

Introducción

En este artículo de final de año vamos a dar cierre al curso de Octave que os hemos dado a lo largo de este primer volumen de la revista con una serie de problemas que podemos resolver con *Octave*. Espero leáis los problemas y antes de mirar la solución os déis un tiempo para hacerlo por vosotros mismos. Es maravilloso encontrar la solución por uno mismo.

Durante el año hemos dado unas nociones básicas de *Octave* y puedes aumentar los conocimientos usando, por ejemplo, el siguiente tutorial http://en.wikibooks.org/wiki/Octave_Programming_Tutorial.

1. Esperanza de vida

Según la organización mundial de la salud la definición de esperanza de vida es como sigue: años que un recién nacido puede esperar vivir si los patrones de mortalidad por edades imperantes en el momento de su nacimiento siguieran siendo los mismos a lo largo de toda su vida. Hacer un cálculo de esta cantidad es muy complicado y se tiene en cuenta factores como por ejemplo la medicina, la higiene, las guerras, etc.

Lo que os proponemos es lo siguiente: dado la esperanza de vida de Angola proporcionada por el banco mundial de datos (<http://datos.bancomundial.org/>) que tenemos datos desde 1980 a 2012, os proponemos resolváis el desafío A.



: ¿En qué año tiene previsto el pueblo angolano tener una esperanza de vida de 80 años?

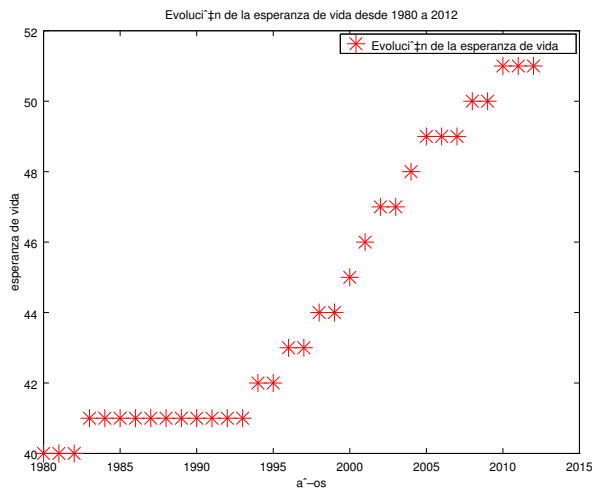
Nuestra Solución

Hemos coleccionado los datos en un archivo *Hopelife.csv*. Los vamos a cargar a Octave y vamos a echarle un vistazo a estos. El código lo podrás visualizar en la sección de códigos de la revista en www.revistasolucoes.com.

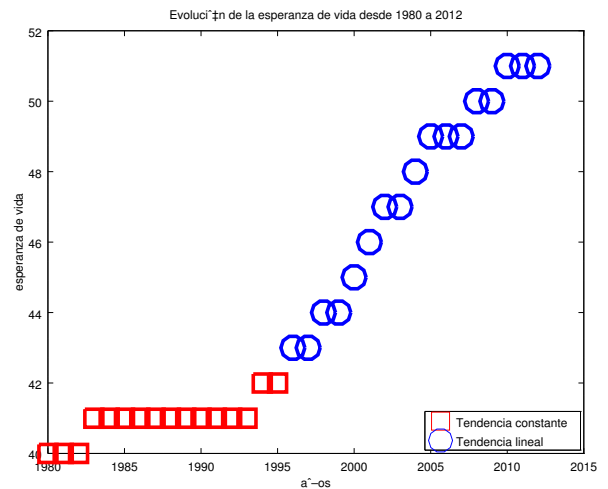
Si observamos la Tabla 1, gráfico (A) tenemos claramente dos tendencias en los datos. Una tendencia constante y una tendencia lineal. Ambas las mostramos en el gráfico:

Y si vamos a la historia de Angola, la tendencia constante corresponde con años de la guerra civil. Esta terminó en 2002 pero vamos a quedarnos con los datos desde el año 1996 ya que parece están en la misma tendencia que de los años posteriores.

La idea de lo que vamos a hacer es lo siguiente: nuestro objetivo es buscar una función $f(n)$ que nos da la esperanza de vida en el año n . Como hemos dicho antes la tendencia que parecen tener los datos (recuerda que nos hemos quedado con los datos azules en la Tabla 1, gráfico (B)) es una tendencia lineal. Existe un concepto en estadística que nos asegura que nuestra intuición está en lo correcto y es el *coeficiente de correlación* o *coeficiente de correlación lineal de Pearson*. Este coeficiente que denotaremos por ρ mide el grado de covariación entre dos distintas variables relacionadas linealmente, denotémoslas x e y . El cálculo del coeficiente de correlación entre dos variables x e y es sencillo y viene dado por la fórmula $\rho = \frac{\sigma_x \sigma_y}{\sigma_{xy}}$ donde σ_x o σ_y son las desviaciones típicas de las variables x e y respectivamente y σ_{xy} la covarianza entre las dos variables. El parámetro ρ cumple que su rango de valores está entre -1 y 1 y lo podemos interpretar como sigue:



(A)



(B)

Cuadro 1: (A): Evolución esperanza de vida desde 1980 a 2012, (B): Clusterización de los datos

- Si $\rho \sim 1$ significa correlación lineal positiva.
- Si $\rho \sim -1$ significa correlación lineal negativa.
- Si $\rho \sim 0$ significa que las variables no tienen correlación lineal.

En Octave existe el comando `cor` que calcula la correlación entre dos variables:

```
x=datos(17:end,1);  
y=datos(17:end,2);  
disp('El coeficiente de correlación')  
cor(x,y)
```

Y el resultado que se obtiene es de 0,98473 por tanto nuestra intuición queda apoyada por este concepto matemático y tiene mucho sentido realizar una regresión lineal. Esto es buscar una función $f(n) = \theta_1 + \theta_2 \cdot n$ para la previsión de la esperanza de vida en el año n . Existen varios métodos para calcular las constantes θ_1 y θ_2 del modelo. No vamos a dar muchos detalles de ello, puedes buscarlo en cualquier libro de estadística pero te diremos que un método rápido que se puede usar para bases de datos pequeñas es el método de las ecuaciones normales que nos dan la solución directamente de la siguiente fórmula:

$$(1.1) \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \text{inv}(X^t \cdot X) \cdot X^t \cdot y$$

$$\text{donde } X = \begin{bmatrix} 1 & \vdots \\ \vdots & x \\ 1 & \vdots \end{bmatrix}.$$

Año	Esperanza de vida
2030	62
2050	73
2063	80

Cuadro 2: Algunas previsiones futuras sobre la esperanza de vida angolana

Si sigues las instrucciones puedes ver que la función resultante y que predice la esperanza de vida en el año n es $f(n) = -1067,5 + 0,55 \cdot n$. Usando esta gráfica podemos augurar que si el crecimiento de Angola va en la dirección que marca el pueblo angolano tendrá una esperanza de vida de 80 años para el año 2063. Damos algunos resultados en la siguiente tabla: