Actividad 1

Valeriano J. Antequera Palacios

10 de junio de 2015

Ejercicio 1.1 *Demuestra que* $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{1}\mathbf{1}^{'}$ *es simétrica e idempotente. Demuestra que* $\boldsymbol{S}$ *es semidefinida positiva , i.e. , para cualquier vector* $\boldsymbol{a}$*, se tiene que* $\boldsymbol{a}^{'}\boldsymbol{S}\boldsymbol{a} \geq \boldsymbol{0}$*. Probar también que* $\boldsymbol{R} \geq \boldsymbol{0}$*.*

Teniendo que $\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{1}\mathbf{1}^{'}$ vamos a demostrar que $\mathbf{H}$ es simétrica:

$$H = \begin{bmatrix} 1 & 0 \cdots & 0 \\ 0 & 1 \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 1 \end{bmatrix} - \frac{1}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \cdots & 1 \end{bmatrix}$$

De aquí obtenemos que:

$$H = \begin{bmatrix} 1 & 0 \cdots & 0 \\ 0 & 1 \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \cdots & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} = \begin{bmatrix} \frac{n-1}{n} & \frac{1}{n} \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{n-1}{n} \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} \cdots & \frac{n-1}{n} \end{bmatrix}$$

Claramente observamos que la matriz $\mathbf{H}$ es simétrica ya que $\mathbf{H} = \mathbf{H}^{'}$.

Ahora para demostrar que es idempotente haremos el producto $H * H = H$.

Para los terminos que no estan en la matriz:

$$2 * (\frac{n-1}{n}) + \sum_{i=1}^{n-2} \frac{1}{n^2} = \frac{-2n+2}{n^2} + \frac{n-2}{n^2} = \frac{-n}{n^2} = \frac{-1}{n}$$

Por último para los terminos de la diagonal:

$$(\frac{n-1}{n})^2 + \sum_{i=1}^{n-1} \frac{1}{n^2} = \frac{(n-1)^2}{n^2} + \frac{n-1}{n^2} = \frac{n-1}{n}$$

Al juntar las cuentas anteriores y ponerlas como una mat'iz obtenemos que:

$$H * H = \begin{bmatrix} \frac{n-1}{n} & \frac{1}{n} \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{n-1}{n} \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} \cdots & \frac{n-1}{n} \end{bmatrix} = H$$

Ahora para demostrar que $\mathbf{S}$ es semidefinida positiva, tenemos $\mathbf{a}^{'}\mathbf{Sa} \geq \mathbf{0}$.

Tomamos la variable unidimensional $y_i = a^{'} * (x_i - \overline{x})$

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} a^{'}(x_i - \overline{x}) = \frac{1}{n} a^{'} \sum_{i=1}^{n} (x_i - \overline{x}) = 0$$

Ya que$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$.

Con esto tenemos que:

$$\mathbf{var(y)} = \frac{1}{n} \sum_{i=1}^{n} y_i^2 - \frac{1}{n} \sum_{i=1} n[a^{'} * (x_i - \overline{x})^{'} + a] \geq 0$$

Además tenemos que $\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})^{'} = S$.

Por tanto queda demostrado que $\mathbf{a}^{'}\mathbf{Sa} \geq \mathbf{0}$ sea cual sea $\mathbf{a}$ con lo que $\mathbf{S}$ es semidefinia positiva.

A continuación para hallar si $\mathbf{R} \geq \mathbf{0}$ tomamos como hemos demostrado anteriormente que $\mathbf{S}$ es semidefinida positiva.

Primero tenemos que:

$$R = D^{'}SD^{'}$$

Despejando la **S** tenemos

$$DRD = DD^{'}SD^{'}D = S \Rightarrow a^{'}DRDa = a^{'}Sa \geq 0$$

Con esto queda demostrado que:

$$DRD \geq 0 \Rightarrow R \geq 0$$

Ejercicio 1.2 *La siguiente tabla representa tres medidas de rentabilidad de 34 accione en bolsa durante un periodo de tiempo. La primera es la rentabilidad efectiva, la segunda la proporción de beneficios y la tercera el cociente entre el precio por acción y los beneficios. Calcular usando R o MatLab el vector de medias, la matriz de covarianzas y de correlaciones, así como el valor de la matriz ,**H**. Comprobar que definiendo adecuadamente **H** para cada caso, **HX** centra la matriz por columnas mientras que **XH** la centra por filas.*

Solución 1.1

1. *Cálculo del vector de medias.*

*Para hallar el vector de medias en R, primero definimos la matriz **X** leyendo los datos de un fichero .csv de la siguiente manera:*

```
>X<-read.csv('libro.csv',header=FALSE)
>X<-as.matrix(X)
```

Ahora seguidamente definimos los nombres de las columnas y obtenemos los valores de n y p.

```
>colnames(X)<-c("X1","X2","X3")
>n<-dim(X)[1]
>p<-dim(X)[2]
```

Después con la orden colMeans se obtiene el vector de medias buscado:

```
> colMeans(X)
       X1        X2        X3
 9.420588 69.526471  9.096765
```

También podemos comprobar la solución aplicando la definición, $\overline{\mathbf{x}}' = \frac{1}{n}\mathbf{X}'\mathbf{1}$*:*

```
> v1<-matrix(1,nrow=n,ncol=1)
> mx<-(1/n)*t(X)%*%v1
> mx
          [,1]
X1   9.420588
X2  69.526471
X3   9.096765
```

2. *<u>Cálculo de la matriz de covarianzas</u>.*

Para obtener S*, matriz de varianzas y covarianzas, podemos usar el siguiente comando de R:*

```
> mv<-cov(X)
          X1         X2         X3
X1  29.09562 100.44065 -15.70281
X2 100.44065 576.22928 -18.54628
X3 -15.70281 -18.54628  22.56599
```

Que nos devuelve la matriz de varianzas muestrales. Dado que no es la matriz que buscamos, la expresamos en términos de la varianza:

$$S = mv\frac{n-1}{n}$$

Que, ejecutado en R nos dió lo siguiente:

```
> S<-((n-1)*mv)/n

           X1        X2        X3
X1  28.23987  97.48651 -15.24096
X2  97.48651 559.28136 -18.00080
X3 -15.24096 -18.00080  21.90229
```

3. *Cálculo de la matriz de correlaciones.*

Aplicando la definición de la matriz de correlaciones, $R = D^{-1}SD^{-1}$*, donde* D *es la diagonal de la matriz de varianzas, operamos con R:*

```
> R<-cov(X)
> R
           X1         X2         X3
X1  1.0000000  0.7757076 -0.6128248
X2  0.7757076  1.0000000 -0.1626417
X3 -0.6128248 -0.1626417  1.0000000
```

4. HX*: centrado por columnas.*

Para este caso, definimos $\boldsymbol{H}$ *como una matriz de dimensión* 34×34.
$H = I_{34\times34} - \frac{1}{34}\mathbf{1}_{34\times1}\mathbf{1}'_{1\times34}$:

```
> H<-diag(n)-(1/n)*v1 %*%t(v1)
```

Y se comprueba que HX *centra la matriz* X *por columnas:*

```
> H %*%X
             X1         X2         X3
 [1,] -6.0205882  20.173529 21.1032353
 [2,] -4.3205882 -13.826471  0.8032353


                   ........
```

```
[33,]   4.9794118  18.273529 -3.8967647
[34,]   5.4794118 -35.026471 -4.4067647
```

Las medias de las columnas $1, 2$ *y* 3 *son* $9'420588$, $69'526471$ *y* $9'096765$ *respectivamente. Se ve claramente que los valores mostrados son los de la matriz* X *menos la media de las columnas.*

5. *XH: centrado por filas.*

*Esta vez se define **H** como una matriz* 3×3 *para centrarlo ahora por filas tal que* $H = I_{3\times3} - \frac{1}{3}\mathbf{1}_{3\times1}\mathbf{1}'_{1\times3}$*:*

```
> v1<-matrix(1,nrow=p,ncol=1)
> H<-diag(p)-(1/p)*v1 %*%t(v1)
```

Definiendo el producto XH *se comprueba que la matriz* X *queda centrada por filas:*

```
> X %*%H
           [,1]      [,2]       [,3]
 [1,] -37.70000 48.60000 -10.900000
 [2,] -18.46667 32.13333 -13.666667
                              ...
[33,] -21.40000 52.00000 -30.600000
[34,]  -3.13000 16.47000 -13.340000
```

Las medias de las filas $1, 2, 33$ *y* 34 *son, respectivamente,* $41'1$, $23'56667$, $35'8$ *y* $18'03$. *Al igual que en el apartado anterior, se ve cómo los valores obtenidos son los mismos de* X *menos la media de sus respectivas filas.*

Queda entonces comprobado que, definiendo adecuadamente H *para cada caso, la matriz* X *queda centrada bien por columnas o bien por filas.*

En esta siguiente página tenemos la tabla de datos que hemos usado en el ejercicio.

Acciones		
Rentabilidad efectiva	Proporción de beneficios	$\frac{Precio}{Beneficios}$
3.4	89.7	30.2
5.1	55.7	9.9
4.5	52.3	11.5
3.5	47	11.2
5.9	42.7	7
5.1	30.6	6.9
4.6	64.4	11.8
5	51	9.6
3.2	54.4	14.7
3.4	45.7	13.2
6.5	39.9	5.2
4.4	40.3	13.7
5.1	52.4	11
5.8	43.9	8
4.6	52.8	14.4
7.2	65.8	7.8
7.2	58.1	7.7
4.4	58.5	12.1
7.8	84.3	11
16	96.5	6
16.7	100	6.8
15.2	92.3	5.2
17.5	99.9	6.8
16.2	93.5	6.1
14.7	100	6.6
15.3	99.9	5.9
15.8	100	6.9
18.3	96.3	5.7
15.9	100	6.1
16.1	92.5	6.1
9.7	87.6	7.7
6.9	53.6	6.6
14.4	87.8	5.2
14.9	34.5	4.69