

Tarea 5

García Ramirez José Antonio

Considere la base de datos para 51 observaciones en el Estado de Washington sobre el gasto en sanidad, *exphlth*, y la renta disponible, *income*, correspondientes al año 1993 ¹

1. Describe los datos.

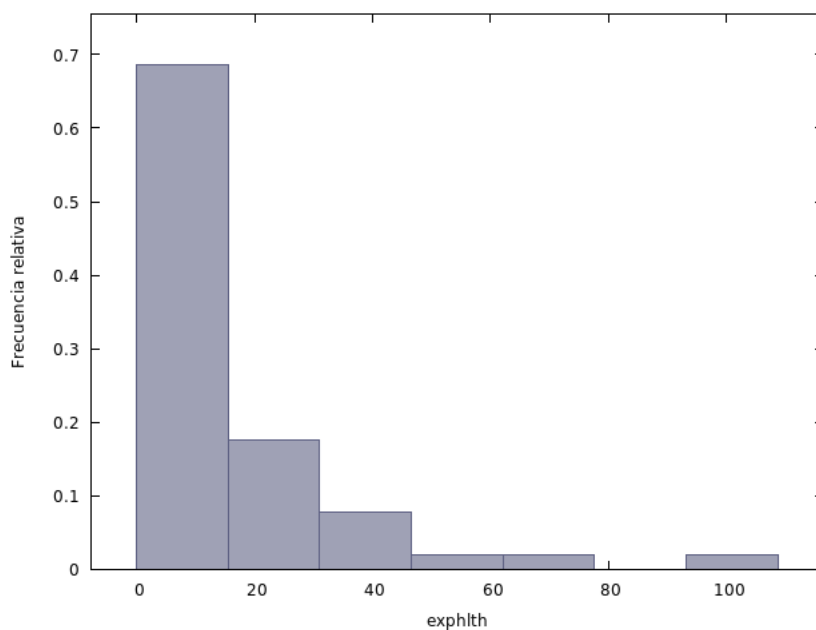
Los datos son de tipo corte transversal donde las unidades de observación son los estados de U.S.A y registran los gastos de asistencia sanitaria por estado así como el ingreso personal.

La variable *exphlth* reporta el gasto en asistencia sanitaria en billones de dólares.

Y la variable *income* reporta el ingreso en billones de dólares.

Ambas variables son cuantitativas y continuas. De los datos cabe destacar que los billones en textos de origen en inglés como en este caso corresponden a mil millones y no a un millón de millones como en textos de origen español.

2. Haz el grafico de frecuencias de la variable *exphlth* e iterpretalo.

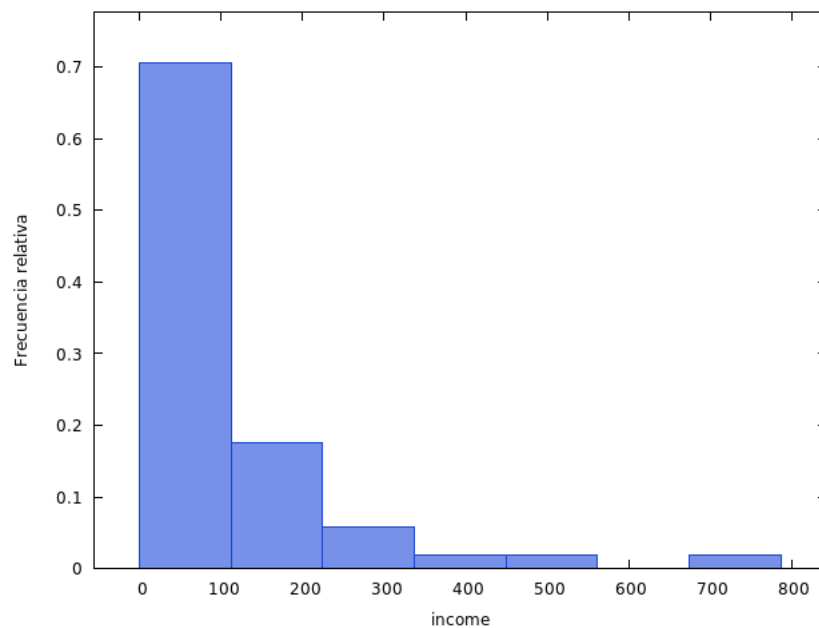


Del histograma de la variable *exphlth* podemos ver que el rango es amplio pues las clases abarcan el intervalo $[0, 100]$. La frecuencia relativa de la primer

¹Fichero data32.gdt. Fuente Statistical Abstract of U.S (1995), recogido en Ramanathan, R. (2002), *Introductory econometrics with applications*, th. Ed., South-Western.

clase es alta, aproximadamente 0,7, por lo que podríamos decir que la media muestral de estos datos está en esta clase. La distribución muestral de la variable *exphlth* no es simétrica y presenta asimetría positiva es decir la media muestral es mayor que la mediana.

3. Haz el grafico de frecuencias de la variable *income* e iterpretalo.



Al igual que en el caso anterior el histograma de la variable *income* presenta asimetría positiva y la distribución empírica se parece mucho, visualmente hablando, a la de la variable *exphlth*. Podemos ver que el rango es menor que en el caso anterior pues al fijar el número de clases a 7 las clases abarcan el intervalo $[0, 8100]$ inclusive la primer clase presenta una mayor frecuencia que la primer clase del caso anterior. Al igual que en el caso anterior existe una clase vacía, la penúltima.

4. Escribe la tabla de estadísticos descriptivos de las variables *exphlth* e *income*. Interpreta los resultados.

Summary Statistics, using the observations 1–51

Variable	Media	Mediana	Mínimum	Máximo
exphlth	15.2649	10.0660	0.998000	94.1780
income	105.133	64.1000	9.30000	683.500
Variable	Std. Dev.	C.V.	Simetría	Ex. Curtosis
exphlth	17.8877	1.17182	2.44968	6.83801
income	125.296	1.19178	2.59308	8.00936
Variable	5 % perc.	95 % perc.	IQ Rango	Obs. perdidas
exphlth	1.54340	56.7028	12.9350	0
income	11.0800	387.240	103.400	0

Como lo sugerían los histogramas en ambos casos la media muestral es mayor a la mediana. Para poder comparar la variación entre ambas variables usamos el coeficiente de variación el cual es mayor en el caso de la variable income y podemos afirmar que la variación en los ingresos en el conjunto de estados es mayor que la variación de gasto en sanidad del mismo conjunto de estados. En ambos casos la distribución tiene una fuerte asimetría positiva pues el coeficiente que lo mide es mayor a 2 y su curtosis es alta, en ambos casos rebaza 6 por lo que podríamos decir que los valores de las colas son importantes, lo cual era de esperarse pues la primer clase en ambos casos encierra una gran cantidad de datos y la última clase concentra un porcentaje bajo pero no despreciable de datos.

5. Busca el coeficiente de correlacion entre las variables e interprétalo.

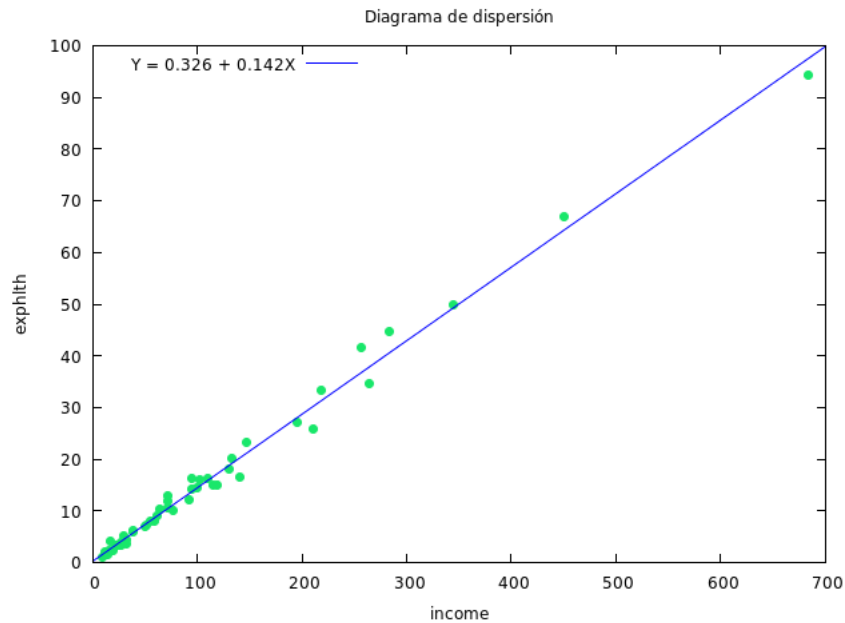
La correlación de Pearson, porque Gretl calcula la correlación de Pearson usando el lenguaje R (lo comprobé), entre las variables que estamos considerando se muestra en la siguiente tabla:

Correlation coefficients, using the observations 1–51
 5 % critical value (two-tailed) = 0.2759 for n = 51

exphlth	income	
1,0000	0,9953	exphlth
	1,0000	income

Lo cual sugiere una fuerte relación lineal positiva entre las variables. La prueba sobre la correlación dice que efectivamente se rechaza la hipótesis nula de que la correlación lineal entre las variables es cero. Esto no es sorprendente pues como vimos en los histogramas ambas variables tienen comportamientos muy parecidos que se ven reflejados en sus respectivos histogramas.

6. Realiza un análisis gráfico de la relación existente entre las variables e interpreta el resultado.



Tal como lo sugiere el coeficiente de correlación de Pearson, **podemos apreciar una fuerte relación lineal positiva entre las variables**. Podemos especular una relación causal entre las variables de tipo *income* \rightarrow *exphlth* pues los ingresos delimitan la cantidad que se destinar a un servicio como la sanidad en este caso. Finalmente podemos decir que existe un dato que llama la atención, el estado de California “CA”, pues tanto su ingreso como su gasto son altos con respecto a los demás estados pero en vista de que este registro mantiene la relación lineal entre las variables no afirmaría que se trata de un dato atípico pero valdría la pena considerar un análisis en donde no se considere este estado.