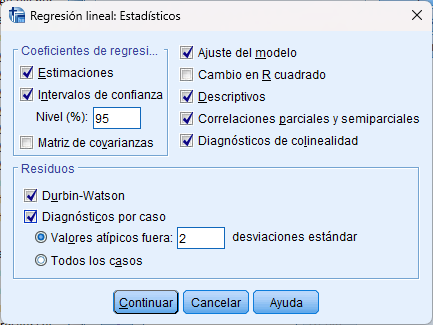
d) Homocedasticidad

Se observa que los puntos no se acumulan alrededor de la línea imaginaria horizontal al 0, por lo cual se verifica que el modelo tiene problemas de Homocedasticidad.

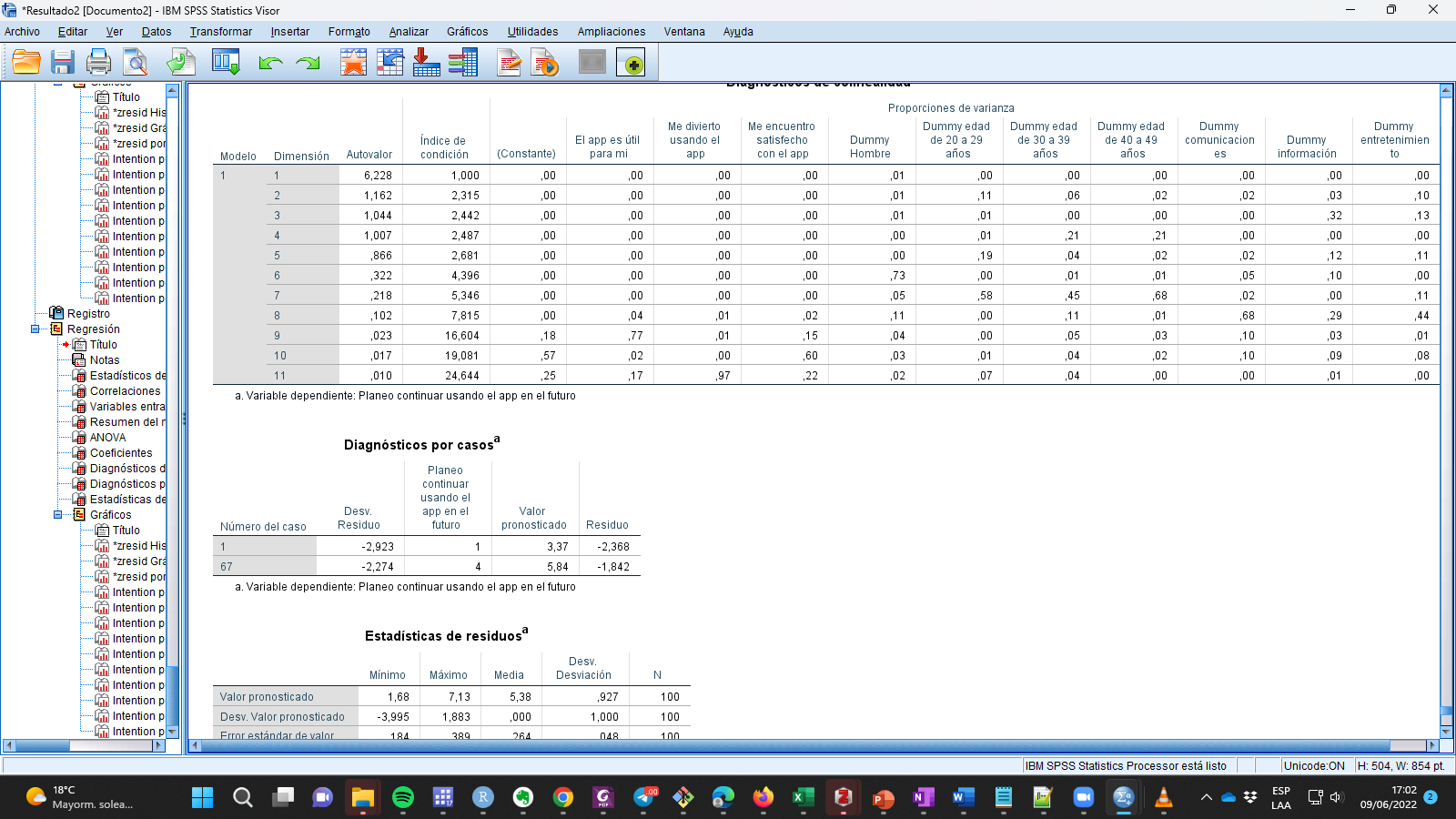


e) Ausencia de outliers

No hay outliers ubicados en +/- 3 SD. También se hizo la verificación para 2 SD, y obtuvimos lo siguiente:

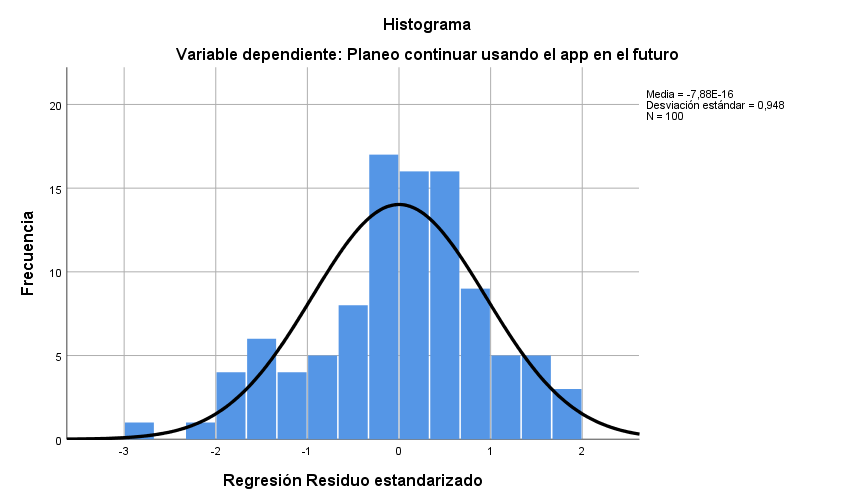


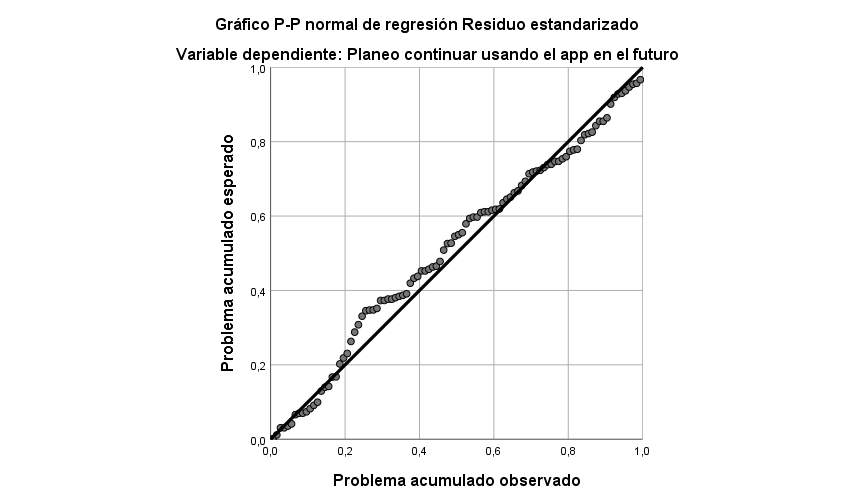
Para este caso se obtuvieron como outliers las observaciones 1 y 67:



f) Distribución normal de los errores

Se observa que la distribución de los errores no se ajusta exactamente a una distribución normal; sino que más bien mantiene un sesgo:

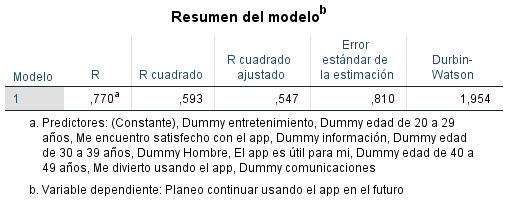




Lo mismo sucede con la gráfica P-P, en donde los puntos no se ajustan en su totalidad a la línea recta. Por lo cual, no hay normalidad de los residuos

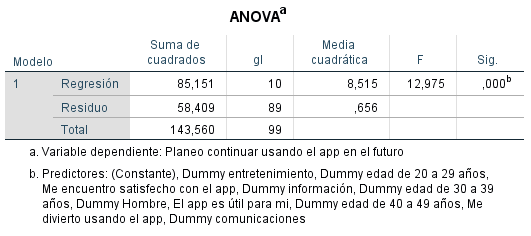
1. Análisis e interpretación del R2 (1 pto), la Prueba F (1 pto), y la significancia de c/u de los coeficientes (3 ptos)

*R2*



El R cuadrado es 0.593, por lo cual el 59% de mi variable dependiente es explicada por las independientes.

*Prueba F*



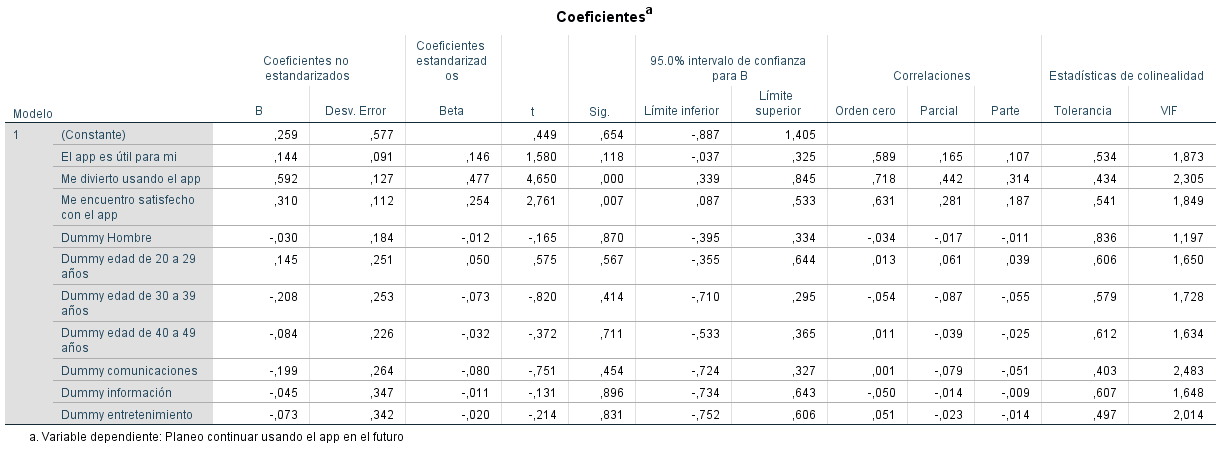
El valor de F resulta en 12.975 con un p-value de 0.000.

H0: Todos los Betas son igual a cero

H1: Al menos un Beta es diferente de cero

Con el valor del p-value se rechaza la hipótesis nula, por lo que tiene sentido continuar con el modelo.

*Significancia de los coeficientes*



Se observa que las únicas variables significativas son “me divierto usando el app” y “me encuentro satisfecho con el app”. El resto de variables tienen un p-value superior a 0.05

1. El valor va de 0 a 4. Se considera valores de 2 o cercanos a 2 (1.5 a 2.5) como adecuados y que denotan independencia de los errores [↑](#footnote-ref-3459)