1. Suponga $\boldsymbol{x} = (x_1, \dots, x_p)^\top$ un vector aleatorio con matriz de covarianza

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}.$$

Encuentre la primera componente principal y el porcentaje que esta explica.

Solución: Nota que las matrices con esta estructuras tienen una regularidad de valores propios. Definamos:

$$A = \frac{1}{\sigma^2}\Sigma = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{bmatrix}_{p\times p}$$

Es una matriz de correlación. Luego se tiene que los valores propios de A vienen dados por:

$$\begin{aligned} \lambda_1 &= 1-(p-1)\rho \\ \lambda_2 &= 1-\rho \\ &\vdots \\ \lambda_p &= 1-\rho \end{aligned}$$

Se sabe que Tr(A) = p lo cual se cumple para la suma de los valores propios descritos anteriormente (que se obtienen por argumento de recursividad). Sin perdida de generalidad supongamos que $\rho \in (0,1)$ entonces:

$$\lambda_1 > \lambda_2 = \lambda_3 = \dots = \lambda_p$$

La primera componente principal viene asociado al valor propio mas grande, por tanto para proseguir notemos que el vector propio asociado a λ_1 viene dado por:

$$V_p(\lambda_1) = (1, \dots, 1)$$

$$\Rightarrow \hat{V}_p(\lambda_1) = \left[\frac{1}{\sqrt{p}}, \dots, \frac{1}{\sqrt{p}}\right] \text{ pues } \hat{v} = \frac{v}{\|v\|}$$

Por lo tanto la primera componente principal viene dada por:

$$\boxed{y_1 = \frac{1}{\sqrt{p}} x_1 + \dots + \frac{1}{\sqrt{p}} x_p}$$

Y si se asume esperanza 0 entonces tenemos que el porcentaje de varianza explicado viene dada por la razón entre el valor propio y la traza.

$$\boxed{r_1 = \frac{\lambda_1}{Tr(A)} = \frac{1+(p-1)\rho}{p}}$$

2. Suponga una muestra de tamaño n de vectores aleatorios $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^\top$ asociados a 4 variables de interés. Se obtuvo además,

$$\boldsymbol{S} = \begin{pmatrix} 2.53 & 3.50 & 2.06 & 1.45 \\ 3.50 & 5.05 & 2.86 & 2.02 \\ 2.06 & 2.86 & 1.86 & 1.19 \\ 1.45 & 2.02 & 1.19 & 0.86 \end{pmatrix},$$

cuyos valores y vectores propios son:

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (10, 0.1, 0.02, 0.01),$$

$$\boldsymbol{T} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3, \boldsymbol{t}_4) = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/(2\sqrt{3}) & 1/(2\sqrt{3}) & 1/(2\sqrt{3}) & -3/(2\sqrt{3}) \end{pmatrix}.$$

Encuentre las PC y sus varianzas ¿Qué porcentaje explica cada una?

Solución:

Nota que toda la información necesaria para el análisis viene dada por la descomposición espectral de S. Por tanto sin mirar la matriz S tenemos que por el metodo de componentes principales la varianza de cada componente principal denotada por yi viene dado por los valores propios. Por ende se tiene que:

Var(y1) = 10
Var(y2) = 0.1
Var(y3) = 0.02
Var(y4) = 0.01

Y los porcentajes que explican es la razón entre la varianza de una componente y la traza de S (equivalente a la suma de los valores propios o la suma de las varianzas individuales). Entonces tenemos que:

Tr(S) = 10 + 0.1+0.02 + 0.01 = 10.13

Por tanto:

Razón 1: 10/10.13 = 0.987166831194472
Razón 2: 0.1/10.13 = 0.00987166831194472
Razón 3: 0.02/10.13 = 0.00197433366238894
Razón 4: 0.01/10.13 = 0.000987166831194472

Finalmente las componentes principales vienen dadas por una combinación lineal de los vectores propios siguiendo el orden de los valores propios de forma decreciente.

$$y_1 = \frac{1}{2}X_1 + \frac{1}{\sqrt{2}}X_2 + \frac{1}{\sqrt{6}}X_3 + \frac{1}{2\sqrt{3}}X_4$$

$$y_2 = \frac{1}{2}X_1 - \frac{1}{\sqrt{2}}X_2 + \frac{1}{\sqrt{6}}X_3 + \frac{1}{2\sqrt{3}}X_4$$

$$y_3 = \frac{1}{2}X_1 - \frac{2}{\sqrt{6}}X_3 + \frac{1}{2\sqrt{3}}X_4$$

$$y_4 = \frac{1}{2}X_1 - \frac{3}{2\sqrt{3}}X_4$$

Obs: nota que en R el análisis cambia puesto que los valores propios de S no son los que se mencionan.

```
eigen() decomposition
$values
[1] 10.03042475  0.18045021  0.06742591  0.02169913

$vectors
           [,1]        [,2]       [,3]        [,4]
[1,] -0.4984305 -0.02217536  0.6892891  0.52531498
[2,] -0.7049798  0.52386006 -0.4776755 -0.02000721
[3,] -0.4143693 -0.85130089 -0.3217723 -0.00688734
[4,] -0.2878692 -0.01912294  0.4395102 -0.85064472
Call:
princomp(x = S)

Standard deviations:
      Comp.1       Comp.2       Comp.3       Comp.4
1.523237e+00 7.328581e-02 2.531623e-02 5.374516e-09

 4  variables and  4 observations.
```

3. El modelo de regresión lineal multivariado equicorrelacionado, dado por

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \qquad \boldsymbol{E} \sim \mathsf{N}_{n,p}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}),$$

con

$$\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p^\top], \qquad \sigma^2 > 0, -1 < \rho < 1.$$

a) Encuentre los MLE de $\boldsymbol{B}$, σ^2 y ρ.

b) Proponga un test para probar la hipótesis $H_0 : \rho = 0$.

Notemos que $Y \sim N_{n,p}(XB, Id \otimes \Sigma)$, luego se tiene que:

$$f(y) = (2\pi)^{\frac{-np}{2}} |\Sigma|^{\frac{-n}{2}} \exp\left(\frac{-1}{2}\mathrm{Tr}(Y - XB)\Sigma^{-1}(Y - XB)^\top\right)$$

Donde la log-verosimilitud viene dada por:

$$\ell(B,\sigma^2,\rho) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}(Tr)\left(\sum_{j=1}^{m}(y_j - x_jB)\Sigma^{-1}(y_j - x_jB)^\top\right)$$

Note que:

$$\begin{aligned}
\mathrm{d}_B\ell &= -\frac{1}{2}\mathrm{Tr}\left(\sum_{j=1}^{m}\left((-x_j\mathrm{d}B)\Sigma^{-1}(y_j - x_j\mathrm{d}B)^\top - (y_j - x_jB)\Sigma^{-1}(x_j\mathrm{d}B)^\top\right)\right)\\
&= \sum_{j=1}^{m}\mathrm{Tr}\left((y_j - x_jB)\Sigma^{-1}(\mathrm{d}B)^\top x_j^\top\right)\\
&= \mathrm{Tr}\left(\sum_{j=1}^{m}x_j^\top(y_j - x_jB)\Sigma^{-1}(\mathrm{d}B)^\top\right)
\end{aligned}$$

Resolviendo $\mathrm{d}_B\ell = 0$ tenemos que:

$$\begin{aligned}
\sum_{j=1}^{m}x_j^\top(y_j - x_jB)\Sigma^{-1}(\mathrm{d}B)^\top = 0 &\Longrightarrow \sum_{j=1}^{m}x_j^\top(y_j - x_jB) = 0\\
&\Longrightarrow \sum_{j=1}^{m}x_j^\top x_jB = \sum_{j=1}^{m}x_j^\top y_j\\
&\Longrightarrow B = \left(\sum_{j=1}^{m}x_j^\top x_j\right)^{-1}\sum_{j=1}^{m}x_j^\top y_j
\end{aligned}$$

Concluyendo que $\widehat{B} = \left(\sum_{j=1}^{m}x_j^\top x_j\right)^{-1}\sum_{j=1}^{m}x_j^\top y_j$, por otro lado $\Sigma = \sigma^2\underbrace{\left[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p^\top\right]}_{A}$, luego es claro que $\det(\Sigma) = \sigma^{2p}\det(A)$ y $\Sigma^{-1} = \sigma^{-2}A^{-1}$ luego:

$$\ell(B,\sigma^2,\rho) = -\frac{np}{2} - \frac{np}{2}\ln(\sigma^2) - \frac{n}{2}\ln|A| - \frac{1}{2\sigma^2}\mathrm{Tr}\left\{\sum_{j=1}^{m}(y_j - x_jB)A^{-1}(y_j - x_jB)^T\right\}$$

Luego:

$$\mathrm{d}_{\sigma^2}\ell = -\frac{np}{2\sigma^2} + \frac{1}{2\sigma^4}\mathrm{Tr}\left\{\sum_{j=1}^{m}(y_j - x_jB)A^{-1}(y_j - x_jB)^\top\right\}$$

Igualando a 0 se tiene que:

$$\sigma^2 = \frac{1}{np}\mathrm{Tr}\left(\sum_{j=1}^{m}(y_j - x_jB)A^{-1}(y_j - x_jB)^\top\right)$$

Por lo tanto $\widehat{\sigma^2} = \dfrac{1}{np}\mathrm{Tr}\left(\sum_{j=1}^{m}(y_j - x_j\widehat{B})\widehat{A}^{-1}(y_j - x_j\widehat{B})^\top\right)$, finalmente por construcción clasica de la normal se sabe que $\widehat{\Sigma} = \frac{1}{n}\sum_{i=j}^{m}(y_j - x_j\widehat{B})(y_j - x_j\widehat{B})^\top$ y por invarianza del estimador tenemos que:

$$\widehat{\sigma^2}\left[(1-\widehat{\rho})I_p + \widehat{\rho}\boldsymbol{1}_p\boldsymbol{1}_p^\top\right] = \frac{1}{n}\sum_{i=j}^{m}(y_j - x_j\widehat{B})(y_j - x_j\widehat{B})^\top$$

Es la ecuación de estimación de ρ. Finalmente para probar el test $\rho = 0$ proponemos un test de razón de verosimilitudes el cuanto toma la siguiente estructura debido a que si $\rho = 0$ entonces $\Sigma = \sigma^2 Id_{p\times p}$.

$$LR = \frac{L(\hat{B},\hat{\sigma}^2)}{L(\hat{B},\hat{\sigma}^2,\hat{\rho})} = \frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}\mathrm{Tr}\left[(y_j - x_j\hat{B})(y_j - x_j\hat{B})^\top\right]\right\}}{|A|^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}\mathrm{Tr}\left[(y_j - x_j\hat{B})\hat{A}^{-1}(y_j - x_j\hat{B})\right]\right\}}$$

Con $\widehat{A} = (1-\widehat{\rho})Id + \widehat{\rho}1_p1_p^\top$.

4. Consider el conjunto de datos

Peso del grano	Peso de la paja	Cantidad de fertilizante
(Y_1)	(Y_2)	(x_1)
40	53	24
17	19	11
9	10	5
15	29	12
6	13	7
12	27	14
5	19	11
9	30	18

a) Ajuste un modelo de regresión bivariado $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$, donde las filas de $\boldsymbol{E}$son IID $\mathrm{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$. Suponga un modelo con el regresor x_1 y un intercepto. ¿Es razonable el supuesto de normalidad?

b) Pruebe la hipótesis de que el vector de respuestas no depende del predictor, x_1. Comente el resultado.

Solución:

(a) Existen dos formas de resolver el problema. La primera es estimar B mediante la formula:

$$(X^T X)^{-1} X^T Y$$

Obteniendo la siguiente matriz:

A matrix: 2 × 2 of type dbl

	Y1	Y2
(Intercept)	-3.752446	-2.296477
x1	1.402153	2.140900

O bien corriendo el siguiente código en R:

```
fit=lm(cbind(Y1,Y2)~x1)
fit$coefficients
```

Para estimar Σ basta resolver la ecuación de estimación:

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B})^T (Y - X\hat{B}) \approx \begin{pmatrix} 47.81 & 17.87 \\ 17.87 & 12.36 \end{pmatrix}$$

En ambos casos se obtienen los mismos resultados. Note que no necesariamente es razonable el supuesto de normalidad dado que en caso de ser normal entonces tanto Y1 como Y2 son normales univaridas. Mediante dos test distintos se puede ver que es razonable que Y1 sea normal pues tiene un p valor relativamente pequeño para los dos test, pero Y2 es todo lo contrario mostrando que no es razonable afirmar que es normal. Por ende no es claro que Y1, Y2 sean normales y por consecuencia los errores no necesariamente son normales.

```
[47]: ###Prueba de Pearson chi-square###
###basada en una distribución Ji cuadrado y que corresponde a una prueba de bondad de ajuste.###
shapiro.test(Y1)
shapiro.test(Y2)

        Shapiro-Wilk normality test

data:  Y1
W = 0.75671, p-value = 0.009688

        Shapiro-Wilk normality test

data:  Y2
W = 0.88883, p-value = 0.2282
```

```
[48]: ###Prueba de Kurtosis###
## 2000 replicaciones por montecarlo
kurtosis.norm.test(Y1, nrepl=2000)
kurtosis.norm.test(Y2, nrepl=2000)

        Kurtosis test for normality

data:  Y1
T = 4.7102, p-value = 0.0105

        Kurtosis test for normality

data:  Y2
T = 3.42, p-value = 0.7915
```

(b) Viendo los resultados note que el coeficiente que acompaña a x1 para Y1 e Y2 tienen valores p pequeños. Con una significancia del 0.0309 o superior se puede decir que rechazamos el hecho de que Y1 e Y2 no dependa de x1. En palabras mas simples y menos formales (no adecuadas para un estadístico) diremos que Y1 e Y2 en realidad dependenden linealmente de x1.

```
summary(fit)
Response Y1 :

Call:
lm(formula = Y1 ~ x1)

Residuals:
    Min      1Q  Median      3Q     Max
-12.486  -4.576   0.932   5.432  10.101

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.7524     6.9668  -0.539   0.6095
x1            1.4022     0.4995   2.807   0.0309 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.985 on 6 degrees of freedom
Multiple R-squared:  0.5677,    Adjusted R-squared:  0.4956
F-statistic: 7.878 on 1 and 6 DF,  p-value: 0.03088


Response Y2 :

Call:
lm(formula = Y2 ~ x1)

Residuals:
   Min     1Q Median     3Q    Max
-6.240 -2.253 -0.183  2.173  5.606

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.296      3.543  -0.648 0.540847
x1             2.141      0.254   8.428 0.000152 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.061 on 6 degrees of freedom
Multiple R-squared:  0.9221,    Adjusted R-squared:  0.9091
F-statistic: 71.03 on 1 and 6 DF,  p-value: 0.0001522
```