



**Universidad Nacional de Colombia**

Facultad de Ciencias  
Departamento de Estadística

# **Segundo Taller**

## **Análisis Multivariado**

Juan Andrés Camacho Zárate

jucamachoz@unal.edu.co

Camilo Alejandro Raba Gomez

craba@unal.edu.co

### **Docente**

Ruben Dario Guevara Gonzalez

Febrero 2025

# 1. Analisis de Componentes Principales

## Fundamentacion.

El **Método de los Componentes Principales** (PCA) es una técnica de reducción de la dimensión que permite transformar un conjunto de variables posiblemente correlacionadas en un conjunto de nuevas variables no correlacionadas, denominadas **componentes principales**. Estas componentes retienen la mayor cantidad de variabilidad presente en los datos originales.

Dado un conjunto de datos con  $p$  variables, se busca una transformación lineal que permita encontrar nuevas variables  $Z_1, Z_2, \dots, Z_p$  tales que:

$$Z_k = a_{k1}X_1 + a_{k2}X_2 + \dots + a_{kp}X_p, \quad (1)$$

donde los coeficientes  $a_{kj}$  son los **autovectores** de la matriz de covarianza  $\Sigma$  o de la matriz de correlación  $R$ , y los valores propios asociados indican la varianza explicada por cada componente principal.

## Pasos del Método

1. **Estandarización:** Si las variables tienen escalas diferentes, se recomienda estandarizar los datos restando la media y dividiendo por la desviación estándar.
2. **Cálculo de la matriz de covarianza o de correlación  $\Sigma$  o  $R$ .**
3. **Cálculo de autovalores y autovectores:** Se resuelve el problema característico:

$$|\Sigma - \lambda I| = 0. \quad (2)$$

4. **Selección de componentes:** Se ordenan los autovalores en orden descendente y se eligen los primeros  $k$  componentes que explican un porcentaje suficiente de la varianza total.
5. **Transformación de los datos:** Se proyectan los datos originales en la nueva base ortogonal formada por los autovectores seleccionados.

## Interpretación

Cada componente principal  $Z_k$  representa una combinación lineal de las variables originales, capturando la máxima varianza posible en cada paso. En la práctica, se eligen los primeros  $k$  componentes que explican al menos un umbral del 80-90% de la variabilidad total.

## Aplicaciones

El PCA se utiliza en diversas áreas como:

- Reducción de la dimensión en modelos de aprendizaje automático.
- Compresión de imágenes y eliminación de ruido.
- Análisis exploratorio de datos para identificar patrones.
- Reducción de multicolinealidad en modelos de regresión.

**Muestre que los componentes principales son combinaciones lineales de las variables que componen un vector aleatorio.**

Sea el vector aleatorio:

$$\mathbf{X} = (X_1, X_2, \dots, X_p)^T, \quad (3)$$

con media  $E[\mathbf{X}] = \boldsymbol{\mu}$  y matriz de covarianza  $\Sigma$ . Los componentes principales se definen como:

$$Z_k = \mathbf{a}_k^T \mathbf{X} = \sum_{j=1}^p a_{kj} X_j, \quad (4)$$

donde  $\mathbf{a}_k$  es un vector de coeficientes que buscamos determinar.

### Criterio de Varianza Máxima

El primer componente principal maximiza la varianza:

$$\max_{\mathbf{a}_1} \text{Var}(Z_1) = \max_{\mathbf{a}_1} \mathbf{a}_1^T \Sigma \mathbf{a}_1, \quad (5)$$

sueto a la restricción de normalización:

$$\mathbf{a}_1^T \mathbf{a}_1 = 1. \quad (6)$$

### Solución por Autovalores

Usando multiplicadores de Lagrange, la condición de primer orden es:

$$\Sigma \mathbf{a}_1 = \lambda_1 \mathbf{a}_1. \quad (7)$$

Esto muestra que  $\mathbf{a}_1$  es un autovector de  $\Sigma$  y  $\lambda_1$  su correspondiente autovalor. Se elige el mayor  $\lambda_1$  para capturar la máxima varianza.

Generalización

Para los siguientes componentes principales, se impone ortogonalidad:

$$\Sigma \mathbf{a}_k = \lambda_k \mathbf{a}_k, \quad \text{con} \quad \mathbf{a}_i^T \mathbf{a}_j = 0, \quad \forall i \neq j. \quad (8)$$

### Demostración de que la varianza total es la suma de los valores propios de la matriz de covarianza

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$  un vector aleatorio de dimensión  $p$  con media  $\boldsymbol{\mu}$  y matriz de covarianza  $\Sigma$ , definida como:

$$\Sigma = \text{Var}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T].$$

### Definición de Varianza Total

La varianza total de  $\mathbf{X}$  se define como la traza de la matriz de covarianza  $\Sigma$ :

$$\text{Var}_{\text{total}}(\mathbf{X}) = \text{tr}(\Sigma).$$

Dado que la traza de una matriz es la suma de sus elementos diagonales, y cada elemento diagonal  $\sigma_{ii}$  representa la varianza de la variable  $X_i$ , tenemos:

$$\text{Var}_{\text{total}}(\mathbf{X}) = \sum_{i=1}^p \sigma_{ii}.$$

### Propiedad de la Traza y Autovalores

La matriz de covarianza  $\Sigma$  es simétrica y semidefinida positiva, lo que implica que tiene una descomposición espectral en términos de autovalores y autovectores:

$$\Sigma = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^\top,$$

donde  $\mathbf{A}$  es una matriz ortonormal de autovectores y  $\mathbf{\Lambda}$  es una matriz diagonal cuyos elementos  $\lambda_i$  son los autovalores de  $\Sigma$ .

Dado que la traza de una matriz es invariante bajo cambio de base ortonormal, se cumple:

$$\text{tr}(\Sigma) = \text{tr}(\mathbf{A}\mathbf{A}^\top \mathbf{A}) = \text{tr}(\mathbf{\Lambda}) = \sum_{i=1}^p \lambda_i.$$

Como  $\text{Var}_{\text{total}}(\mathbf{X}) = \text{tr}(\Sigma)$ , obtenemos:

$$\text{Var}_{\text{total}}(\mathbf{X}) = \sum_{i=1}^p \lambda_i.$$

### Demostración de que el valor esperado de los Componentes Principales es cero

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$  un vector aleatorio de dimensión  $p$  con media  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  y matriz de covarianza  $\Sigma$ . Los Componentes Principales se definen como combinaciones lineales de  $\mathbf{X}$ :

$$Z_k = \mathbf{a}_k^\top \mathbf{X}, \quad k = 1, 2, \dots, p,$$

donde  $\mathbf{a}_k$  es el autovector correspondiente al autovalor  $\lambda_k$  de la matriz de covarianza  $\Sigma$ .

### Cálculo del Valor Esperado

Tomando esperanza en ambos lados de la ecuación de los componentes principales:

$$\mathbb{E}[Z_k] = \mathbb{E}[\mathbf{a}_k^\top \mathbf{X}].$$

$$\mathbb{E}[Z_k] = \mathbf{a}_k^\top \mathbb{E}[\mathbf{X}].$$

Si centramos los datos restando la media a cada variable, obtenemos el vector centrado:

$$\mathbf{X}_c = \mathbf{X} - \boldsymbol{\mu}.$$

Esto implica que  $\mathbb{E}[\mathbf{X}_c] = \mathbf{0}$ , y como los Componentes Principales se calculan a partir de los datos centrados, se cumple:

$$\mathbb{E}[Z_k] = \mathbf{a}_k^\top \mathbb{E}[\mathbf{X}_c] = \mathbf{a}_k^\top \mathbf{0} = 0.$$

## Demostración de que la varianza del $i$ -ésimo componente principal $Z_i$ es el $i$ -ésimo valor propio de la matriz de covarianzas

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$  un vector aleatorio de dimensión  $p$  con media  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ . Los Componentes Principales se definen como combinaciones lineales de  $\mathbf{X}$ :

$$Z_i = \mathbf{a}_i^\top \mathbf{X}, \quad i = 1, 2, \dots, p,$$

donde  $\mathbf{a}_i$  es el autovector correspondiente al autovalor  $\lambda_i$  de la matriz de covarianza  $\boldsymbol{\Sigma}$ .

### Cálculo de la Varianza de $Z_i$

Usando la propiedad de la varianza para transformaciones lineales:

$$\text{Var}(Z_i) = \mathbf{a}_i^\top \boldsymbol{\Sigma} \mathbf{a}_i.$$

### Propiedad de los Autovectores

Dado que  $\mathbf{a}_i$  es un autovector de  $\boldsymbol{\Sigma}$  asociado al autovalor  $\lambda_i$ , se cumple:

$$\boldsymbol{\Sigma} \mathbf{a}_i = \lambda_i \mathbf{a}_i.$$

Multiplicando por  $\mathbf{a}_i^\top$  en ambos lados:

$$\mathbf{a}_i^\top \boldsymbol{\Sigma} \mathbf{a}_i = \mathbf{a}_i^\top (\lambda_i \mathbf{a}_i).$$

Dado que los autovectores están normalizados, es decir,  $\mathbf{a}_i^\top \mathbf{a}_i = 1$ , se concluye:

$$\mathbf{a}_i^\top \boldsymbol{\Sigma} \mathbf{a}_i = \lambda_i.$$

Esto demuestra el resultado deseado.

## Importancia de la $k$ -ésima variable en el $i$ -ésimo componente principal

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$  un vector aleatorio de dimensión  $p$  con media  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ . Los componentes principales se definen como combinaciones lineales de las variables originales:

$$Z_i = \mathbf{a}_i^\top \mathbf{X} = \sum_{k=1}^p a_{ik} X_k, \quad i = 1, 2, \dots, p,$$

donde  $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{ip})^\top$  es el autovector asociado al autovalor  $\lambda_i$  de la matriz de covarianza  $\boldsymbol{\Sigma}$ .

- **Cargas factoriales:** Si los datos han sido estandarizados, las cargas factoriales son:

$$\text{Carga}(X_k, Z_i) = a_{ik} \sqrt{\lambda_i}.$$

Estas indican la correlación entre la variable original  $X_k$  y el componente principal  $Z_i$ .

## **Demostración de que la distancia de Mahalanobis al cuadrado es la distancia euclidiana de los componentes principales estandarizados al cuadrado**

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$  un vector aleatorio de dimensión  $p$  con media  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ . La **distancia de Mahalanobis** de un punto  $\mathbf{X}$  con respecto a la media  $\boldsymbol{\mu}$  se define como:

$$d_M^2(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}).$$

### **Expresión en términos de Componentes Principales**

Los componentes principales están definidos como:

$$\mathbf{Z} = \mathbf{A}^\top (\mathbf{X} - \boldsymbol{\mu}),$$

donde:

- $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p]$  es la matriz de autovectores de  $\boldsymbol{\Sigma}$ ,
- $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)^\top$  es el vector de componentes principales.

La matriz de covarianza en la base de los componentes principales es diagonal:

$$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p),$$

donde  $\lambda_i$  son los autovalores de  $\boldsymbol{\Sigma}$ . Como  $\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^\top$ , su inversa es:

$$\boldsymbol{\Sigma}^{-1} = \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^\top.$$

Sustituyendo en  $d_M^2(\mathbf{X})$ :

$$d_M^2(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^\top (\mathbf{X} - \boldsymbol{\mu}).$$

Definiendo los componentes principales estandarizados como:

$$\tilde{Z}_i = \frac{Z_i}{\sqrt{\lambda_i}},$$

y utilizando el hecho de que  $\mathbf{A}^\top (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}$ :

$$d_M^2(\mathbf{X}) = \sum_{i=1}^p \frac{Z_i^2}{\lambda_i}.$$

### **Expresión en términos de la Distancia Euclidiana**

Dado que la distancia euclidiana en el espacio de los componentes principales estandarizados es:

$$d_E^2(\tilde{\mathbf{Z}}) = \sum_{i=1}^p \tilde{Z}_i^2 = \sum_{i=1}^p \frac{Z_i^2}{\lambda_i},$$

se concluye que:

$$d_M^2(\mathbf{X}) = d_E^2(\tilde{\mathbf{Z}}).$$

## **Demostración de que el vector de variables estandarizadas tiene media cero y matriz de covarianza igual a la matriz de correlación**

Sea  $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$  un vector aleatorio de dimensión  $p$  con media  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma} = (\sigma_{ij})$ . Definimos la versión estandarizada de cada variable como:

$$Z_k = \frac{X_k - \mu_k}{\sigma_k}, \quad k = 1, 2, \dots, p,$$

donde  $\sigma_k = \sqrt{\sigma_{kk}}$  es la desviación estándar de  $X_k$ . El vector de variables estandarizadas se denota como:

$$\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)^\top.$$

### **Media del Vector Estandarizado**

Calculamos la esperanza de cada componente  $Z_k$ :

$$\mathbb{E}[Z_k] = \mathbb{E}\left[\frac{X_k - \mu_k}{\sigma_k}\right].$$

Usando la linealidad de la esperanza:

$$\mathbb{E}[Z_k] = \frac{1}{\sigma_k} \mathbb{E}[X_k - \mu_k] = \frac{1}{\sigma_k} (\mathbb{E}[X_k] - \mu_k) = \frac{1}{\sigma_k} (\mu_k - \mu_k) = 0.$$

Dado que esto se cumple para cada  $k = 1, \dots, p$ , concluimos que:

$$\mathbb{E}[\mathbf{Z}] = \mathbf{0}.$$

### **Matriz de Covarianza del Vector Estandarizado**

La matriz de covarianza de  $\mathbf{Z}$  se define como:

$$\text{Cov}(\mathbf{Z}) = \mathbb{E}[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])^\top].$$

Dado que  $\mathbb{E}[\mathbf{Z}] = \mathbf{0}$ , esto se simplifica a:

$$\text{Cov}(\mathbf{Z}) = \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top].$$

El elemento  $(i, j)$  de esta matriz es:

$$\text{Cov}(Z_i, Z_j) = \mathbb{E}\left[\left(\frac{X_i - \mu_i}{\sigma_i}\right)\left(\frac{X_j - \mu_j}{\sigma_j}\right)\right].$$

Factorizando constantes:

$$\text{Cov}(Z_i, Z_j) = \frac{1}{\sigma_i \sigma_j} \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)].$$

Dado que la covarianza de  $X_i$  y  $X_j$  es:

$$\text{Cov}(X_i, X_j) = \sigma_{ij},$$

obtenemos:

$$\text{Cov}(Z_i, Z_j) = \frac{\sigma_{ij}}{\sigma_i \sigma_j}.$$

Por definición, la matriz de correlación de  $\mathbf{X}$  está dada por:

$$\mathbf{R} = (r_{ij}), \quad r_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}.$$

Por lo tanto, concluimos que:

$$\text{Cov}(\mathbf{Z}) = \mathbf{R}.$$

## 8.1

Determine los componentes principales de la población  $Y_1$  y  $Y_2$  para la siguiente matriz de covarianza:

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

Además, calcule la proporción de la varianza total de la población explicada por el primer componente principal.

### Solución:

Computamos la matriz:

```
sigma <- matrix(c(5,2,
                  2,2), ncol = 2, byrow = T)
```

Se hayan los valores y vectores propios de  $\Sigma$

```
sigma_eigen <- eigen(sigma)
```

```
print(sigma_eigen$values)
```

```
## [1] 6 1
```

```
print(sigma_eigen$vectors)
```

```
##           [,1]      [,2]
## [1,] -0.8944272  0.4472136
## [2,] -0.4472136 -0.8944272
```

De lo cual concluimos que las componentes principales poblacionales tiene la siguiente forma:

$$Y_1 = -0.894X_1 - 0.4472X_2$$

$$Y_2 = 0.4472X_1 - 0.8944X_2$$

Procedemos a calcular la proporción de varianza explicada por la primera componente principal:

```
sigma_values[1]/sum(sigma_values)
```

```
## [1] 0.8571429
```

Concluimos que el primer componente principal explica aproximadamente un 86% de la varianza total.



## 8.2

Convierta la matriz de covarianza en el Ejercicio 8.1 en una matriz de correlación  $\rho$ .

### Solución:

De 8.1 obtenemos la siguiente matriz desviación estándar :

$$\mathbb{V}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

Luego:

```
(R <- solve(V) %*% sigma %*% solve(V))
```

```
##           [,1]      [,2]  
## [1,] 1.0000000 0.6324555  
## [2,] 0.6324555 1.0000000
```

$$\rho = \begin{pmatrix} 1 & 0.63 \\ 0.63 & 1 \end{pmatrix}$$

(a)

Determine los componentes principales  $Y_1$  y  $Y_2$  a partir de  $\rho$  y calcule la proporción de la varianza total de la población explicada por  $Y_1$ .

Se hallan los autovalores y autovectores de  $\rho$ :

```
R_eigen <- eigen(R)
```

```
(R_values <- R_eigen$values)
```

```
## [1] 1.6324555 0.3675445
```

```
(R_vectors <- R_eigen$vectors)
```

```
##           [,1]      [,2]  
## [1,] -0.7071068  0.7071068  
## [2,] -0.7071068 -0.7071068
```

Así, los componentes principales son:

$$Y_1 = -0.7071X_1 - 0.7071X_2$$

$$Y_2 = 0.7071X_1 - 0.7071X_2$$

```
R_values[1]/2
```

```
## [1] 0.8162278
```

Y la proporción poblacional de varianza explicada por el primer componente principal es aproximadamente del 81.62% .

(b)

Compare los componentes calculados en la Parte (a) con aquellos obtenidos en el Ejercicio 8.1.

¿Son iguales? ¿Deberían serlo?

Los componentes principales de  $\Sigma$  y  $\rho$  no necesariamente son iguales, ya que la estandarización afecta la dirección de los autovectores y la magnitud de los vectores propios lo cual lleva a modificar los componentes

principales. Por ello, los resultados pueden diferir al comparar la matriz de covarianza con la matriz de correlación.

(c)

Compute las correlaciones  $\rho_{Y_1, Z_1}$ ,  $\rho_{Y_1, Z_2}$ , y  $\rho_{Y_2, Z_1}$ .

Ya que estamos trabajando con la matriz de correlacion (datos estandarizados), a partir del resultado 8.4 tenemos que:

```
rho11 <- R_vectors[1,1]*sqrt(R_values[1])
rho12 <- R_vectors[2,1]*sqrt(R_values[1])
rho21 <- R_vectors[1,2]*sqrt(R_values[2])
```

Los resultados obtenidos son

$$\begin{aligned}\rho_{Y_1, Z_1} &= -0.9034532 \\ \rho_{Y_1, Z_2} &= -0.9034532 \\ \rho_{Y_2, Z_1} &= 0.4286866\end{aligned}$$

Los resultados muestran que  $Y_1$  está altamente correlacionado con  $Z_1$  y  $Z_2$  ( $\rho = -0.9034$ ), mientras que  $Y_2$  tiene una correlación más baja con  $Z_1$  ( $\rho = 0.4289$ ).

Esto nos indica que valores bajos en  $Y_1$  se ven representados en valores altos en el primer plano factorial.

### 8.3

Sea

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Determine las componentes principales  $Y_1, Y_2$ , y  $Y_3$ . ¿Qué puede decir acerca de los autovectores (y componentes principales) asociados con los autovalores que no son distintos?

**Solución:**

Primero, computamos la matriz:

```
Sigma <- matrix(c(2, 0, 0,
                  0, 4, 0,
                  0, 0, 4),
                nrow = 3, ncol = 3, byrow = TRUE)
```

Los valores propios obtenidos son:

```
eigen(Sigma)$values
```

```
## [1] 4 4 2
```

Note que hay un valor con multiplicidad algebraica de 2. Para  $\lambda = 4$ , tenemos lo siguiente:

$$\Sigma - 4I = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Buscamos los vectores  $v = (x, y, z)^T$  tal que:

$$\begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Esto nos da el sistema:

$$\begin{aligned} -2x &= 0 \Rightarrow x = 0 \\ 0y &= 0 \Rightarrow y \in \mathbb{R} \\ 0z &= 0 \Rightarrow z \in \mathbb{R} \end{aligned}$$

Por lo tanto, los vectores propios asociados a  $\lambda = 4$  son todos los vectores de la forma:

$$v = C_1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + C_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

donde  $C_1, C_2 \in \mathbb{R}$  son constantes arbitrarias.

Por lo que se concluye que, los vectores propios de sigma son:

```
eigen(Sigma)$vectors
```

```
##      [,1] [,2] [,3]
## [1,]    0    0    1
## [2,]    0    1    0
## [3,]    1    0    0
```

Y los componentes principales serán de la forma

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_2 \\ Y_3 &= X_3 \end{aligned}$$

Dado que la matriz de covarianza  $\Sigma$  tiene autovalores  $\lambda = 4, 4, 2$ , observamos que  $\lambda = 4$  es un autovalor repetido. En este caso, los autovectores asociados a  $\lambda = 4$  no están completamente determinados, ya que cualquier combinación lineal de ellos también es un autovector válido. Dado que la combinación lineal  $\mathbf{e}_i' \mathbf{X} = X_i$ , el conjunto de componentes principales es simplemente el conjunto original de variables aleatorias no correlacionadas.

Para una matriz de covarianza con tal patrón, no se obtiene ninguna ganancia al extraer las componentes principales.

## 8.4

Encuentre los componentes principales y la proporción de la varianza total de la población explicada por cada uno cuando la matriz de covarianza es:

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\rho & 0 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ 0 & \sigma^2\rho & \sigma^2 \end{bmatrix}, \quad -\frac{1}{\sqrt{2}} < \rho < \frac{1}{\sqrt{2}}$$

### Solución:

Se realiza el proceso para hallar los valores propios de  $\Sigma$ :

Al hallar el determinante de sigma e igualarlo a cero obtenemos el siguiente polinomio característico:

$$P(\Sigma) = -\lambda^3 + 3\sigma^2\lambda^2 - (\sigma^4 - \sigma^4\rho^2 + \sigma^4\rho^2)\lambda + (\sigma^6 - 2\sigma^6\rho^2) = 0$$

Realizando el álgebra necesaria se obtiene que:

$$P(\Sigma) = -(\lambda - \sigma^2)(\lambda^2 - 2\sigma^2\lambda + \sigma^4 - 2\sigma^4\rho^2)$$

Así para hallar los autovalores que solucionan la ecuación cuadrática usamos:

$$\begin{aligned} \lambda &= \frac{2\sigma^2 \pm \sqrt{4\sigma^4 - 4(\sigma^4 - 2\sigma^4\rho^2)}}{2} \\ &= \sigma^2(1 \pm \rho\sqrt{2}) \end{aligned}$$

La solución a  $(\Sigma - I\sigma^2)\vec{v} = \vec{0}$  será el vector propio asociado a  $\sigma^2$ :

$$\begin{bmatrix} 0 & \sigma^2\rho & 0 \\ \sigma^2\rho & 0 & \sigma^2\rho \\ 0 & \sigma^2\rho & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

La solución que obtenemos es:  $\vec{v}_1 = [1 \quad 0 \quad -1]^T$ .

Tomando  $\lambda_2 = \sigma^2(1 + \rho\sqrt{2})$ , procedemos a resolver  $\Sigma - I\lambda_2$ :

$$\begin{bmatrix} -\sigma^2\rho\sqrt{2} & \sigma^2\rho & 0 \\ \sigma^2\rho & -\sigma^2\rho\sqrt{2} & \sigma^2\rho \\ 0 & \sigma^2\rho & -\sigma^2\rho\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

La solución que obtenemos es:  $\vec{v}_2 = \left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}} \quad 0\right]^T$

Tomando  $\lambda_3 = \sigma^2(1 - \rho\sqrt{2})$  procedemos a resolver  $\Sigma - I\lambda_3$ :

$$\begin{bmatrix} \sigma^2\rho\sqrt{2} & \sigma^2\rho & 0 \\ \sigma^2\rho & \sigma^2\rho\sqrt{2} & \sigma^2\rho \\ 0 & \sigma^2\rho & \sigma^2\rho\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Con solución  $\vec{v}_3 = \left[\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}} \quad 0\right]^T$

Con lo cual concluimos que las componentes principales tienen la siguiente estructura:

$$Z_1 = \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_3$$

$$Z_2 = \frac{1}{2}X_1 + \frac{1}{\sqrt{2}}X_2 + \frac{1}{2}X_3$$

$$Z_3 = \frac{1}{2}X_1 - \frac{1}{\sqrt{2}}X_2 + \frac{1}{2}X_3$$

Cualculando la proporción de varianza explicada:

Para  $Z_1$ :

$$\frac{\sigma^2}{\sigma^2 + 2\sigma^2} = \frac{1}{3} \approx 0.33$$

Para  $Z_2$  y  $Z_3$ :

$$0 < \frac{1 \pm \rho\sqrt{2}}{3} < \frac{2}{3} \approx 0.66$$

## 8.5

(a)

Halle los autovalores de la matriz de correlación

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

¿Los resultados son consistentes con (8-16) y (8-17)?

**Solución:**

Igualando el determinante de  $\rho$  a 0 obtenemos la siguiente ecuación propia:

$$\begin{aligned} \det(\boldsymbol{\rho} - \lambda I) &= \begin{vmatrix} 1-\lambda & \rho & \rho \\ \rho & 1-\lambda & \rho \\ \rho & \rho & 1-\lambda \end{vmatrix} \\ &= (1-\lambda)^3 + 2\rho^2 - 3\rho^2(1-\lambda) \\ &= -\lambda^3 + 3\lambda^2 + 3\lambda(\rho^2 - 1) + (2\rho + 1)(\rho - 1)^2 \end{aligned}$$

Realizando división sintética por  $\lambda - (2\rho + 1)$  obtenemos que:  $(\lambda - (1 - \rho))^2(\lambda - (2\rho + 1)) = 0$

De donde concluimos que los autovalores son:

$$\begin{aligned} \lambda_1 &= 2\rho + 1 \\ \lambda_2 &= \lambda_3 = 1 - \rho \end{aligned}$$

El resultado obtenido es consistente con (8-16) para  $p = 3$ .

(b)

Verifique los pares autovalor-autovector para la matriz  $p \times p$   $\rho$  dada en (8-15).

Al operar

$$\rho \left( \frac{1}{\sqrt{p}} \mathbf{1} \right)$$

se tiene que:

$$\begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \left( \frac{1}{\sqrt{p}} \mathbf{1} \right) = \left( (1 + 2\rho) \frac{1}{\sqrt{p}} \mathbf{1} \right)$$

Así  $\left( \frac{1}{\sqrt{p}} \mathbf{1} \right)$  es el vector propio asociado al valor propio  $1 + 2\rho$  como en (8-15) para  $p = 3$ .

Adicionalmente los restantes autovectores son de la forma:

$$\rho e_i = (1 - \rho) e_i, \quad i = 2, 3$$

Es decir, existen  $3 - 1 = 2$  autovectores ortogonales al vector de unos, y que todos comparten el mismo autovalor  $1 - \rho$ .

Esto último también es consistente con (8-15).

## 8.6

Los datos de  $x_1 =$  ventas y  $x_2 =$  beneficios para las 10 empresas más grandes del mundo fueron listados en el Ejercicio 1.4 del Capítulo 1.

Del Ejemplo 4.12:

$$\bar{\mathbf{x}} = \begin{bmatrix} 155.60 \\ 14.70 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{bmatrix}$$

(a)

Determine los componentes principales muestrales y sus varianzas para estos datos. (Es posible que necesite la fórmula cuadrática para resolver los autovalores de  $\mathbf{S}$ ).

**Solución:**

Primero hallamos los valores y vectores propios estimados:

```
S_eigen <- eigen(S)
```

```
(S_values <- S_eigen$values)
```

```
## [1] 7488.80293 13.83707
```

```
(S_vectors <- cbind(S_eigen$vectors[1:2, 1]*(-1), S_eigen$vectors[1:2, 2]))
```

```
##           [,1]      [,2]
## [1,] 0.99917337 0.04065185
## [2,] 0.04065185 -0.99917337
```

De donde obtenemos que:

$$\hat{\lambda}_1 = 7488.80293, \quad \hat{\mathbf{e}}_1 = \begin{bmatrix} 0.99917337 \\ 0.04065185 \end{bmatrix}$$

$$\hat{\lambda}_2 = 13.83707, \quad \hat{\mathbf{e}}_2 = \begin{bmatrix} 0.04065185 \\ -0.99917337 \end{bmatrix}$$

Concluimos que los componentes principales son de la forma:

$$\begin{aligned}\hat{y}_1 &= 0.99917x_1 + 0.0407x_2 \\ \hat{y}_2 &= 0.04065x_1 - 0.9991x_2\end{aligned}$$

Usando las propiedades de los componentes principales, se tiene que  $\text{Var}(\hat{\mathbf{Z}}_i) = \lambda_i$ . Así la varianza muestral de los componentes principales es la siguiente:

	$\hat{y}_1$	$\hat{y}_2$
$\hat{\lambda}_i$	7488.8	13.8

(b)

Encuentre la proporción de la varianza total de la muestra explicada por  $\hat{y}_1$ .

**Solución:**

Calculamos:

```
S_values[1]/sum(S_values)
```

```
## [1] 0.9981557
```

Como resultado tenemos que  $\hat{y}_1$  explica aproximadamente el 99.8% de la varianza total.

(c)

Dibujar la elipse de densidad constante

$$(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = 1.4$$

Indique los componentes principales  $\hat{y}_1$  y  $\hat{y}_2$  en su gráfico.

**Solución**

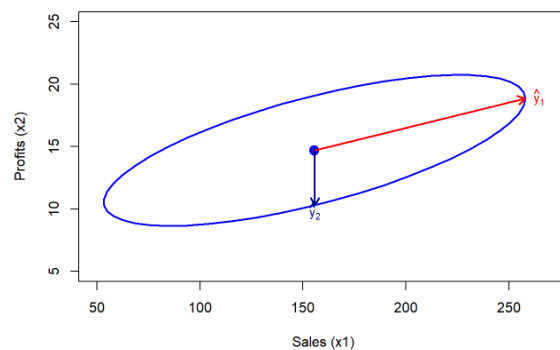


Figure 1: Elipse y Componentes principales.

(d)

Calcule los coeficientes de correlación  $r_{\hat{y}_1, x_k}$ ,  $k = 1, 2$ .

¿Qué interpretación, si alguna, puede dar al primer componente principal?

**Solución:**

Usando el resultado 8.3:

$$r_{\hat{y}_1, x_k} = \frac{\hat{e}_{1k} \sqrt{\hat{\lambda}_1}}{\sqrt{s_{kk}}}, \quad k = 1, 2$$

$$r_{\hat{y}_1, x_1} = 0.9999985$$

$$r_{\hat{y}_1, x_2} = 0.6874146$$

Se ve que ambas variables influyen sustancialmente sobre el primer componente principal, lo cual es coherente con el porcentaje de varianza explicada que se hayó. Al ser prácticamente uno la correlación entre  $x_1$  y  $\hat{y}_1$  se evidencia que la variabilidad de las ventas aporta e influencia mucho al primer componente principal.

## 8.7

Convierta la matriz de covarianza **S** en el Ejercicio 8.6 en una matriz de correlación muestral **R**.

**Solución:** Primero hallamos la matriz desviación estándar:

$$\mathbb{V}^{\frac{1}{2}} = \begin{bmatrix} 86.46647 & 0 \\ 0 & 5.117617 \end{bmatrix}$$

Luego, estandarizando obtenemos:

```
(R <- solve(V) %*% S %*% solve(V))
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.6861434
## [2,] 0.6861434 1.0000000
```

Por tanto, la matriz de correlación muestral es:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.6861434 \\ 0.6861434 & 1 \end{pmatrix}$$

(a)

Encuentre las componentes principales  $\hat{y}_1, \hat{y}_2$  y sus varianzas.

Calculamos los vectores y valores propios de **R**:

```
R_values <- (eigen(R))$values
```

```
R_vectors <- (eigen(R))$vectors
```

$$\hat{\lambda}_1 = 1.6861434$$

$$\hat{\lambda}_2 = 0.3138566$$



$$\hat{\mathbf{e}}_1 = (0.7071068 \quad 0.7071068)^T$$

$$\hat{\mathbf{e}}_2 = (-0.7071068 \quad 0.7071068)^T$$

Dado el resultado anterior concluimos que los componentes principales son de la forma:

$$\hat{y}_1 = 0.7071068z_1 + 0.7071068z_2$$

$$\hat{y}_2 = -0.7071068z_1 + 0.7071068z_2$$

Y sus respectivas varianzas muestrales son

	$\hat{y}_1$	$\hat{y}_2$
$\hat{\lambda}_i$	1.6861434	0.3138566

(b)

Compute la proporción de varianza total explicada por  $\hat{y}_1$ .

```
R_values[1]/sum(R_values)
```

```
## [1] 0.8430717
```

Obtenemos como resultado que aproximadamente un 84% de la varianza total es explicada por el componente  $\hat{y}_1$ .

(c)

Compute los coeficientes de correlación  $r_{\hat{y}_i, x_k}, k = 1, 2$ . Interprete  $\hat{y}_1$ .

$$r_{\hat{y}_1, z_1} = 0.9181894$$

$$r_{\hat{y}_1, z_2} = 0.9181894$$

Ambos componentes son iguales y positivos, por ende, contribuyen de forma equitativa al primer componente principal  $\hat{y}_1$ .

(d)

Compare los componentes obtenidos en la Parte (a) con aquellos obtenidos en el Ejercicio 8.6(a). Dado los datos originales presentados en el Ejercicio 1.4, ¿considera que es mejor determinar los componentes principales a partir de la matriz de covarianza muestral o de la matriz de correlación muestral? Explique su respuesta.

Al realizar el proceso de componentes principales en la parte a. usando la matriz de covarianza se debió tener en cuenta la diferencia de unidades ya que esto puede dañar el análisis. Al ser las unidades de  $x_1$  mucho mayores, su aporte a las componentes principales se vio desproporcionado. En contraparte, cuando se uso un análisis basado en la matriz de correlaciones se observó como ese efecto se mitigó y se logró recoger la variación de mejor forma.

En conclusión, el segundo análisis es mejor ya que en este se representa mejor la variabilidad misma mientras que en el primer análisis gran parte de la variabilidad se vio absorbida por las notables diferencias de unidades.

## 8.8

Use los resultados del ejemplo 8.5.

$$\bar{\mathbf{x}}' = [0.0011, 0.0007, 0.0016, 0.0040, 0.0040]$$

$$\mathbf{R} = \begin{bmatrix} 1.000 & 0.632 & 0.511 & 0.115 & 0.155 \\ 0.632 & 1.000 & 0.574 & 0.322 & 0.213 \\ 0.511 & 0.574 & 1.000 & 0.183 & 0.146 \\ 0.115 & 0.322 & 0.183 & 1.000 & 0.683 \\ 0.155 & 0.213 & 0.146 & 0.683 & 1.000 \end{bmatrix}$$

$$\hat{\lambda}_1 = 2.437, \quad \hat{e}_1 = [0.469, 0.532, 0.465, 0.387, 0.361]'$$

$$\hat{\lambda}_2 = 1.407, \quad \hat{e}_2 = [-0.368, -0.236, -0.315, 0.585, 0.606]'$$

$$\hat{\lambda}_3 = 0.501, \quad \hat{e}_3 = [-0.604, -0.136, 0.772, 0.093, -0.109]'$$

$$\hat{\lambda}_4 = 0.400, \quad \hat{e}_4 = [0.363, -0.629, 0.289, -0.381, 0.493]'$$

$$\hat{\lambda}_5 = 0.255, \quad \hat{e}_5 = [0.384, -0.496, 0.071, 0.595, -0.498]'$$

(a)

Calcule las correlaciones  $r_{y_i, z_k}$  para  $i = 1, 2$  y  $k = 1, 2, \dots, 5$ . ¿Estas correlaciones refuerzan las interpretaciones dadas a los dos primeros componentes? Explique su respuesta.

**Solución:**

Para hallar las correlaciones usamos el resultado 8.4:

$$r_{\hat{y}_i, z_k} = \hat{e}_{ik} \sqrt{\hat{\lambda}_i}, \quad i = 1, 2, k = 1, 2, 3, 4, 5$$

```
r <- matrix(NA, ncol = 5, nrow = 2)

for (i in c(1, 2)) {
  for (k in 1:5) {
    r[i, k] <- eigenvectors[k, i] * sqrt(lambdas[i])
  }
}

r <- as.data.frame.matrix(r)

colnames(r) <- paste0("k=", 1:5)
rownames(r) <- paste0("i=", 1:2)

r

##           k=1          k=2          k=3          k=4          k=5
## i=1  0.7321509  0.8304995  0.7259066  0.6041416  0.5635533
## i=2 -0.4365107 -0.2799362 -0.3736436  0.6939096  0.7188192
```

Estos resultados respaldan las interpretaciones presentadas en el ejemplo, ya que la primera componente principal muestra correlaciones positivas con todas las variables, lo que indica la tendencia general del mercado. Por otro lado, la segunda componente principal evidencia un contraste entre los precios de las acciones bancarias y las acciones del sector petrolero, reflejando que los valores dentro de un mismo sector tienden a comportarse de manera similar.

(b)

$$H_0 : \rho = \rho_0 = \begin{bmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{bmatrix}$$

contra

$$H_1 : \rho \neq \rho_0$$

al nivel de significancia del 5%. Enumere cualquier supuesto requerido para llevar a cabo esta prueba.

Para llevar a cabo la prueba de hipótesis sobre la estructura de correlación uniforme, se deben cumplir los siguientes supuestos:

1. **Normalidad multivariada:** Los datos provienen de una distribución normal multivariada
2. **Muestra aleatoria:** Los datos deben ser una muestra aleatoria simple.
3. **Tamaño de muestra suficientemente grande:** Para que la aproximación asintótica a la distribución  $\chi^2$  sea válida, el tamaño de la muestra  $n$  debe ser lo suficientemente grande en relación con el número de variables  $p$ .

La base cuenta con un total de 103 observaciones y no se nos menciona que los datos provienen de una muestra aleatoria. Se puede realizar el test.

Para realizar el test calculamos:

$$\bar{r}_k = \frac{1}{p-1} \sum_{\substack{i=1 \\ i \neq k}}^p r_{ik}, \quad k = 1, 2, \dots, p; \quad \bar{r} = \frac{2}{p(p-1)} \sum_{i < k} r_{ik}$$

$$\hat{\gamma} = \frac{(p-1)^2 [1 - (1 - \bar{r})^2]}{p - (p-2)(1 - \bar{r})^2}$$

Y la regla de decisión para rechazar  $H_0$  es:

$$T = \frac{(n-1)}{(1-\bar{r})^2} \left[ \sum_{i < k} (r_{ik} - \bar{r})^2 - \hat{\gamma} \sum_{k=1}^p (\bar{r}_k - \bar{r})^2 \right] > \chi_{\frac{(p+1)(p-2)}{2}}^2(\alpha)$$

$\bar{r}_1$	$\bar{r}_2$	$\bar{r}_3$	$\bar{r}_4$	$\bar{r}_5$
0.3534	0.43525	0.35350	0.32575	0.29925

$$\bar{r} = 0.3534$$

$$\sum_{i < k} (r_{ik} - \bar{r})^2 = 0.4487024$$

$$\sum_{k=1}^5 (\bar{r}_k - \bar{r})^2 = 0.0103962$$

$$T = 103.1636 > 16.91898 = \chi_9^2(0.05)$$

Luego, con una significancia del 5% existe evidencia estadística suficiente en contra de  $H_0$  y se rechaza la hipótesis nula de equicorrelación entre las variables.

## 8.9.

(a)

Consideremos que la prueba de razón de verosimilitud bajo la teoría normal para  $H_0 : \Sigma$  es la matriz diagonal.

$$\begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{pp} \end{bmatrix}, \quad \sigma_{ii} > 0$$

Mostremos que la prueba es la siguiente: Rechazar  $H_0$  si

$$\Lambda = \frac{|S|^{n/2}}{\prod_{i=1}^p s_{ii}^{n/2}} = |R|^{n/2} < c.$$

Para un tamaño de muestra grande,  $-2 \ln \Lambda$  es aproximadamente  $\chi_{p(p-1)/2}^2$ . Bartlett [?] sugiere que el estadístico de prueba

$$-2 \left[ 1 - \frac{(2p+11)}{6n} \right] \ln \Lambda$$

se use en lugar de  $-2 \ln \Lambda$ . Esto da como resultado una mejor aproximación a la chi-cuadrado. El punto crítico para una muestra grande con nivel  $\alpha$  es

$$\chi_{p(p-1)/2}^2(\alpha).$$

Nótese que probar  $\Sigma = \Sigma_0$  es equivalente a probar  $\rho = I$ .

### Solución:

Mostremos que la prueba de razón de verosimilitud para  $H_0 : \Sigma = \sigma^2 I$  rechaza  $H_0$  si

$$\Lambda = \frac{|S|^{n/2}}{(\text{tr}(S)/p)^{np/2}} = \left[ \frac{\text{media geométrica } \hat{\lambda}_i}{\text{media geométrica } \hat{\lambda}_i} \right]^{np/2} < c.$$

Para un tamaño de muestra grande, Bartlett [?] sugiere que

$$-2 \left[ 1 - \frac{(2p^2 + p + 2)}{6pn} \right] \ln \Lambda$$

es aproximadamente  $\chi_{p(p+2)(p-1)/2}^2$ . Así, el punto crítico para una muestra grande con nivel  $\alpha$  es

$$\chi_{p(p+2)(p-1)/2}^2(\alpha).$$

Esta prueba se denomina prueba de esfericidad, ya que los contornos de densidad constante son esferas cuando  $\Sigma = \sigma^2 I$ .

Con lo proporcionado se tiene que

$$\begin{aligned}
\max_{\mu, \Sigma} L(\mu, \Sigma) &= \frac{e^{-np/2}}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}} \\
&= \frac{e^{-np/2}}{(2\pi)^{np/2} n^{-np/2} \prod_{i=1}^p \left| \sum_{j=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t \right|} \\
&= \frac{e^{-np/2}}{(2\pi)^{np/2} n^{-np/2} \prod_{i=1}^p |(n-1)\mathbf{S}|}
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
\max_{\mu, \Sigma_0} L(\mu, \Sigma_0) &= \prod_{i=1}^p \frac{e^{-n/2}}{(2\pi)^{n/2} n^{-n/2} (\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2)^{n/2}} \\
&= \frac{e^{-np/2}}{(2\pi)^{np/2} n^{-np/2} \prod_{i=1}^p ((n-1)s_{ii})^{n/2}}
\end{aligned}$$

Entonces,

$$\begin{aligned}
\Lambda &= \frac{\max_{\mu, \Sigma_0} L(\mu, \Sigma_0)}{\max_{\mu, \Sigma} L(\mu, \Sigma)} \\
&= \frac{(n-1)^{np/2} |\mathbf{S}|^{n/2}}{(n-1)^{np/2} \prod_{i=1}^p s_{ii}^{n/2}} \\
&= \frac{|\mathbf{S}|^{n/2}}{\prod_{i=1}^p s_{ii}^{n/2}} = |\mathbf{R}|^{n/2} < c.
\end{aligned}$$

(b)

**Solución:** Cuando  $\Sigma = \sigma^2 I$ , usando (4-16) y (4-17), obtenemos:

$$\max_{\mu} L(\mu, \sigma^2 I) = \frac{1}{(2\pi)^{\frac{np}{2}} (\sigma^2)^{\frac{np}{2}}} \exp \left( -\frac{1}{2\sigma^2} \text{tr}[(n-1)S] \right)$$

Así,

$$\begin{aligned}
\max_{\mu, \sigma^2} L(\mu, \sigma^2 I) &= \frac{(np)^{np/2} e^{-np/2}}{(2\pi)^{np/2} (n-1)^{np/2} [\text{tr}(S)]^{np/2}} \\
&= \frac{e^{-np/2}}{(2\pi)^{np/2} (n-1)^{np/2} \left( \frac{1}{p} \text{tr}(S) \right)^{np/2}}
\end{aligned}$$

Bajo  $H_0$ , hay  $p$  medias  $\mu_i$  y una varianza, por lo que la dimensión del espacio de parámetros es:

$$Y_0 = p + 1$$

El caso no restringido tiene dimensión:

$$p + \frac{p(p+1)}{2}$$

Así, la estadística  $\chi^2$  tiene:

$$\frac{p(p+1)}{2} - 1 = \frac{(p+2)(p-1)}{2}$$

grados de libertad.

## 8.10

Las tasas semanales de rendimiento para cinco acciones listadas en la Bolsa de Valores de Nueva York se presentan en la Tabla 8.4. (Consulte los datos de precios de acciones en el siguiente sitio web: [www.prenhall.com/statistics](http://www.prenhall.com/statistics).)

(a)

Construya la matriz de covarianza muestral  $S$  y encuentre los componentes principales muestrales en (8-20). (Nótese que el vector de medias muestrales  $\bar{x}$  se muestra en el Ejemplo 8.5.)

**Solución:**

Usando la matriz  $R$  y la matriz desviación estándar  $D$  hallamos la matriz de varianzas y covarianzas  $S$ :

```
varianzas <- c(0.00043327, 0.00043872, 0.00022398, 0.00072251, 0.00076568)

sigma <- sqrt(varianzas)

R <- matrix(c(1.000, 0.632, 0.511, 0.115, 0.155,
              0.632, 1.000, 0.574, 0.322, 0.213,
              0.511, 0.574, 1.000, 0.183, 0.146,
              0.115, 0.322, 0.183, 1.000, 0.683,
              0.155, 0.213, 0.146, 0.683, 1.000),
            nrow = 5, byrow = TRUE)
D <- diag(sigma)

# Calcular la matriz de covarianzas S
S <- D %*% R %*% D
```

0.0004333	0.0002755	0.0001592	0.0000643	0.0000893
0.0002755	0.0004387	0.0001799	0.0001813	0.0001234
0.0001592	0.0001799	0.0002240	0.0000736	0.0000605
0.0000643	0.0001813	0.0000736	0.0007225	0.0005080
0.0000893	0.0001234	0.0000605	0.0005080	0.0007657

Para hallar los componentes principales:

```
knitr::kable(data.frame(Variable=c("JP_Morgan", "Citibank", "Wells_Fargo",
                                   "Royal_Dutch_Shell", "Exxon_Mobil"), round(eigen(S)$vectors, 3)),
              col.names = c("Variable", "CP1", "CP2", "CP3", "CP4", "CP5"))
```

Variable	CP1	CP2	CP3	CP4	CP5
JP_Morgan	0.223	0.625	0.326	0.664	0.111
Citibank	0.307	0.570	-0.249	-0.421	-0.584
Wells_Fargo	0.155	0.345	-0.039	-0.488	0.786
Royal_Dutch_Shell	0.639	-0.248	-0.643	0.310	0.144
Exxon_Mobil	0.651	-0.322	0.646	-0.217	-0.090

Variable	CP1	CP2	CP3	CP4	CP5
----------	-----	-----	-----	-----	-----

De esta forma, por ejemplo, para el primer componente principal  $\hat{y}_1$ , su estructura es de la forma:

$$\hat{y}_1 = 0.223x_1 + 0.307x_2 + 0.155x_3 + 0.639x_4 + 0.651x_5$$

(b)

Determine la proporción de la varianza muestral total explicada por los tres primeros componentes principales. Interprete estos componentes.

**Solución:**

De (a) tenemos que:

$\lambda_1$	$\lambda_2$	$\lambda_3$
0.00137	7e-04	0.00025

Recordando que la varianza total está dada por la suma de todos los autovalores de la matriz de covarianzas. Al calcular ese valor, tenemos que, la varianza total es:

```
## [1] 0.00258
```

Por tanto, el porcentaje de varianza explicado por los tres primeros componentes principales es de:

```
## [1] "89.9 %"
```

Por otro lado, apoyandonos en el ejemplo 8.5 el primer componente puede interpretarse como una medida de, el segundo tiene una relación importante con el campo industrial, y el tercero no es tan fácil de interpretar como los dos primeros.

(c)

Construya intervalos de confianza simultáneos de Bonferroni al 90% para las varianzas  $\lambda_1, \lambda_2$ , y  $\lambda_3$  de los tres primeros componentes principales de la población  $Y_1, Y_2$ , y  $Y_3$ .

**Solución:**

Usando

$$\frac{\hat{\lambda}_i}{\left(1 + z(\alpha/2p)\sqrt{2/n}\right)} \leq \lambda_i \leq \frac{\hat{\lambda}_i}{\left(1 - z(\alpha/2p)\sqrt{2/n}\right)}$$

Computamos los intervalos:

```
n<-103
alpha<-.1
p<-3

a<-matrix(NA, nrow = p, ncol = 3)
Eigenvalue = paste0("$$\\lambda_", 1:p, "$$")

for (i in 1:p){
  a[i,1]<-Eigenvalue[i]
  a[i,2]<-round(eigen(S)$values[i]/(1+qnorm(1-alpha/p)*sqrt(2/n)),4)
  a[i,3]<-round(eigen(S)$values[i]/(1-qnorm(1-alpha/p)*sqrt(2/n)),4)
```

```
}
knitr::kable(a, col.names = c("Lambda", "LI", "LS"))
```

Lambda	LI	LS
$\lambda_1$	0.00106	0.00195
$\lambda_2$	0.00054	0.00100
$\lambda_3$	0.00019	0.00036

(d)

Dado los resultados en las Partes (a)-(c), ¿considera que los datos de tasas de rendimiento de las acciones pueden resumirse en menos de cinco dimensiones? Explique.

**Solución:**

Sí se pueden resumir, dado que:

- Los tres primeros componentes principales explican el 89.9% de la varianza total.
- En datos financieros, generalmente dos componentes resumen tendencias del mercado e industria.
- Los últimos dos componentes tienen valores propios pequeños, aportando poca información adicional.
- Esto sugiere que se pueden reducir las dimensiones a dos o tres sin perder información importante.
- Reducir las dimensiones mejora la interpretación sin afectar el análisis de los datos.

## 8.12

Considere los datos de contaminación del aire listados en la Tabla 1.5. Su tarea es resumir estos datos en menos de  $p = 7$  dimensiones si es posible.

Realice un análisis de componentes principales (ACP) utilizando tanto la matriz de covarianza  $S$  como la matriz de correlación  $R$ .

¿Qué ha aprendido del análisis? ¿Hace alguna diferencia qué matriz se elija para el análisis? ¿Es posible resumir los datos en tres o menos dimensiones? ¿Puede interpretar los componentes principales obtenidos?

**Solución:**

Primero, observemos el porcentaje de varianza que explican los componentes usando la matriz de varianzas y covarianzas muestral  $S_X$ :

Componente Principal	Porcentaje Explicado	Porcentaje Acumulado
1	87.295	87.295
2	8.112	95.407
3	3.289	98.969
4	0.724	99.421
5	0.367	99.788
6	0.152	99.939
7	0.060	100.000

Table 8: Porcentaje de varianza explicada por cada componente principal



En la tabla se observa que es conveniente tomar únicamente los dos primeros componentes principales. Ahora, realizando el proceso con la matriz de correlación:

Componente Principal	Porcentaje Explicado	Porcentaje Acumulado
1	33.382	33.382
2	19.800	53.182
3	17.200	70.384
4	10.387	80.705
5	9.334	90.159
6	7.667	97.773
7	2.227	100.000

Table 9: Porcentaje de varianza explicada por cada componente principal

Como se puede ver, no es óptimo reducir la dimensión y usar dos componentes.

Se concluye que gran parte de la variación recogida por los componentes principales hechos a través de  $S_X$  se ve influenciada por las diferencias de escala. En casos como estos es apropiado estandarizar y trabajar con la matriz de correlación.

## 8.25

Refiérase a los datos de horas extra de la policía en el Ejemplo 8.10. Construya un gráfico de control alternativo, basado en la suma de cuadrados  $d_{ij}^2$ , para monitorear la variación no explicada en las observaciones originales resumidas por los componentes principales adicionales.

### Solución:

Para trabajar con las componentes que no fueron tomadas en cuenta en el ejemplo 8.10 usamos el siguiente resultado del libro:

$$d_{ij}^2 = (x_j - \bar{x} - \hat{y}_{j1}\hat{e}_1 - \hat{y}_{j2}\hat{e}_2)'(x_j - \bar{x} - \hat{y}_{j1}\hat{e}_1 - \hat{y}_{j2}\hat{e}_2)$$

Notemos que, al insertar la identidad  $\hat{E}\hat{E}' = I$ , también obtenemos:

$$d_{ij}^2 = (x_j - \bar{x} - \hat{y}_{j1}\hat{e}_1 - \hat{y}_{j2}\hat{e}_2)'\hat{E}\hat{E}'(x_j - \bar{x} - \hat{y}_{j1}\hat{e}_1 - \hat{y}_{j2}\hat{e}_2) = \sum_{k=3}^p \hat{y}_{jk}^2$$

Esto es, haremos la carta de control usando la distancia entre los datos centrados y sus proyecciones sobre las dos primeras componentes principales y graficamos. Para definir el límite superior de la carta de control se hallan:

$$\overline{d_U^2} = \frac{1}{n} \sum_{j=1}^n d_{Uj}^2 = c\nu \quad , \quad s_{d^2}^2 = \frac{1}{n-1} \sum_{j=1}^n (d_{Uj}^2 - \overline{d_U^2})^2 = 2c^2\nu$$

$$c = \frac{s_{d^2}^2}{2\overline{d_U^2}} \quad , \quad \nu = \frac{2(\overline{d_U^2})^2}{s_{d^2}^2}.$$

El límite de control está dado por:

$$c\chi_\nu^2(\alpha),$$

En este caso usaremos una significancia del 5%.

```
d2 <- rowSums(principal_components[, 4:6]^2)

c=var(d2)/(2*mean(d2))

v=2*(mean(d2)^2/var(d2))

colors<-rep("black",nrow(police_hours))
colors[which(d2>=c*qchisq(.95,v))]="red"

plot(1:nrow(police_hours), d2, type = "b", main = "Suma de cuadrados de \n los componentes no explicados",
     xlab = "Periodo", ylab = "SS", pch=19, col=colors)

abline(h=c*qchisq(.95,v), col="red")
text(15, y =c*qchisq(.95,v), labels="UCL", col="red", pos=3)
```

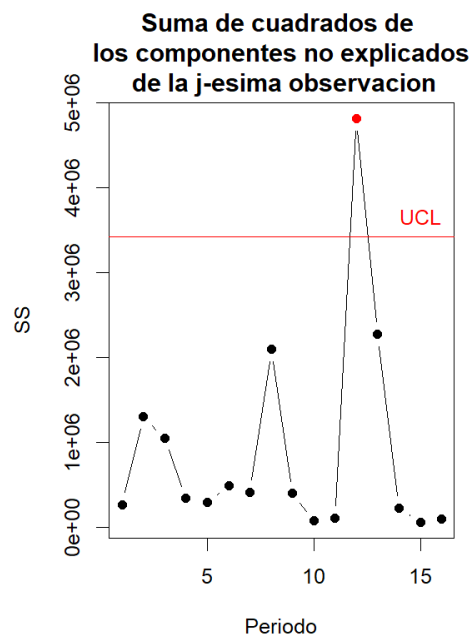


Figure 2: Carta de control con los Componentes principales restantes.

La carta de control nuevamente nos arroja una señal en el periodo 12. Se concluye lo mismo que se concluyo usando únicamente los dos primeros componentes principales, el proceso se salió de control para  $t = 12$ .

### III. LECTURAS

Lea las siguientes dos referencias

- Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2), 265-286.

- Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005), “ROBPCA: A New Approach to Robust Principal Component Analysis,” *Technometrics*, 47, 64–79.

De acuerdo con estas lecturas:

## Deficiencias del PCA Clásico

El análisis de componentes principales (PCA) clásico presenta varias deficiencias importantes:

- **Sensibilidad a valores atípicos:** El PCA clásico se basa en la matriz de covarianza, que es altamente sensible a observaciones atípicas. Esto puede sesgar las componentes principales hacia los valores atípicos, reduciendo su capacidad para representar correctamente la estructura de los datos
- **Falta de interpretabilidad:** Cada componente principal es una combinación lineal de todas las variables originales, lo que dificulta la interpretación de los resultados
- **Limitaciones en altas dimensiones:** En conjuntos de datos donde el número de variables es mayor que el número de observaciones ( $p > n$ ), la estimación de la matriz de covarianza se vuelve inestable y puede llevar a resultados poco fiables.
- **No considera la esparsidad:** En muchos problemas, como en el análisis de datos genómicos, es deseable obtener componentes principales con cargas dispersas para facilitar la interpretación, lo que el PCA clásico no logra.
- **Asunción de normalidad:** El PCA clásico asume que los datos siguen una distribución normal, lo cual no siempre es cierto en la práctica. Métodos robustos son necesarios para manejar datos que se desvían de esta suposición.

## Deficiencias Resueltas y Propuestas de los Artículos

### ROBPCA: Robust Principal Component Analysis

El método ROBPCA aborda las deficiencias del PCA clásico relacionadas con la sensibilidad a valores atípicos y la inestabilidad en alta dimensión. Su propuesta se basa en:

- **Combinación de estimación robusta de covarianza y proyección robusta:** Utiliza una combinación de estimación de matriz de covarianza robusta y técnicas de *projection pursuit* para reducir la influencia de valores atípicos.
- **Preprocesamiento de datos:** Reduce la dimensionalidad inicial usando una proyección robusta para evitar la inestabilidad en problemas con  $p > n$ .
- **Detección de valores atípicos:** Introduce un *diagnostic plot* que clasifica los datos en observaciones regulares, valores atípicos ortogonales y valores atípicos de leverage.

Este método permite obtener componentes principales más robustos que no se ven influenciados por valores extremos y mantiene una buena eficiencia computacional en datos de alta dimensión.

### SPCA: Sparse Principal Component Analysis

El método SPCA soluciona la falta de interpretabilidad del PCA clásico y su incapacidad para generar componentes principales esparsos. Su propuesta consiste en:

- **Formulación de PCA como un problema de regresión:** Reformula el cálculo de componentes principales como un problema de optimización basado en regresión con penalización.
- **Uso de penalización Lasso y Elastic Net:** Aplica la penalización Lasso o Elastic Net sobre las cargas de los componentes principales, lo que permite obtener coeficientes esparsos.
- **Reducción de dimensionalidad efectiva:** Al forzar a muchas cargas a ser cero, selecciona subconjuntos de variables relevantes, facilitando la interpretación de los componentes principales.

SPCA permite conservar la varianza explicada mientras genera componentes más interpretables y adecuados para aplicaciones en datos de alta dimensión, como el análisis genético.

## Comparación de ROBPCA y SPCA

Método	Deficiencia Resuelta	Enfoque
ROBPCA	Sensibilidad a valores atípicos Inestabilidad en alta dimensión	Estimación robusta de covarianza y proyección robusta Preprocesamiento de reducción de dimensión
SPCA	Falta de interpretabilidad No considera la esparsidad	Penalización Lasso/Elastic Net para obtener esparsidad Reformulación de PCA como problema de regresión penalizada

Table 10: Comparación entre ROBPCA y SPCA

Ambos métodos mejoran el PCA clásico, pero con enfoques distintos: ROBPCA prioriza la robustez frente a valores atípicos, mientras que SPCA mejora la interpretabilidad al inducir esparsidad en los componentes principales.

## 2. Parte de Análisis Factorial

### 9.1

Muestre que la matriz de covarianzas

$$\rho = \begin{bmatrix} 1.0 & 0.63 & 0.45 \\ 0.63 & 1.0 & 0.35 \\ 0.45 & 0.35 & 1.0 \end{bmatrix}$$

para las  $p = 3$  variables aleatorias estandarizadas  $Z_1, Z_2$ , y  $Z_3$  puede ser generada por el modelo de un solo factor ( $m = 1$ ):

$$Z_1 = 0.9F_1 + \epsilon_1$$

$$Z_2 = 0.7F_1 + \epsilon_2$$

$$Z_3 = 0.5F_1 + \epsilon_3$$

donde  $\text{Var}(F_1) = 1$ ,  $\text{Cov}(\epsilon, F_1) = 0$ , y

$$\Psi = \text{Cov}(\epsilon) = \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}$$

Esto es, escribir  $\rho$  en la forma

$$\rho = \mathbf{L}\mathbf{L}' + \Psi$$

**Solución:**

Procedemos a computar las matrices para comprobar la igualdad:

```
rho <- matrix(c(1.0, 0.63, 0.45,
               0.63, 1.0, 0.35,
               0.45, 0.35, 1.0),
             nrow = 3, ncol = 3, byrow = TRUE)
```

Se tienen:

```
L <- cbind(c(.9, .7, .5))
psi <- diag(c(.19, .51, .75))
```

Vea que

$$\mathbf{LL}^t + \Psi$$

es igual a

```
L%*%t(L) + psi

##      [,1] [,2] [,3]
## [1,] 1.00 0.63 0.45
## [2,] 0.63 1.00 0.35
## [3,] 0.45 0.35 1.00
all.equal(L%*%t(L) + psi, rho)

## [1] TRUE
```

Con esto concluimos que se cumple la igualdad.

## 9.2.

Use la información del ejercicio 9.1.

(a)

Calcule las comunalidades  $h_i^2$ ,  $i = 1, 2, 3$ , e interprete estos valores.

**Solución:**

Hallamos:  $h_i^2 = \ell_{i1}^2$  para  $i = 1, 2, 3$

```
h_i <- L^2
knitr::kable(h_i, col.names = "Comunalidades")
```

Comunalidades
0.81
0.49
0.25

Estos resultados nos dicen que el factor común explica un 81% de la varianza de la primera variable, 49% de la varianza de la segunda variable y finalmente el 25% de la varianza de la tercera variable.

(b)

Calcule  $\text{Corr}(Z_i, F_1)$  para  $i = 1, 2, 3$ . ¿Cuál variable podría tener mayor peso en la “denominación” del factor común? ¿Por qué?

**Solución:**

Para variables ya estandarizadas se cumple que,  $\text{Corr}(Z_i, F_1) = \text{Cov}(Z_i, F_1)$  para  $i = 1, 2, 3$ .

En ese orden, tenemos que:

$$\text{Cov}(Z_1, F_1) = \ell_{11} = 0.9$$

$$\text{Cov}(Z_2, F_1) = \ell_{12} = 0.7$$

$$\text{Cov}(Z_3, F_1) = \ell_{13} = 0.5$$

Al tener  $Z_1$  la correlación más alta con el factor común podemos decir que es la variable con un mayor peso y por tanto la principal candidata para determinar una denominación del factor común.

### 9.3

Los autovalores y autovectores de la matriz de correlación  $\rho$  del ejercicio 9.1 son:

$$\lambda_1 = 1.96, \quad \mathbf{e}_1 = [.625, .593, .507]'$$

$$\lambda_2 = 0.68, \quad \mathbf{e}_2 = [-0.219, -0.491, 0.843]'$$

$$\lambda_3 = 0.36, \quad \mathbf{e}_3 = [0.749, -0.638, -0.177]'$$

(a)

Suponiendo un modelo de factor con  $m = 1$ , calcule la matriz de cargas  $\mathbf{L}$  y la matriz de varianzas específicas  $\mathbf{\Psi}$  utilizando el método de solución por componentes principales. Compare los resultados con los del Ejercicio 9.1.

**Solución:**

Realizamos el calculo de la matriz  $L$ :

```
L <- cbind(sqrt(lambdas[1]) * vectors[,1])
```

```
L_L_t <- L %*% t(L)
```

```
psi <- matrix(0, 3, 3)
```

```
for (i in 1:3) {  
  psi[i,i] <- rho[i,i] - L[i]^2  
}
```

```
psi <- as.data.frame(psi)
```

```
knitr::kable(L, col.names = c("L1"))
```

$$\frac{\overline{\text{L1}}}{0.8750}$$

L1
0.8302
0.7098

```
knitr::kable(psi, col.names = c("SP1", "SP2", "SP3"))
```

SP1	SP2	SP3
0.234375	0.000000	0.000000
0.000000	0.310768	0.000000
0.000000	0.000000	0.496184

Los resultados son claramente diferentes en especial en las varianzas específicas, para la primera variable aumenta un poco y disminuye en una magnitud importante para las dos restantes.

(b)

¿Qué proporción de la varianza total de la población es explicada por el primer factor común?

**Solución:**

Para hallar la proporción de varianza explicada por el factor común computamos:

```
lambdas[1]/sum(diag(rho))
```

```
## [1] 0.6533333
```

El resultado nos dice que el factor común explica aproximadamente un 65.3% de la varianza total.

## 9.9

En un estudio sobre la preferencia de licores en Francia, Stoetzel [?] recopiló clasificaciones de preferencia de  $p = 9$  tipos de licores de una muestra de  $n = 1442$  individuos. Un análisis factorial de la matriz de correlación muestral  $9 \times 9$  de los rangos de ordenamiento proporcionó las cargas factoriales estimadas.

Variable	F1	F2	F3
Liquors	0.64	0.02	0.16
Kirsch	0.50	-0.06	-0.10
Mirabelle	0.46	-0.24	-0.19
Rum	0.17	0.74	0.97
Marc	-0.29	0.66	-0.39
Whiskey	-0.29	-0.08	0.09
Calvados	-0.49	-0.20	-0.04
Cognac	-0.52	-0.03	0.42
Armagnac	-0.60	-0.17	0.14

Dado estos resultados, Stoetzel concluyó lo siguiente:

El principio principal de la preferencia por los licores en Francia es la distinción entre licores dulces y fuertes. El segundo factor motivador es el precio, lo que se puede entender al recordar que el licor es tanto un bien costoso como un artículo de consumo ostentoso.

Excepto en el caso de los dos licores más populares y menos costosos (rum y marc), este segundo factor juega un papel mucho menor en la producción de juicios de preferencia.

El tercer factor tiene un componente sociológico y se relaciona principalmente con la variabilidad regional de los juicios. (Ver [?], p. 11).

(a)

Dado lo que sabe sobre los distintos tipos de licores, ¿parece razonable la interpretación de Stoetzel?

Analizando los factores tenemos que:

- **Primer Factor (Dulce vs. Fuerte)**
  - Distingue bien entre licores dulces (Mirabelle, Kirsch) y licores fuertes (Whiskey, Cognac, Armagnac). Esto es coherente con la primera interpretación de Stoetzel sobre la distinción entre licores dulces y fuertes.
- **Segundo Factor (Precio y Estatus Social)**
  - Está fuertemente relacionado con el precio y el prestigio de los licores. Cognac y Armagnac son conocidos por ser caros y símbolo de estatus. En contraste, Rum y Marc son más baratos, pero siguen siendo populares. Respecot a lo que Stoetzel menciona, que el precio no siempre es determinante, es razonable considerarlo porque existen otros factores que influyen en la preferencia.
- **Tercer Factor (Variabilidad Regional y Sociológica)**

Algunos licores tienen fuertes asociaciones regionales, como el Calvados en Normandía o el Cognac en la región de Cognac. Esto respalda la idea de que las preferencias de licor están influenciadas por factores culturales y geográficos.

(b)

Graficar los pares de cargas factoriales para los dos primeros factores.

Realizar una rotación ortogonal de los ejes de los factores. Generar cargas factoriales rotadas aproximadas. Interpretar las cargas factoriales rotadas de los dos primeros factores. ¿Su interpretación concuerda con la interpretación de Stoetzel de estos factores a partir de las cargas no rotadas? Explique.

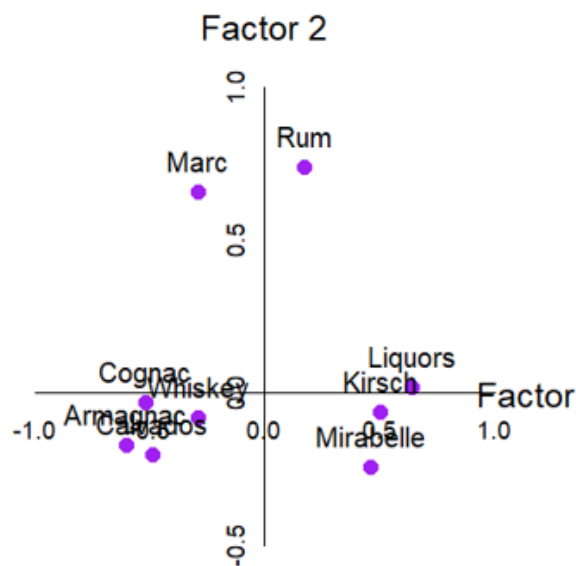


Figure 3: Gráfico de las cargas factoriales en los primeros dos factores.

Ahora realizamos la rotación usando Varimax:

```
library(GPArotation)

factor_matrix <- as.matrix(factor_loadings[, 2:4])
```



```

rotated_factors <- varimax(factor_matrix)
rotated_loadings <- as.matrix(rotated_factors$loadings)

rotated_df <- data.frame(Variable = factor_loadings$Variable, rotated_loadings)
rotated_df<-rotated_df[1:9,]

Factors<-rotated_df$rotated_loadings

rotated_df<-data.frame(Variable=factor_loadings$Variable,Factors)

knitr::kable(rotated_df)

```

Variable	F1	F2	F3
Liquors	0.5613772	-0.2775749	0.2083456
Kirsch	0.4826245	-0.1629616	-0.0641646
Mirabelle	0.4298834	-0.2472650	-0.2436397
Rum	0.0603274	0.0405831	1.2296803
Marc	-0.0220187	0.8190013	0.0234945
Whiskey	-0.3137325	-0.0130337	-0.0014186
Calvados	-0.4925576	0.0237677	-0.1962704
Cognac	-0.6073666	-0.0642172	0.2732800
Armagnac	-0.6376320	-0.0071093	-0.0432997

Después de aplicar la rotación Varimax, analizamos las nuevas cargas factoriales y las comparamos con la interpretación inicial de Stoetzel sobre las preferencias de licores en Francia.

**Factor 1 (F1):** Diferenciación entre licores dulces y fuertes. - Licores dulces (Kirsch, Mirabelle, Liquors) con cargas positivas. - Licores fuertes (Cognac, Armagnac, Calvados) con cargas negativas.

**Factor 2 (F2):** Variabilidad regional - Marc y Calvados tienen altas cargas positivas. - Whiskey y Cognac tienen cargas cercanas a 0.

Por tanto, se mantiene la interpretación original de Stoetzel sobre la distinción para los precios en el primer factor y confirma la teoría sobre la importancia de la regionalidad para el segundo factor.

## 9.10

La matriz de correlación para las mediciones de huesos de pollo (ver Ejemplo 9.14) es:

$$\begin{bmatrix}
 1.000 & 0.505 & 0.569 & 0.602 & 0.621 & 0.603 \\
 0.505 & 1.000 & 0.422 & 0.467 & 0.482 & 0.450 \\
 0.569 & 0.422 & 1.000 & 0.926 & 0.877 & 0.878 \\
 0.602 & 0.467 & 0.926 & 1.000 & 0.874 & 0.894 \\
 0.621 & 0.482 & 0.877 & 0.874 & 1.000 & 0.937 \\
 0.603 & 0.450 & 0.878 & 0.894 & 0.937 & 1.000
 \end{bmatrix}$$

Las siguientes cargas factoriales estimadas fueron extraídas mediante el procedimiento de máxima verosimilitud:

Variable	$F_1$	$F_2$	$F_1^*$	$F_2^*$
1. Skull length	0.602	0.200	0.484	0.411
2. Skull breadth	0.467	0.154	0.375	0.319
3. Femur length	0.926	0.143	0.603	0.717
4. Tibia length	1.000	0.000	0.519	0.855

Variable	$F_1$	$F_2$	$F_1^*$	$F_2^*$
5. Humerus length	0.874	0.476	0.861	0.499
6. Ulna length	0.894	0.327	0.744	0.594

Usando las cargas factoriales estimadas **sin rotar**, obtenga las estimaciones de máxima verosimilitud de lo siguiente:

(a)

Las varianzas específicas

**Solución:**

Computamos el calculo de las varianzas específicas:

```
Var_Esp <- diag(R - L %*% t(L))
knitr::kable(Var_Esp,col.names = "Var. Específicas")
```

Var. Específicas
0.597596
0.758195
0.122075
0.000000
0.009548
0.093835

(b)

Las comunalidades.

**Solución:**

Obtenemos el valor de las comunalidades aplicando  $1 - \hat{\psi}_i$

```
knitr::kable(cbind(1 - Var_Esp),col.names = "Comunalidades")
```

Comunalidades
0.402404
0.241805
0.877925
1.000000
0.990452
0.906165

(c)

La proporción de varianza explicada para cada factor.

**Solución:**

Usando que al sumar por columnas en la matriz  $\hat{\mathbf{L}}$  obtenemos los valores propios tenemos:

```
p <- 6
lambda_1 <- sum(L[,1]^2)
```

```
lambda_2 <- sum(L[,2]^2)
```

Así:

```
lambda_1/p
```

```
## [1] 0.6668468
```

```
lambda_2/p
```

```
## [1] 0.06961167
```

Los factores 1 y 2 explican el aproximadamente el 66.7% y 6.96% de la varianza total respectivamente.

(d)

La matriz de residuales.

**Solución:**

La matriz de residuales se puede obtener operando las matrices  $\hat{L}$  y  $\hat{\psi}$  de la siguiente forma:

$$\mathbf{R} = \hat{\mathbf{L}}_z \hat{\mathbf{L}}_z' - \hat{\Psi}_z.$$

```
knitr::kable(R - L %*% t(L) - psi)
```

0.000000	0.193066	-0.017052	0	-0.000348	-0.000588
0.193066	0.000000	-0.032464	0	0.000538	-0.017856
-0.017052	-0.032464	0.000000	0	-0.000392	0.003395
0.000000	0.000000	0.000000	0	0.000000	0.000000
-0.000348	0.000538	-0.000392	0	0.000000	-0.000008
-0.000588	-0.017856	0.003395	0	-0.000008	0.000000

## 9.11

Refiérase al Ejercicio 9.10. Calcule el valor del criterio varimax utilizando tanto las cargas factoriales estimadas sin rotar como las rotadas. Comente sobre los resultados.

**Solución:**

Usando (9-45), tenemos que:

$$\nu = \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p \tilde{\ell}_{ij}^4 - \left( \frac{\sum_{i=1}^p \tilde{\ell}_{ij}^2}{p} \right)^2 \right]$$

Donde

$$\tilde{\ell}_{ij}^* = \frac{\hat{\ell}_{ij}}{\hat{h}_i}$$

Con base en esa formula se computan los calculos.

1. Para los factores de carga originales.

```

p<-length(communalities)
a <- factor_loadings$F1_Estimated / sqrt(communalities)
b <- factor_loadings$F2_Estimated / sqrt(communalities)
c<-(sum(a^2))^2/p
d<-(sum(b^2))^2/p

V=(1/p)*(sum(a^4)-c+sum(b^4)-d)

round(V,4)

## [1] 0.0109

```

2. Para los factores rotados:

```

a <- factor_loadings$F1_Rotated / sqrt(communalities)
b <- factor_loadings$F2_Rotated / sqrt(communalities)
c<-(sum(a^2))^2/p
d<-(sum(b^2))^2/p

V=(1/p)*(sum(a^4)-c+sum(b^4)-d)

round(V,4)

## [1] 0.0469

```

En este caso, las cargas no rotadas parecen ofrecer una mejor claridad conceptual. A pesar de que la varianza explicada aumenta con la rotación, los factores resultantes pueden ser menos intuitivos, lo que sugiere que la estructura sin rotar refleja mejor la naturaleza de los datos.

## 9.12

La matriz de covarianza para los logaritmos de las mediciones de tortugas (ver Ejemplo 8.4) es:

$$\mathbf{S} = 10^{-3} \begin{bmatrix} 11.072 & 8.019 & 6.417 \\ 8.160 & 6.005 & 6.773 \end{bmatrix}$$

Se obtuvieron las siguientes estimaciones de máxima verosimilitud de las cargas factoriales para un modelo con  $m = 1$ :

Variable	Estimated Factor Loadings $F_1$
1. ln(length)	0.1022
2. ln(width)	0.0752
3. ln(height)	0.0765

Usando las cargas factoriales estimadas, obtenga las estimaciones de máxima verosimilitud de cada uno de los siguientes elementos:

(a)

Varianzas específicas.

**Solución:**

Se calcula  $\mathbf{S}$  usando  $\mathbf{S}_n$ :

```

n <- 24

S <- 10^(-3) * matrix(c(
  11.072, 8.019, 8.160,
  8.019, 6.417, 6.005,
  8.160, 6.005, 6.773
), nrow = 3, byrow = TRUE)

S_n <- (n-1)/n * S

L <- matrix(c(0.1022, 0.0752, 0.0765), nrow = 3, ncol = 1, byrow = FALSE)

```

A partir de  $\mathbf{S}_n$  y  $\hat{\mathbf{L}}$  hallamos las varianzas específicas:

```

psi <- diag(S_n - L%*%t(L))
psi<-diag(psi)

Var_Esp <- diag(psi)

knitr::kable(Var_Esp,col.names = "Varianzas Específicas")

```

Varianzas Específicas
0.0001658
0.0004946
0.0006385

(b)

Comunalidades.

**Solución:**

Dado que en este caso sólo estamos considerando un factor común, el calculo se reduce a  $\hat{h}_i = \hat{l}_{i1}^2$ ,

```

Comunalidades <- L^2

knitr::kable(Comunalidades, col.names = "Comunalidades")

```

Comunalidades
0.0104448
0.0056550
0.0058522

(c)

Proporción de varianza explicada por el factor.

**Solución:**

Para calcular el valor de la proporción de varianza explicada por  $F_1$  basta con sumar los  $\hat{h}_{i1}^2$  y lo dividirlos sobre la traza de  $\mathbf{S}_n$

```

p <- 3

sum(L^2)/sum(diag(S_n))

```

```
## [1] 0.9441336
```

Concluimos que el factor común que estamos considerando explica el 94% de la varianza total.

(d)

La matriz de residuales  $S_n - \hat{L}\hat{L}' - \hat{\Psi}$

*Pista:* Convertir  $\mathbf{S}$  a  $\mathbf{S}_n$ .

```
knitr::kable(S_n - L %*% t(L) - psi)
```

0.0e+00	-6e-07	1.7e-06
-6.0e-07	0e+00	2.0e-06
1.7e-06	2e-06	0.0e+00

## 9.15.

Hirschey y Wichern [?] investigan la consistencia, los determinantes y los usos de las medidas contables y de valor de mercado de la rentabilidad. Como parte de su estudio, se realizó un análisis factorial de las medidas de beneficios contables y las estimaciones de beneficios económicos según el mercado. La matriz de correlación de las medidas de rentabilidad contables históricas, contables de reemplazo y de valor de mercado para una muestra de empresas operativas en 1977 es la siguiente:

$$\mathbf{R} = \begin{bmatrix} 1.000 & 0.738 & 0.731 & 0.828 & 0.681 & 0.712 & 0.625 & 0.604 \\ 0.738 & 1.000 & 0.520 & 0.688 & 0.831 & 0.543 & 0.322 & 0.303 \\ 0.731 & 0.520 & 1.000 & 0.652 & 0.513 & 0.826 & 0.579 & 0.617 \\ 0.828 & 0.688 & 0.652 & 1.000 & 0.887 & 0.867 & 0.639 & 0.563 \\ 0.681 & 0.831 & 0.513 & 0.887 & 1.000 & 0.692 & 0.419 & 0.352 \\ 0.712 & 0.543 & 0.826 & 0.867 & 0.692 & 1.000 & 0.608 & 0.610 \\ 0.625 & 0.322 & 0.579 & 0.639 & 0.419 & 0.608 & 1.000 & 0.937 \\ 0.604 & 0.303 & 0.617 & 0.563 & 0.352 & 0.610 & 0.937 & 1.000 \end{bmatrix}$$

Se obtuvieron las siguientes estimaciones de las cargas factoriales mediante componentes principales rotados para un modelo con  $m = 3$  factores:

Variable	$F_1$	$F_2$	$F_3$
Historical return on assets	0.433	0.612	0.499
Historical return on equity	0.125	0.892	0.234
Historical return on sales	0.296	0.238	0.887
Replacement return on assets	0.406	0.708	0.483
Replacement return on equity	0.198	0.895	0.283
Replacement return on sales	0.331	0.414	0.789
Market Q ratio	0.928	0.160	0.294
Market relative excess value	0.910	0.079	0.355

Proporción de varianza total acumulada:

$F_1$	$F_2$	$F_3$
0.287	0.628	0.908

(a)

Usando las cargas factoriales estimadas, determine las varianzas específicas y las comunialidades. **Solución:**

Usando los resultados del libro calculamos:

```
psi <- diag(R - L%*%t(L))
psi<-diag(psi)
```

Por consiguiente, los valores para las varianzas específicas y comunialidades son:

```
Var_Esp <- diag(psi)
Comunalidades <- NULL

for (i in 1:nrow(L)) {
  Comunalidades[i] <- sum(L[i,]^2)
}

knitr::kable(cbind( Var_Esp,Comunalidades),col.names =c("Varianzas Específicas","Comunalidades"))
```

Varianzas Específicas	Comunalidades
0.188966	0.811034
0.133955	0.866045
0.068971	0.931029
0.100611	0.899389
0.079682	0.920318
0.096522	0.903478
0.026780	0.973220
0.039634	0.960366

(b)

Determine la matriz de residuales

$$\mathbf{R} - \hat{\mathbf{L}}_z \hat{\mathbf{L}}_z' - \hat{\Psi}_z.$$

Dada esta información y la proporción acumulada de la varianza total explicada en la tabla anterior, ¿parece adecuado un modelo de  $m = 3$  factores para estos datos?

```
knitr::kable(R-L %*% t(L)-psi )
```

0.000000	0.021205	0.014563	-0.022111	-0.093691	-0.078402	-0.021450	-0.015523
0.021205	0.000000	0.063146	-0.107308	-0.058312	-0.052289	-0.005516	0.035712
0.014563	0.063146	0.000000	-0.065101	-0.009639	-0.070351	0.005454	0.013953
-0.022111	-0.107308	-0.065101	0.000000	0.036263	0.058415	0.006950	-0.033857
-0.093691	-0.058312	-0.009639	0.036263	0.000000	0.032645	0.008854	0.000650
-0.078402	-0.052289	-0.070351	0.058415	0.032645	0.000000	0.002626	-0.004011
-0.021450	-0.005516	0.005454	0.006950	0.008854	0.002626	0.000000	-0.024490
-0.015523	0.035712	0.013953	-0.033857	0.000650	-0.004011	-0.024490	0.000000

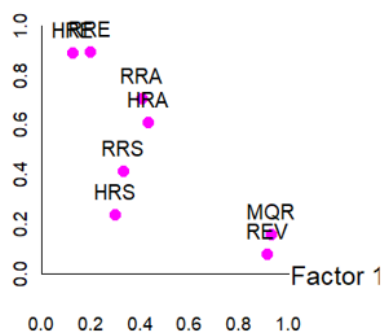
Los valores de los residuales son pequeños, adicionalmente, los valores de las comunialidades son altos lo cual refleja una muy buena descripción de la varianza total por parte del modelo. También la proporción de varianza total explicada por los factores comunes es alta, así que parece que el modelo con  $m = 3$  es adecuado.

(c)

Suponiendo que las cargas factoriales estimadas menores a 0.4 son pequeñas, interprete los tres factores. ¿Parece, por ejemplo, que las medidas de valor de mercado proporcionan evidencia de rentabilidad distinta de la ofrecida por las medidas contables? ¿Es posible separar las medidas contables históricas de rentabilidad de las medidas contables de reemplazo?

**Solución:**

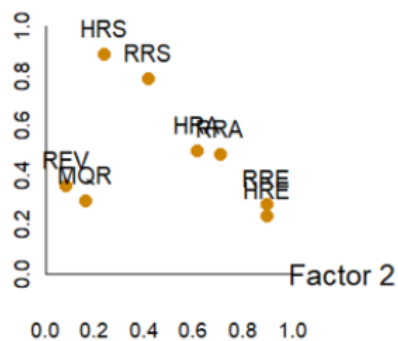
Factor 2



Factor 3



Factor 3



El primer factor está relacionado con las medidas de valor de mercado (Q, REV). El segundo factor se asocia con medidas contables históricas sobre el patrimonio (HRE, RRE). El tercer factor está vinculado con



medidas contables históricas sobre las ventas (HRS, RRS). Las medidas contables históricas sobre activos (HRA, RRA) tienen una relación débil con todos los factores.

Las medidas de valor de mercado reflejan la rentabilidad de manera diferente a las medidas contables. Sin embargo, no se pueden separar claramente las medidas contables históricas de rentabilidad de las medidas contables de reemplazo

### 3. Regresion Multivariada

#### 7.3

Sea

$$\mathbf{Y}_{(n \times 1)} = \mathbf{Z}_{(n \times (r+1))} \beta_{(r+1) \times 1} + \boldsymbol{\varepsilon}_{(n \times 1)}$$

donde  $E(\boldsymbol{\varepsilon}) = 0$  pero  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2 \mathbf{V}$ , con  $\mathbf{V}$  de dimensión  $(n \times n)$  conocida y definida positiva. Para  $\mathbf{V}$  de rango completo, demostrar que el **estimador de mínimos cuadrados ponderados** es

$$\hat{\beta}_W = (\mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Y}$$

Si  $\sigma^2$  es desconocido, puede estimarse de manera insesgada como:

$$(n - r - 1)^{-1} \times (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W).$$

$\mathbf{V}^{-1/2} \mathbf{Y} = (\mathbf{V}^{-1/2} \mathbf{Z}) \beta + \mathbf{V}^{-1/2} \boldsymbol{\varepsilon}$  es de la forma clásica de regresión lineal:

$$\mathbf{Y}^* = \mathbf{Z}^* \beta + \boldsymbol{\varepsilon}^*,$$

con  $E(\boldsymbol{\varepsilon}^*) = 0$  y  $E(\boldsymbol{\varepsilon}^* \boldsymbol{\varepsilon}^{*T}) = \sigma^2 \mathbf{I}$ . Por lo tanto,

$$\hat{\beta}_W = \hat{\beta}^* = (\mathbf{Z}^{*T} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*T} \mathbf{Y}^*.$$

**Solución:**

El modelo es:

$$\mathbf{Y} = \mathbf{Z} \beta + \boldsymbol{\varepsilon},$$

donde  $E(\boldsymbol{\varepsilon}) = 0$  y  $E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) = \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{V}$ .  $\mathbf{V}_{n \times n}$  es una matriz conocida, definida positiva y de rango completo.

Para encontrar los mínimos cuadrados ponderados, queremos transformar los datos observados de tal manera que los errores transformados sean incorrelacionados e idénticamente distribuidos, es decir, si  $\boldsymbol{\varepsilon}^*$  es el vector de errores transformados, necesitamos que:

$$\text{Var}(\boldsymbol{\varepsilon}^*) = \sigma^2 \mathbf{I}.$$

Observemos que, dado que  $\mathbf{V}$  es una matriz definida positiva y de rango completo, existe la raíz de Cholesky  $\mathbf{V}^{1/2}$  de  $\mathbf{V}$ .  $\mathbf{V}^{1/2}$  es simétrica e invertible. Además,  $\mathbf{V}^{-1/2}$  es simétrica, ya que:

$$(\mathbf{V}^{-1/2})^T = [(\mathbf{V}^{1/2})^{-1}]^T = [(\mathbf{V}^{1/2})^T]^{-1} = (\mathbf{V}^{1/2})^{-1} = \mathbf{V}^{-1/2}.$$

Premultiplicamos la ecuación del modelo por  $\mathbf{V}^{-1/2}$ :

$$\mathbf{V}^{-1/2}\mathbf{Y} = \mathbf{V}^{-1/2}\mathbf{Z}\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\epsilon}$$

$$\Rightarrow \mathbf{Y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*,$$

donde  $\mathbf{Y}^* = \mathbf{V}^{-1/2}\mathbf{Y}$ ,  $\mathbf{Z}^* = \mathbf{V}^{-1/2}\mathbf{Z}$  y  $\boldsymbol{\epsilon}^* = \mathbf{V}^{-1/2}\boldsymbol{\epsilon}$ . Entonces,

$$\text{Var}(\boldsymbol{\epsilon}^*) = \text{Var}(\mathbf{V}^{-1/2}\boldsymbol{\epsilon}).$$

$$= \mathbf{V}^{-1/2}\text{Var}(\boldsymbol{\epsilon})(\mathbf{V}^{-1/2})^T.$$

$$= \mathbf{V}^{-1/2}(\sigma^2\mathbf{V})\mathbf{V}^{-1/2}.$$

$$= \sigma^2\mathbf{I}.$$

Por lo tanto, los errores transformados son incorrelacionados e idénticamente distribuidos. Así, podemos usar  $\mathbf{Y}^*$  como vector de respuesta y  $\mathbf{Z}^*$  como matriz de diseño para encontrar el estimador de mínimos cuadrados ponderados  $\hat{\boldsymbol{\beta}}_W$ .

Según el resultado 7.1

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{Z}^{*T}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*T}\mathbf{Y}^*.$$

$$= \left[ (\mathbf{V}^{-1/2}\mathbf{Z})^T \mathbf{V}^{-1/2}\mathbf{Z} \right]^{-1} (\mathbf{V}^{-1/2}\mathbf{Z})^T \mathbf{V}^{-1/2}\mathbf{Y}.$$

$$= [\mathbf{Z}^T \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{Z}]^{-1} \mathbf{Z}^T \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{Y}.$$

$$= [\mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Z}]^{-1} \mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Y}.$$

Esto demuestra la primera parte.

Ahora bien, en el modelo  $\mathbf{Y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$ , donde  $\text{Var}(\boldsymbol{\epsilon}^*) = \sigma^2\mathbf{I}$ , el parámetro  $\sigma^2$  puede ser estimado de manera insesgada por:

$$s^2 = \frac{\boldsymbol{\epsilon}^{*T}\boldsymbol{\epsilon}^*}{n - (r + 1)}.$$

donde  $\boldsymbol{\epsilon}^*$  es el vector de residuos:

$$\boldsymbol{\epsilon}^* = \mathbf{Y}^* - \hat{\mathbf{Y}}^*.$$

Sustituyendo  $\hat{\mathbf{Y}}^* = \mathbf{Z}^*\hat{\boldsymbol{\beta}}_W$ :

$$\boldsymbol{\epsilon}^* = \mathbf{Y}^* - \mathbf{Z}^*\hat{\boldsymbol{\beta}}_W.$$

$$= \mathbf{V}^{-1/2} \mathbf{Y} - \mathbf{V}^{-1/2} \mathbf{Z} \hat{\beta}_W.$$

$$= \mathbf{V}^{-1/2} (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W).$$

Sustituyendo en la expresión de  $s^2$ :

$$\begin{aligned} s^2 &= \frac{[\mathbf{V}^{-1/2} (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)]^T [\mathbf{V}^{-1/2} (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)]}{n - (r + 1)}. \\ &= \frac{(\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)^T \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)}{n - (r + 1)}. \\ &= \frac{(\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z} \hat{\beta}_W)}{n - (r + 1)}. \end{aligned}$$

Esto completa la demostración.

## 7.4

Utilice el estimador de mínimos cuadrados ponderados en el Ejercicio 7.3 para derivar una expresión para la estimación de la pendiente  $\beta$  en el modelo  $Y_j = \beta Z_j + \varepsilon_j$ , para  $j = 1, 2, \dots, n$ , cuando:

1.  $\text{Var}(\varepsilon_j) = \sigma^2$ ,
2.  $\text{Var}(\varepsilon_j) = \sigma^2 Z_j$ ,
3.  $\text{Var}(\varepsilon_j) = \sigma^2 + \sigma^2 Z_j^2$ .

Comente sobre la forma en que las varianzas desiguales de los errores influyen en la elección óptima de  $\hat{\beta}_W$ .

### Solución:

En todos los casos, el modelo es:

$$Y_j = \beta z_j + \varepsilon_j, \quad j = 1, 2, \dots, n.$$

Las varianzas  $\text{Var}(\varepsilon_j)$  son diferentes en los tres casos. En consecuencia, la matriz de varianza-covarianza  $\sigma^2 V$  de  $\varepsilon$  es diferente en cada uno de los casos.

En cada caso, el vector de respuesta observado  $\mathbf{Y}$  y la matriz de diseño  $\mathbf{Z}$  son, respectivamente:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$

(a)

$$\text{Var}(\varepsilon_j) = \sigma^2$$

Entonces, la matriz de covarianza de varianza  $\sigma^2 V$  de  $\varepsilon$  es  $\sigma^2 I$ . Por lo tanto,  $V^{-1} = I$ .

El estimador de mínimos cuadrados ponderados se obtiene como:

$$\begin{aligned}\hat{\beta}_W &= (\mathbf{Z}^T V^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T V^{-1} \mathbf{Y} \\ &= \left( \sum_{i=1}^n z_i^2 \right)^{-1} \sum_{i=1}^n z_i Y_i.\end{aligned}$$

Así, la estimación de la pendiente del modelo es:

$$\hat{\beta}_W = \frac{\sum_{i=1}^n z_i Y_i}{\sum_{i=1}^n z_i^2}.$$

Aquí, los valores de  $Y_i$  asociados a valores altos de  $z_i$  tienen mayor influencia en la estimación.

(b)

$$\text{Var}(\varepsilon_j) = \sigma^2 z_j$$

Entonces, la matriz de covarianza de varianza  $\sigma^2 V$  de  $\varepsilon$  es diagonal y tiene la forma:

$$V = \begin{bmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_n \end{bmatrix}.$$

La inversa de  $V$  es:

$$V^{-1} = \begin{bmatrix} \frac{1}{z_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{z_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{z_n} \end{bmatrix}.$$

El estimador de mínimos cuadrados ponderados se obtiene como:

$$\begin{aligned}\hat{\beta}_W &= (\mathbf{Z}^T V^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T V^{-1} \mathbf{Y} \\ &= \left( \sum_{i=1}^n \frac{z_i^2}{z_i} \right)^{-1} \sum_{i=1}^n \frac{z_i Y_i}{z_i} \\ &= \left( \sum_{i=1}^n z_i \right)^{-1} \sum_{i=1}^n Y_i.\end{aligned}$$

Por lo tanto, la estimación de la pendiente del modelo es:

$$\hat{\beta}_W = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n z_i}.$$

Todos los  $Y_i$  tienen la misma influencia en la estimación.

(c)

$$\text{Var}(\varepsilon_j) = \sigma^2 z_j^2$$

Entonces, la matriz de covarianza de varianza  $\sigma^2 V$  de  $\varepsilon$  es diagonal y tiene la forma:

$$V = \begin{bmatrix} z_1^2 & 0 & \cdots & 0 \\ 0 & z_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_n^2 \end{bmatrix}.$$

La inversa de  $V$  es:

$$V^{-1} = \begin{bmatrix} \frac{1}{z_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{z_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{z_n^2} \end{bmatrix}.$$

El estimador de mínimos cuadrados ponderados se obtiene como:

$$\begin{aligned} \hat{\beta}_W &= (\mathbf{Z}^T V^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T V^{-1} \mathbf{Y} \\ &= \left( \sum_{i=1}^n \frac{z_i^2}{z_i^2} \right)^{-1} \sum_{i=1}^n \frac{z_i Y_i}{z_i^2} \\ &= \left( \sum_{i=1}^n 1 \right)^{-1} \sum_{i=1}^n \frac{Y_i}{z_i}. \end{aligned}$$

Dado que  $\sum_{i=1}^n 1 = n$ , obtenemos:

$$\hat{\beta}_W = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{z_i}.$$

Así, la estimación de la pendiente del modelo es:

$$\hat{\beta}_W = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{z_i}.$$

En este caso, los valores altos de  $z_i$  reducen la influencia de  $Y_i$  en la estimación.

## 7.6

(a)

mUESTRE QUE

$$(Z^T Z)^- = \sum_{i=1}^{r_1+1} \lambda_i^{-1} e_i e_i^T$$

es una inversa generalizada de  $Z^T Z$ .

Como los valores propios de  $(Z^T Z)$  cumplen que:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{r+1} > 0$$

$(Z^T Z)$  es definida positiva y se puede descomponer como:

$$Z^T Z = P \Lambda P^T$$

Donde

$$P = [\vec{e}_1^T, \dots, \vec{e}_{r+1}^T]^T, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_{r+1})$$

Así,

$$Z^T Z = \sum_{i=1}^{r+1} \lambda_i \vec{e}_i \vec{e}_i^T$$

Sea

$$G = \sum_{i=1}^{r+1} \lambda_i^{-1} \vec{e}_i \vec{e}_i^T$$

Luego,

$$(Z^T Z)G(Z^T Z) =$$

$$\left( \sum_{i=1}^{r+1} \lambda_i \vec{e}_i \vec{e}_i^T \right) \left( \sum_{i=1}^{r+1} \lambda_i^{-1} \vec{e}_i \vec{e}_i^T \right) \left( \sum_{i=1}^{r+1} \lambda_i \vec{e}_i \vec{e}_i^T \right)$$

$$= \sum_{i=1}^{r+1} \lambda_i^2 (\lambda_i^{-1} \vec{e}_i \vec{e}_i^T) \vec{e}_i \vec{e}_i^T$$

$$= \sum_{i=1}^{r+1} \lambda_i \vec{e}_i (\vec{e}_i^T \vec{e}_i) (\vec{e}_i^T \vec{e}_i) \vec{e}_i^T \quad (\text{los vectores propios son ortonormales})$$

$$= \sum_{i=1}^{r+1} \lambda_i \vec{e}_i \vec{e}_i^T$$

$$= (Z^T Z)$$

Así,

$$G = \sum_{i=1}^{r+1} \lambda_i^{-1} \vec{e}_i \vec{e}_i^T$$

es una inversa generalizada de  $Z^T Z$ .

(b)

Los coeficientes  $\hat{\beta}$  que minimizan la suma de los errores cuadráticos  $(y - Z\beta)^T (y - Z\beta)$  satisfacen las *ecuaciones normales*  $(Z^T Z)\hat{\beta} = Z^T y$ . Mostremos que estas ecuaciones se satisfacen para cualquier  $\hat{\beta}$  tal que  $Z\hat{\beta}$  sea la proyección de  $y$  sobre las columnas de  $Z$ .

Si  $Z\hat{\beta}$  es la proyección de  $y$  sobre el espacio generado por las columnas de  $Z$ , es decir,

$$C(Z)$$

se tiene que:

$$y - Z\hat{\beta} \perp C(Z)$$

Esto implica que:

$$Z^T (y - Z\hat{\beta}) = 0$$

Expandiendo la ecuación:

$$Z^T y - Z^T Z\hat{\beta} = 0$$

Finalmente, obtenemos la ecuación normal:

$$Z^T Z\hat{\beta} = Z^T y$$

(c)

Mostremos que

$$Z\hat{\beta} = Z(Z^T Z)^{-} Z^T y$$

es la proyección de  $y$  sobre las columnas de  $Z$ .

Consideremos:

$$q_i = \lambda_i^{-1/2} Z e_i, \quad i = 1, \dots, r+1$$

Se tiene que:

$$\begin{aligned} q_i^T q_i &= \lambda_i^{-1} e_i^T (Z^T Z) e_i \\ &= \lambda_i^{-1} e_i^T \left( \sum_{j=1}^{r+1} e_j \lambda_j e_j^T \right) e_i \\ &= \lambda_i^{-1} e_i^T (\lambda_i e_i e_i^T) e_i \\ &= 1 \end{aligned}$$

Similarmente,

$$\begin{aligned}
q_i^T q_k &= \lambda_i^{-1} e_i^T (Z^T Z) e_k \\
&= \lambda_i^{-1} e_i^T \left( \sum_{j=1}^{r+1} e_j \lambda_j e_j^T \right) e_k \\
&= \lambda_i^{-1} e_i^T (\lambda_i e_i e_i^T) e_k \\
&= 0
\end{aligned}$$

Así, los  $q_i$  para  $i = 1, \dots, r+1$  son ortonormales.

Luego,

$$\begin{aligned}
Z(Z^T Z)^{-1} Z^T &= Z \left( \sum_{i=1}^{r+1} \lambda_i^{-1} e_i e_i^T \right) Z^T \\
&= \sum_{i=1}^{r+1} \lambda_i^{-1} (Z e_i e_i^T Z^T) \\
&= \sum_{i=1}^{r+1} ([\lambda_i^{-1/2} Z e_i][\lambda_i^{-1/2} Z e_i]^T) \\
&= \sum_{i=1}^{r+1} q_i q_i^T
\end{aligned}$$

Así,  $Z(Z^T Z)^{-1} Z^T$  es una suma de proyecciones ortogonales sobre el espacio columna de  $Z$  y, por lo tanto, la proyección de  $y$  sobre  $C(Z)$  es:

$$Z(Z^T Z)^{-1} Z^T y$$

(d)

Mostremos directamente que

$$\hat{\beta} = (Z^T Z)^{-1} Z^T y$$

es una solución a las ecuaciones normales.

$$(Z^T Z)[(Z^T Z)^{-1} Z^T y] = Z^T y.$$

Queremos demostrar que:

$$(Z^T Z)[(Z^T Z)^{-1} Z^T y] = Z^T y.$$

Sustituyendo  $\hat{\beta} = (Z^T Z)^{-1} Z^T y$  en la ecuación:

$$(Z^T Z)\hat{\beta} = (Z^T Z)(Z^T Z)^{-1} Z^T y.$$



Dado que  $(Z^T Z)^-$  es una inversa generalizada de  $Z^T Z$ , se cumple que:

$$(Z^T Z)(Z^T Z)^- = I.$$

Por lo tanto, la ecuación se reduce a:

$$I Z^T y = Z^T y.$$

Como la igualdad es válida, concluimos que  $\hat{\beta} = (Z^T Z)^- Z^T y$  es una solución de las ecuaciones normales.

## 7.9

Una empresa está intentando evaluar la calidad de su personal de ventas y está buscando un examen o una serie de pruebas que puedan revelar el potencial para un buen desempeño en ventas.

La empresa ha seleccionado una muestra aleatoria de 50 vendedores y ha evaluado a cada uno en 3 medidas de desempeño: crecimiento de las ventas, rentabilidad de las ventas y ventas de nuevas cuentas. Estas medidas se han convertido en una escala en la que 100 indica un rendimiento "promedio". Cada uno de los 50 individuos tomó cada una de las 4 pruebas, que supuestamente miden creatividad, razonamiento mecánico, razonamiento abstracto y habilidad matemática, respectivamente. Las  $n = 50$  observaciones sobre  $p = 7$  variables están listadas en la Tabla 9.12 en la página 536.

(a)

Suponga un modelo factorial ortogonal para las variables estandarizadas.

$$Z_i = \frac{(X_i - \mu_i)}{\sqrt{\sigma_{ii}}}, \quad i = 1, 2, \dots, 7.$$

Obtenga la solución de componentes principales o la solución de máxima verosimilitud para  $m = 2$  y  $m = 3$  factores comunes.

**Solución:**

El calculo para  $m = 2$

	Factor1	Factor2	Comunalidad
Sales_growth	0.632	0.742	0.950
Sales_profitability	0.611	0.731	0.907
New_account_sales	0.693	0.588	0.826
Creativity_test	0.993	-0.099	0.095
Mechanical_reasoning_test	0.702	0.375	0.634
Abstract_reasoning_test	0.231	0.662	0.491
Mathematics_test	0.490	0.757	0.813

Table 25: Resultados del modelo factorial ortogonal con  $m = 2$

Calculo para  $m = 3$

	<b>Factor1</b>	<b>Factor2</b>	<b>Factor3</b>	<b>Comunalidad</b>
Sales_growth	0.871	0.362	0.215	0.935
Sales_profitability	0.895	0.437	-0.053	0.995
New_account_sales	0.856	0.192	0.312	0.866
Creativity_test	0.851	-0.520	0.005	0.995
Mechanical_reasoning_test	0.814	0.065	-0.027	0.668
Abstract_reasoning_test	0.430	0.397	0.754	0.911
Mathematics_test	0.772	0.473	0.052	0.822

Table 26: Resultados del modelo factorial con  $m=3$

(b)

Dada su solución en (a), obtenga las cargas rotadas para  $m = 2$  y  $m = 3$ . Compare los dos conjuntos de cargas rotadas. Interprete las soluciones factoriales para  $m = 2$  y  $m = 3$ .

**Solución:**

Usando los mecanismos computacionales hallamos:

**Con  $m = 2$ :**

<b>Variable</b>	<b>Unicidad</b>
Sales_growth	0.050
Sales_profitability	0.093
New_account_sales	0.174
Creativity_test	0.005
Mechanical_reasoning_test	0.366
Abstract_reasoning_test	0.509
Mathematics_test	0.187

Table 27: Unidades de las variables para  $m = 2$

	<b>Factor1</b>	<b>Factor2</b>
Sales_growth	0.880	0.418
Sales_profitability	0.864	0.401
New_account_sales	0.748	0.517
Creativity_test	0.162	0.984
Mechanical_reasoning_test	0.544	0.581
Abstract_reasoning_test	0.699	
Mathematics_test	0.858	0.277

Table 28: Cargas factoriales rotadas para  $m = 2$

	<b>Factor1</b>	<b>Factor2</b>
SS loadings	3.628	1.989
Proportion Var	0.518	0.284
Cumulative Var	0.518	0.802

Table 29: Varianza explicada por los factores para  $m = 2$

**Con  $m = 2$ :**

Variable	Unicidad
Sales_growth	0.065
Sales_profitability	0.005
New_account_sales	0.134
Creativity_test	0.005
Mechanical_reasoning_test	0.332
Abstract_reasoning_test	0.089
Mathematics_test	0.178

Table 30: Unidades de las variables ( $m = 3$ )

	Factor1	Factor2	Factor3
Sales_growth_x1	0.763	0.342	0.486
Sales_profitability_x2	0.918	0.292	0.258
New_account_sales_x3	0.612	0.459	0.530
Creativity_test_x4	0.260	0.961	
Mechanical_reasoning_test_x5	0.616	0.509	0.169
Abstract_reasoning_test_x6	0.299		0.906
Mathematics_test_x7	0.820	0.187	0.339

Table 31: Cargas factoriales rotadas ( $m = 3$ )

	Factor1	Factor2	Factor3
SS loadings	3.009	1.633	1.550
Proportion Var	0.430	0.233	0.221
Cumulative Var	0.430	0.663	0.885

Table 32: Varianza explicada por los factores ( $m = 3$ )

La prueba de la hipótesis de que tres factores son suficientes arrojó un estadístico  $\chi^2$  de 6.94 con 3 grados de libertad y un valor  $p$  de 0.0738.

La prueba de la hipótesis de que dos factores son suficientes arrojó un estadístico  $\chi^2$  de 43.63 con 8 grados de libertad y un valor  $p$  de 6.68e-07.

Sobre las cargas factoriales rotadas concluimos:

- En la solución con  $m = 2$ , los factores agrupan las variables en dos dimensiones principales: desempeño en ventas y habilidades cognitivas generales. - En la solución con  $m = 3$ , se logra una mayor diferenciación entre tipos de habilidades cognitivas, separando razonamiento abstracto y mecánico. - La varianza acumulada aumenta de 80.2% para  $m = 2$  a 88.5% para  $m = 3$ , lo que indica que la solución con tres factores captura más información.

Respecto a los factores comunes:

#### Para $m = 2$

- **Factor 1:** Representa el desempeño en ventas, con altas cargas en Sales Growth, Sales Profitability y New Account Sales.
- **Factor 2:** Representa habilidades cognitivas generales, incluyendo creatividad, razonamiento mecánico y abstracto.

#### Para $m = 3$

- **Factor 1:** Explica la capacidad comercial y desempeño en ventas.
- **Factor 2:** Captura habilidades creativas y de razonamiento mecánico.
- **Factor 3:** Diferencia la habilidad matemática y razonamiento abstracto.

(c)

Lista de comunalidades estimadas, varianzas específicas y  $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$  para  $m = 2$  y  $m = 3$

Compare los resultados. ¿Qué elección de  $m$  prefiere en este punto? ¿Por qué?

**Solución:**

Para **m=2**:

Variable	Comunalidad	Varianza Específica
Sales_growth	0.950	0.050
Sales_profitability	0.907	0.093
New_account_sales	0.826	0.174
Creativity_test	0.995	0.005
Mechanical_reasoning_test	0.634	0.366
Abstract_reasoning_test	0.491	0.509
Mathematics_test	0.812	0.188

Table 33: Comunalidades y varianzas específicas  $m = 2$

Matriz estimada ( $\mathbf{L} \mathbf{L}' + \mathbf{\Psi}$ ,  $m=2$ ):

```
## Sales_growth_x1 Sales_profitability_x2
## Sales_growth_x1      1.0000000      0.9260758
## Sales_profitability_x2 0.9260758      1.0000358
## New_account_sales_x3   0.8817155      0.8332024
## Creativity_test_x4     0.5542547      0.5341223
## Mechanical_reasoning_test_x5 0.7267483      0.7585336
## Abstract_reasoning_test_x6 0.6805357      0.5187499
## Mathematics_test_x7    0.8544378      0.8949230
## New_account_sales_x3 Creativity_test_x4
## Sales_growth_x1      0.8817155      0.5542547
## Sales_profitability_x2 0.8332024      0.5341223
## New_account_sales_x3   1.0000002      0.6297578
## Creativity_test_x4     0.6297578      1.0000053
## Mechanical_reasoning_test_x5 0.7007363      0.6594334
## Abstract_reasoning_test_x6 0.6795282      0.1633138
## Mathematics_test_x7    0.7676301      0.4117714
## Mechanical_reasoning_test_x5
## Sales_growth_x1      0.7267483
## Sales_profitability_x2 0.7585336
## New_account_sales_x3   0.7007363
## Creativity_test_x4     0.6594334
## Mechanical_reasoning_test_x5 1.0000002
## Abstract_reasoning_test_x6 0.3557349
## Mathematics_test_x7    0.6579484
## Abstract_reasoning_test_x6 Mathematics_test_x7
## Sales_growth_x1      0.6805357      0.8544378
## Sales_profitability_x2 0.5187499      0.8949230
## New_account_sales_x3   0.6795282      0.7676301
## Creativity_test_x4     0.1633138      0.4117714
## Mechanical_reasoning_test_x5 0.3557349      0.6579484
## Abstract_reasoning_test_x6 0.9999999      0.5589944
## Mathematics_test_x7    0.5589944      1.0000008
```

Para **m=3**:

Variable	Comunalidad	Varianza Específica
Sales_growth_x1	0.936	0.064
Sales_profitability_x2	0.995	0.005
New_account_sales_x3	0.866	0.134
Creativity_test_x4	0.995	0.005
Mechanical_reasoning_test_x5	0.669	0.331
Abstract_reasoning_test_x6	0.910	0.089
Mathematics_test_x7	0.822	0.178

Matriz estimada (L L' + Psi, m=3):

	Sales_growth_x1	
Sales_growth_x1	1.0000000	
Sales_profitability_x2	0.9260758	
New_account_sales_x3	0.8817155	
Creativity_test_x4	0.5542547	
Mechanical_reasoning_test_x5	0.7267483	
Abstract_reasoning_test_x6	0.6805357	
Mathematics_test_x7	0.8544378	
	Sales_profitability_x2	
Sales_growth_x1	0.9260758	
Sales_profitability_x2	1.0000358	
New_account_sales_x3	0.8332024	
Creativity_test_x4	0.5341223	
Mechanical_reasoning_test_x5	0.7585336	
Abstract_reasoning_test_x6	0.5187499	
Mathematics_test_x7	0.8949230	
	New_account_sales_x3	
Sales_growth_x1	0.8817155	
Sales_profitability_x2	0.8332024	
New_account_sales_x3	1.0000002	
Creativity_test_x4	0.6297578	
Mechanical_reasoning_test_x5	0.7007363	
Abstract_reasoning_test_x6	0.6795282	
Mathematics_test_x7	0.7676301	
	Creativity_test_x4	
Sales_growth_x1	0.5542547	
Sales_profitability_x2	0.5341223	
New_account_sales_x3	0.6297578	
Creativity_test_x4	1.0000053	
Mechanical_reasoning_test_x5	0.6594334	
Abstract_reasoning_test_x6	0.1633138	
Mathematics_test_x7	0.4117714	
	Mechanical_reasoning_test_x5	
Sales_growth_x1	0.7267483	
Sales_profitability_x2	0.7585336	
New_account_sales_x3	0.7007363	
Creativity_test_x4	0.6594334	
Mechanical_reasoning_test_x5	1.0000002	
Abstract_reasoning_test_x6	0.3557349	
Mathematics_test_x7	0.6579484	
	Abstract_reasoning_test_x6	
Sales_growth_x1	0.6805357	
Sales_profitability_x2	0.5187499	
New_account_sales_x3	0.6795282	

Creativity_test_x4	0.1633138
Mechanical_reasoning_test_x5	0.3557349
Abstract_reasoning_test_x6	0.9999999
Mathematics_test_x7	0.5589944
	Mathematics_test_x7
Sales_growth_x1	0.8544378
Sales_profitability_x2	0.8949230
New_account_sales_x3	0.7676301
Creativity_test_x4	0.4117714
Mechanical_reasoning_test_x5	0.6579484
Abstract_reasoning_test_x6	0.5589944
Mathematics_test_x7	1.0000008

- Para  $m = 3$ , las comunales son ligeramente más altas, indicando una mejor explicación de la varianza compartida.
- En  $m = 2$ , la varianza específica es mayor en algunas variables, sugiriendo que quedan aspectos sin capturar.

**Matriz  $LL' + \Psi$ :**

- Para  $m = 3$ , la matriz  $LL' + \Psi$  se ajusta mejor a la matriz de correlaciones original.
- Para  $m = 2$ , hay mayores discrepancias en la estimación de correlaciones.

(d)

Realice una prueba de

$$H_0 : \Sigma = LL' + \Psi$$

frente a

$$H_1 : \Sigma \neq LL' + \Psi$$

para  $m = 2$  y  $m = 3$  al nivel de significancia  $\alpha = 0.01$ .

Con estos resultados y los obtenidos en las Partes (b) y (c), ¿qué elección de  $m$  parece ser la mejor?

**Solución:**

Dado el p-valor que se vio en la salida anterior decimos que existe evidencia estadística suficiente para rechazar la hipótesis nula  $H_0$  con  $m=2$ .

El resultado nos dice que 2 factores no son suficientes para explicar la matriz de correlaciones.

De forma similar, con base en el p-valor de la salida de arriba cuando  $m=3$ , no se rechaza  $H_0$ .

Esto indica que 3 factores son suficientes para representar a la matriz de correlación.

(e)

Supongamos que un nuevo vendedor, seleccionado al azar, obtiene los siguientes puntajes en las pruebas:

$$\mathbf{x}' = [x_1, x_2, \dots, x_7] = [110, 98, 105, 15, 18, 12, 35].$$

Calcule la puntuación factorial del vendedor utilizando el método de mínimos cuadrados ponderados y el método de regresión.

**Nota:** Los componentes de  $\mathbf{x}$  deben ser estandarizados utilizando las medias muestrales y las varianzas calculadas a partir de los datos originales.

**Solución:**

Variable	Factor 1	Factor 2	Factor 3
Sales_growth_x1	0.871	0.362	0.215
Sales_profitability_x2	0.895	0.437	-0.053
New_account_sales_x3	0.856	0.192	0.312
Creativity_test_x4	0.851	-0.520	0.005
Mechanical_reasoning_test_x5	0.814	0.065	-0.027
Abstract_reasoning_test_x6	0.430	0.397	0.754
Mathematics_test_x7	0.772	0.473	0.052

Table 34: Matriz de cargas factoriales  $L$

Variable	$\Psi$ (Varianza Específica)
Sales_growth_x1	0.065
Sales_profitability_x2	0.005
New_account_sales_x3	0.134
Creativity_test_x4	0.005
Mechanical_reasoning_test_x5	0.332
Abstract_reasoning_test_x6	0.089
Mathematics_test_x7	0.178

Table 35: Matriz diagonal de varianzas específicas  $\Psi$

El nuevo vendedor obtuvo los siguientes puntajes estandarizados:

$$\mathbf{z} = [1.522, -0.852, 0.433, 1.046, 1.142, 0.609, 0.405].$$

**Resultados de los métodos de estimación:**

- **Mínimos Cuadrados Ponderados (WLS):**

$$\hat{F}_{WLS} = \begin{bmatrix} 0.171 \\ -1.741 \\ 2.537 \end{bmatrix}$$

- **Método de Regresión:**

$$\hat{F}_{REG} = \begin{bmatrix} 0.170 \\ -1.724 \\ 2.267 \end{bmatrix}$$

Los resultados muestran que el vendedor presenta un desempeño:

- **Ligeramente superior al promedio** en el Factor 1.
- **Significativamente menor** en el Factor 2.
- **Claramente elevado** en el Factor 3.

Esto sugiere que el vendedor tiene fortalezas destacadas en las dimensiones asociadas al Factor 3, una posición moderadamente buena en el Factor 1 y ciertas debilidades en lo que representa el Factor 2.

## 7.11

Sea  $\mathbf{A}$  una matriz definida positiva, de modo que

$$d_j^2(\mathbf{B}) = (\mathbf{y}_j - \mathbf{Bz}_j)^T \mathbf{A} (\mathbf{y}_j - \mathbf{Bz}_j)$$

es una distancia estadística cuadrática desde la  $j$ -ésima observación  $\mathbf{y}_j$  hasta su regresión  $\mathbf{B}\mathbf{z}_j$ . Mostremos que la elección

$$\mathbf{B} = \hat{\mathbf{B}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y}$$

minimiza la suma de las distancias estadísticas cuadráticas,

$$\sum_{j=1}^n d_j^2(\mathbf{B}),$$

para cualquier elección de  $\mathbf{A}$  definida positiva. Algunas elecciones para  $\mathbf{A}$  incluyen  $\mathbf{\Sigma}^{-1}$  y  $\mathbf{I}$ .

*Sugerencia:* Repita los pasos en la demostración del Resultado 7.10 con  $\mathbf{\Sigma}^{-1}$  reemplazada por  $\mathbf{A}$ .

Sea  $\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y}$  y

$$\begin{aligned} \sum_{j=1}^n d_j^2(B) &= \sum_{j=1}^n (\mathbf{Y}_j - \mathbf{B}\mathbf{Z}_j)^T \mathbf{A} (\mathbf{Y}_j - \mathbf{B}^T \mathbf{Z}_j) \\ &= (\mathbf{Y} - \mathbf{Z}\mathbf{B})^T \mathbf{A} (\mathbf{Y} - \mathbf{Z}\mathbf{B}) \end{aligned}$$

Se tiene que:

$$\begin{aligned} \text{Traza} \left( \sum_{j=1}^n d_j^2(B) \right) &= \text{Traza} ((\mathbf{Y} - \mathbf{Z}\mathbf{B})^T \mathbf{A} (\mathbf{Y} - \mathbf{Z}\mathbf{B})) \\ &= \text{Traza} (\mathbf{A} (\mathbf{Y} - \mathbf{Z}\hat{\beta})^T (\mathbf{Y} - \mathbf{Z}\hat{\beta})) \\ &= \text{Traza} \left( \mathbf{A} (\mathbf{Y} + \mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta})^T (\mathbf{Y} + \mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta}) \right) \\ &= \text{Traza} \left( \mathbf{A} \left( (\mathbf{Y} - \mathbf{Z}\hat{\beta}) + (\mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta}) \right)^T \left( (\mathbf{Y} - \mathbf{Z}\hat{\beta}) + (\mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta}) \right) \right) \\ &= \text{Traza} \left( \mathbf{A} \left( \hat{\varepsilon} + (\mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta}) \right)^T \left( \hat{\varepsilon} + (\mathbf{Z}\hat{\beta} - \mathbf{Z}\hat{\beta}) \right) \right) \\ &= \text{Traza} \left( \mathbf{A} \left( \hat{\varepsilon}\hat{\varepsilon}^T + \mathbf{Z}(\hat{\beta} - \hat{\beta})^T \hat{\varepsilon} + (\hat{\beta} - \hat{\beta})^T \mathbf{Z}^T \hat{\varepsilon} + (\hat{\beta} - \hat{\beta})^T \mathbf{Z}^T \mathbf{Z} (\hat{\beta} - \hat{\beta}) \right) \right) \end{aligned}$$

Dado que  $\mathbf{Z}$  y  $\mathbf{Y} - \mathbf{Z}\hat{\beta}$  son ortogonales, el calculo se reduce a:

$$\text{Traza} \left( \sum_{j=1}^n d_j^2(B) \right) = \text{Traza} (\mathbf{A} \hat{\varepsilon}^T \hat{\varepsilon}) + \text{Traza} \left( \mathbf{A} (\hat{\beta} - \hat{\beta})^T \mathbf{Z}^T \mathbf{Z} (\hat{\beta} - \hat{\beta}) \right)$$

Note que el primer termino no depende de la elección de  $\beta$ , centrandonos en el segundo termino:

$$\text{Traza} \left( \mathbf{A} (\hat{\beta} - \beta)^T \mathbf{Z}^T \mathbf{Z} (\hat{\beta} - \beta) \right)$$



$$\begin{aligned}
&= \text{Traza} \left( (\hat{\beta} - \beta)^T Z^T Z (\hat{\beta} - \beta) A \right) \\
&= \text{Traza} \left( (Z^T Z (\hat{\beta} - \beta)) A (\hat{\beta} - \beta)^T \right) \\
&= \text{Traza} \left( Z (\hat{\beta} - \beta) A (\hat{\beta} - \beta)^T Z^T \right) \\
&= \text{Traza} (C^T A C)
\end{aligned}$$

$$\geq 0 \quad (A \text{ es definida positiva})$$

Esta expresión solo se anula si  $\beta = \hat{\beta}$ , por lo que  $\hat{\beta}$  es la mejor opción para cualquier  $A$  definida positiva.

## 7.19

Satellite applications motivated the development of a silver-zinc battery. Table 7.5 contains failure data collected to characterize the performance of the battery during its life cycle. Use these data.

(a)

Encuentre la regresión lineal estimada de  $\ln(Y)$  sobre un subconjunto adecuado (“óptimo”) de variables predictoras.

```
fit1<-lm(log(Y)~Z2+Z4, data = battery_data)
round(fit1$coefficients,3)
```

```
## (Intercept)      Z2      Z4
##      2.756    -0.322    0.114
```

(b)

Grafique los residuos del modelo ajustado seleccionado en la Parte (a) para verificar el supuesto de normalidad.

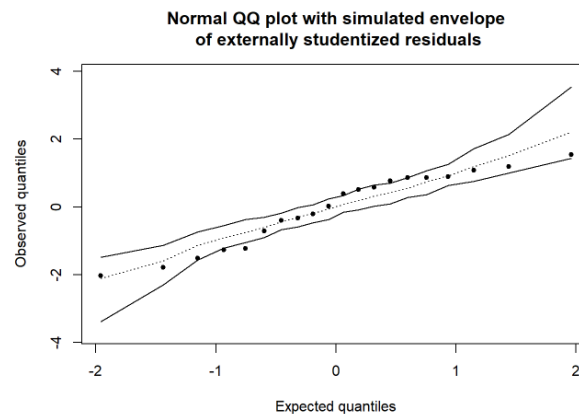


Figure 4: QQ-plot de los residuos del modelo propuesto.

Los residuos se ajustan bastante bien al supuesto de normalidad.

## 7.20

Usando los datos de fallos de batería en la Tabla 7.5, realice una regresión de  $\ln(Y)$  sobre el primer componente principal de las variables predictoras  $z_1, z_2, \dots, z_5$ . (Véase la Sección 8.3.) Compare el resultado con el modelo ajustado obtenido en el Ejercicio 7.19(a).

Primero realizamos el acp:

```
R<-cor(battery_data[, -6])

lambdas<-eigen(R)$values
e<-eigen(R)$vectors

PCA<-scale(battery_data[, -6])%*%e

knitr::kable(PCA, col.names = c("CP1", "CP2", "CP3", "CP4", "CP5"))
```

	CP1	CP2	CP3	CP4	CP5
-1.3145022	-1.3008095	-0.7467948	1.2627956	0.7766756	
-0.9258234	-1.1361096	0.4063741	0.1053045	-0.6854285	
0.0361085	0.0393150	-0.0338813	0.1398604	0.1796813	
0.0453077	-1.3593990	1.1639773	-0.7947477	0.2491085	
1.7968554	1.3582494	0.0058230	0.5193101	0.1920623	
0.8349235	0.1828248	0.4460784	0.4847541	-0.6730475	
0.8257243	1.5815388	-0.7517802	1.4193623	-0.7424747	
-1.6945173	1.8415278	-0.0947659	-0.3316148	0.6716481	
0.4947645	0.3418442	1.6674461	-0.3189967	1.4322586	
-0.0662304	0.5439435	0.2366337	2.0834403	1.1070969	
-2.1207327	0.8561807	0.8093741	-0.1360071	-1.3402690	
-0.0876879	0.7298336	2.0630119	-0.5764356	-0.9643823	
-0.6668264	0.7870787	-1.9872610	-1.2163643	0.0338665	
1.5224553	-0.7126049	-0.2250490	-1.2037462	0.7944771	
-0.4152074	-0.9653831	-1.1413260	0.1290541	-1.0398832	
1.1031198	0.7072992	-0.7496983	-1.7715228	0.1156747	
1.7694747	-1.7657097	0.0219567	0.6089763	-0.3139862	
1.3501392	-0.3458057	-0.5026925	0.0411997	-0.9927886	
-1.7246385	-1.2796194	-0.0735855	-0.2395892	0.1673003	
-0.7627066	-0.1041948	-0.5138409	-0.2050332	1.0324101	

El primer componente principal explica el siguiente porcentaje de varianza:

```
## [1] "28.93 %"
```

Realizamos la regresión con el primer componente principal:

```
## (Intercept)    PCA[, 1]
##      1.7383      -0.0701
```

## 7.21

Considere los datos de contaminación del aire en la Tabla 1.5. Sea  $Y_1 = \text{NO}_2$  y  $Y_2 = \text{O}_3$  las dos respuestas (contaminantes) correspondientes a las variables predictoras  $Z_1 = \text{viento}$  y  $Z_2 = \text{radiación solar}$ .

(a)

Realice un análisis de regresión utilizando solo la primera respuesta  $Y_1$ .

(i) Sugerir y ajustar modelos de regresión lineal adecuados.

```
fit<-lm(x5~x2+I(x2^2), data = data)
summary(fit)

##
## Call:
## lm(formula = x5 ~ x2 + I(x2^2), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2751 -2.0022 -0.5498  2.0400  9.8151
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.380812   5.322651  -1.387  0.17342
## x2             0.528148   0.159459   3.312  0.00200 **
## I(x2^2)       -0.003754   0.001166  -3.219  0.00259 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.052 on 39 degrees of freedom
## Multiple R-squared:  0.2205, Adjusted R-squared:  0.1805
## F-statistic: 5.517 on 2 and 39 DF,  p-value: 0.007766
```

(ii) Analice los residuales.

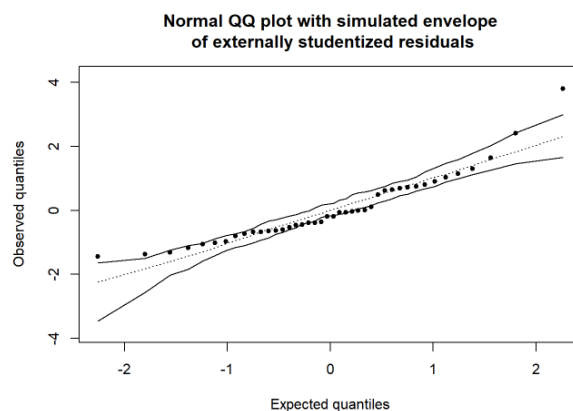


Figure 5: QQ-plot de los residuos del modelo propuesto.

Hay varios puntos fuera de las bandas de confianza. Se sospecha que el supuesto de normalidad en los residuos puede fallar.

(iii) Construya un intervalo de predicción del 95% para  $\text{NO}_2$  correspondiente a  $z_1 = 10$  y  $z_2 = 80$ .

Se usa:

$$z_0 \hat{\beta}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)} \sqrt{(1 + z_0^T (Z^T Z)^{-1} z_0) \left(\frac{n}{n-r-1} \hat{\sigma}_{ii}\right)}$$

```

z0 <- c(1,80,80^2) # Nuevo punto de predicción
z0beta<-t(z_0) %*%fit$coefficients
n<-nrow(PCA)
sigma_sq <- sum(residuals(fit)^2) / fit$df.residual
a<-sqrt(qf(.95,1,42-3))
X_mat <- model.matrix(fit) # Matriz de diseño
X_inv <- solve(t(X_mat) %*% X_mat) # Inversa de X'X
b<-sqrt((1+t(z0)%*%X_inv%*%z0)(n*sigma_sq/fit$df.residual)
knitr::kable(z0beta+c(a*b,-a*b),col.names = c("LI","LS"))

```

	LI	LS
	5.316944	16.3735

(b)

Realice un análisis de regresión múltiple multivariante utilizando ambas respuestas  $Y_1$  y  $Y_2$ .

(i) Sugerir y ajustar modelos de regresión lineal adecuados.

```

fit1<-lm(cbind(x5,x6)~x1+x2, data = data)
summary(fit1)

```

```

## Response x5 :
##
## Call:
## lm(formula = x5 ~ x1 + x2, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.7521 -2.2053 -0.5917  1.6852 10.4623
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.11454    3.62607   2.789  0.00813 **
## x1          -0.21129    0.33917  -0.623  0.53694
## x2           0.02055    0.03094   0.664  0.51042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.416 on 39 degrees of freedom
## Multiple R-squared:  0.02311,    Adjusted R-squared:  -0.02698
## F-statistic: 0.4614 on 2 and 39 DF,  p-value: 0.6338
##
##
## Response x6 :
##
## Call:
## lm(formula = x6 ~ x1 + x2, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.808 -3.256 -1.557  2.143 18.327
##

```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.42347    5.12061   0.473   0.639
## x1           0.05890    0.47897   0.123   0.903
## x2           0.04857    0.04369   1.112   0.273
##
## Residual standard error: 4.824 on 39 degrees of freedom
## Multiple R-squared:  0.03072,    Adjusted R-squared:  -0.01898
## F-statistic: 0.6181 on 2 and 39 DF,  p-value: 0.5442
```

(ii) Analyze the residuals.

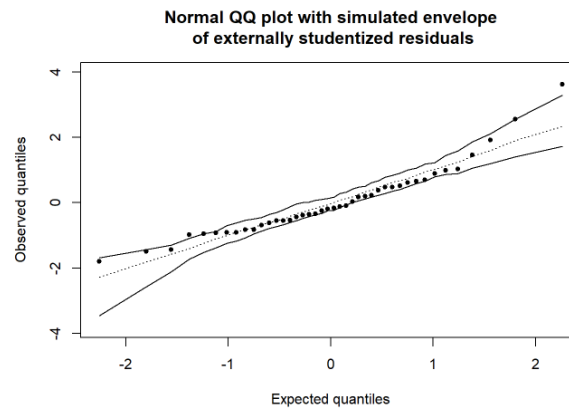


Figure 6: QQ-plot de los residuos del modelo propuesto para  $x_5$ .

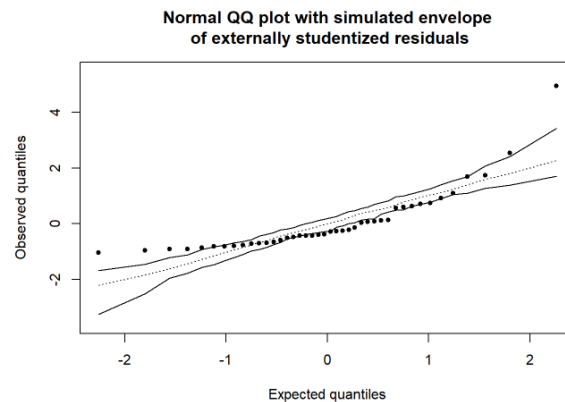


Figure 7: QQ-plot de los residuos del modelo propuesto para  $x_6$ .

Hay varios puntos fuera de las bandas de confianza. Se sospecha que el supuesto de normalidad en los residuos puede fallar.

(iii) Construya una elipse de predicción del 95% para  $\text{NO}_2$  y  $\text{O}_3$  correspondiente a  $z_1 = 10$  y  $z_2 = 80$ . Compare esta elipse con el intervalo de predicción en la Parte (a)(iii) y comente.

```
library(ellipse)
library(car)
```

```

z0<-c(1,10,80)

m<-2
n<-nrow(data)
p<-fit1$df.residual

Z<-model.matrix(fit1)

m_inv<-solve(t(Z)%*%Z)

center<-t(z0)%*%fit1$coefficients

F_crit <- qf(0.95, m, p - m)

# Varianza de predicción (usando matriz de covarianza y sigma estimado)
sigma_sq <- c(3.416, 4.824) # Estimaciones de varianza para NO2 y O3

# Calcular el radio de la elipse
ratio <- sqrt(m * p * F_crit / (p - m)) *
  sqrt((1 + t(z0) %*% m_inv %*% z0) * (n * sigma_sq / p))

```

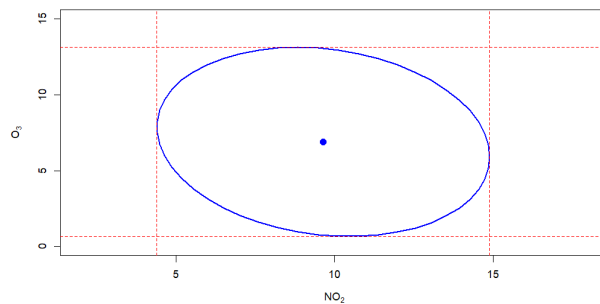


Figure 8: Elipse de confianza para observaciones futuras.

Los límites para la confianza de las observaciones futuras varían especialmente porque en este caso se considera un modelo completamente lineal.