ejemplo: para los datos de los billetes del banco suizo, el vector de medias

$$\bar{x}_{(n)} = \begin{pmatrix} 214.9 \\ 130.1 \\ 129.9 \\ 9.4 \\ 10.6 \\ 140.5 \end{pmatrix}$$

Los valores propios de $\hat{\Sigma}$, es decir, los elementos de la diagonal en la matriz $\mathcal{L}$ son

$$\begin{pmatrix} 2.985 \\ 0.931 \\ 0.242 \\ 0.194 \\ 0.085 \\ 0.035 \end{pmatrix}$$

Los vectores propios de $\hat{\Sigma}$, es decir las columnas $g_1, \ldots, g_p$ de $\mathcal{G}$

$$\mathcal{G} = \begin{pmatrix} -0.0437 & 0.011 & -0.326 & 0.562 & 0.753 & 0.098 \\ 0.112 & 0.071 & -0.259 & 0.455 & -0.347 & -0.767 \\ 0.139 & 0.066 & -0.345 & 0.415 & -0.535 & 0.632 \\ 0.768 & -0.563 & -0.218 & -0.186 & 0.100 & -0.022 \\ 0.202 & 0.659 & -0.557 & -0.451 & 0.102 & -0.035 \\ -0.579 & -0.489 & -0.592 & -0.258 & -0.085 & -0.046 \end{pmatrix}$$

Como ejemplo calculamos las componentes principales para los datos de los billetes del banco suizo. La figura C nos muestra gráficos de dispersión para: $(Y_1, Y_2)$ (Arriba a la izquierda), $(Y_2, Y_3)$ (Arriba a la derecha), $(Y_1, Y_3)$ (Abajo a la izquierda). Los caracteres "o" ó "+" corresponden a "verdaderos" ó "falsos". La gráfica abajo a la izquierda en la figura C, corresponde a los valores propios $l_1, l_2, \ldots, l_6$ de $\hat{\Sigma}$.

<u>Ejercicio</u>: Calcular de nuevo las componentes pero usando un reescalamiento de los datos $\tilde{\mathcal{X}}$. Por ejemplo si se asume que las variables $X_1, X_2, X_3$ y $X_6$ fueron medidos en cms. y que $X_4$ y $X_5$ se quedan como estaban originalmente, o sea en escala de mm., esto sería equivalente a re-escalar $\tilde{X}_1 = X_1/10$ ; $\tilde{X}_2 = X_2/10$ ; $\tilde{X}_3 = \frac{X_3}{10}$ y $\tilde{X}_6 = X_6/10$ y usar la matriz de datos $\tilde{\mathcal{X}}$

Este ejercicio tiene el fin de ilustrar que los gráficos pueden cambiar mucho. Dicho de otra forma: "LAS COMPONENTES PRINCIALES SON SENSIBLES A CAMBIOS DE ESCALA EN LAS VARIABLES".

Basta re-escalar una de las columnas en $\mathcal{X}$ para que los valores propios $l_1, \ldots, l_p$ y los vectores propios $g_1, g_2, \ldots, g_p$ cambien

Veremos más adelante que una forma de evitar este problema es trabajar con la matriz de correlaciones $\mathcal{R}$ de los datos $\mathcal{X}$.

Usando las primeras dos columnas de $\Gamma$ en la página 38, la primera y segunda componentes principales para los datos del billetes del banco suizo son

$$y_1 = 0.044\,x_1 + 0.112\,x_2 + 0.139\,x_3 + 0.768\,x_4 + 0.202\,x_5 - 0.579\,x_6$$

$$y_2 = 0.011\,x_1 + 0.071\,x_2 + 0.066\,x_3 - 0.563\,x_4 + 0.659\,x_5 - 0.489\,x_6$$

$x_1$ = longitud del billete.

$x_2$ = ancho del billete (a la izquierda).

$x_3$ = ancho del billete (a la derecha).

$x_4$ = distancia de la figura en el billete al borde inferior del billete

$x_5$ = distancia de la figura en el billete al borde superior del billete.

$x_6$ = longitud de la diagonal del billete

De forma que la primera componente $y_1$, esencialmente corresponde a la diferencia entre el ancho del borde inferior del billete y la longitud de la diagonal

La segunda componente principal $y_2$, corresponde a la diferencia entre el ancho del borde superior del billete y la suma de: el ancho del borde inferior y la longitud de la diagonal en los billetes.

Una medida que nos dice qué tan bien explican la variabilidad las primeras q componentes principales, está dada por

$$\psi_q \equiv \frac{\sum_{j=1}^{q} \lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\sum_{j=1}^{q} VAR(Y_j)}{\sum_{j=1}^{p} VAR(Y_j)}$$

La figura D muestra una gráfica de las parejas ordenadas $\left(i, \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}\right)$ ; $i=1,2,\ldots,6$.

Podemos darnos cuenta de que la 1ª componente principal explica un 66 %. de la variabilidad y las primeras dos componentes explican 88 %. . En algunos libros se sugiere seleccionar $i_0 \in \{1,2,\ldots,6\}$ en donde la gráfica de

La figura D [1] presenta un "codo" o se dobla. Este valor de io es el número de componentes principales que se usarán para representar a los datos originales $\mathcal{X}$.

El siguiente teorema resultará de interés

TEOREMA Sea $\mathbb{X}$ un vector aleatorio de dimensión $p \times 1$ tal que $\mathbb{E}(\mathbb{X}) = \mu$ y $VAR(\mathbb{X}) = \Sigma$. Sea $\mathbb{Y}$ el vector de componentes principales de $\mathbb{X}$, entonces

(1) $COV(\mathbb{X}, \mathbb{Y}) = \Gamma \Lambda$,

donde $\Gamma$ es la matriz de dimensiones $p \times p$ cuyas columnas son los vectores propios de $\Sigma$ (en la descomposición de Jordan de $\Sigma$) y $\Lambda$ es una matriz diagonal (de dimensiones $p \times p$) que tiene los valores propios de $\Sigma$, $\lambda_1, \dots, \lambda_p$ como elementos de la diagonal.

(1) Scree Plot

(2) La correlación $\rho_{X_iY_j}$ entre la variable $X_i$ y la componente principal $Y_j$ está dada por

$$\rho_{X_iY_j} = \gamma_{ij}\left(\frac{\lambda_j}{\sigma^2_{X_iX_i}}\right)^{1/2}, \qquad \text{..... (rho)}$$

donde $\sigma^2_{X_iX_i} = VAR(X_i) = \Sigma_{ii}$ la entrada $i,i$ de la matriz $\Sigma$ y $\gamma_{ij}$ es la entrada $i,j$ de la matriz $\Gamma$.

---

(☆) Sea $\gamma_i$ la columna $i$ en $\Gamma$, notemos que

$\sum_{j=1}^{p} \lambda_j \gamma_{ij}^2 = \gamma_i' \Delta \gamma_i$ = elemento $(i,i)$ de la matriz $\Gamma\Delta\Gamma' = \Sigma$, entonces tenemos

$$\sum_{j=1}^{p} \rho^2_{X_iY_j} = \sum_{j=1}^{p} \gamma_{ij}^2 \frac{\lambda_j}{\Sigma_{ii}} = \frac{\sum_{j=1}^{p} \gamma_{ij}^2 \lambda_j}{\Sigma_{ii}}$$

$$= \frac{\Sigma_{ii}}{\Sigma_{ii}} = 1.$$

La anterior propiedad nos inspira a interpretar la "correlación" $\rho^2_{X_iY_j}$ como "la proporción de de la varianza de la variable $X_i$ explicada por la j-ésima componente principal $Y_j$".

Siguiendo estas ideas, el porcentaje de la varianza de $X_i$ explicado por las primeras q componentes principales $Y_1, \dots, Y_q$ está dado por $\sum_{j=1}^{q} \rho^2_{X_iY_j}$.

Como todas las cantidades mencionadas son poblacionales, necesitamos definir como calcular la correlación en la ecuación (rho) para unos datos $\mathbb{x}_1, \dots, \mathbb{x}_n$, esta correlación muestral está dada por

$$r_{X_iY_j} = g_{ij}\left(\frac{l_j}{\hat{\Sigma}_{ii}}\right)^{1/2}.$$

Notemos que podemos argumentar como en (☆) pero con cantidades muestrales:

$\sum_{j=1}^{p} l_j g_{ij}^2 = \mathbb{g}_i' \mathbb{L} \mathbb{g}_i$ = elemento (i,i) de la matriz $\mathbb{G}\mathbb{L}\mathbb{G}' = \hat{\Sigma}$ ($\mathbb{g}_i$ = columna i en $\mathbb{G}$), de lo anterior tenemos que

$$\sum_{j=1}^{p} r^2_{X_iY_j} = \sum_{j=1}^{p} g^2_{ij} \left( \frac{l_j}{\hat{\Sigma}_{ii}} \right) = \frac{\sum_{j=1}^{p} g^2_{ij} l_j}{\hat{\Sigma}_{ii}}$$

$$= \frac{\hat{\Sigma}_{ii}}{\hat{\Sigma}_{ii}} = 1 .$$

nuevamente podemos interpretar $r^2_{X_iY_j}$ como "la proporción de la varianza de de $X_i$ explicada por la j-ésimo componente principal $Y_j$".

Para incorporar este análisis en una gráfica, podemos graficar las parejas ordenadas $(r_{X_iY_1}, r_{X_i,Y_2})$ ; $i=1,2,\dots,p$ en el plano. Dadas las observaciones arriba, tenemos que $r^2_{X_iY_1} + r^2_{X_iY_2} \leq 1$, de forma que las parejas $(r_{X_iY_1}, r_{X_iY_2})$ yacen dentro del círculo de radio 1 en el plano. La **figura** E muestra estas parejas para los datos de los billetes del banco Suizo.

Ahí observamos que para las variables $X_4$ $X_5$ y $X_6$ las correspondientes parejas $(r_{X_i,Y_1}, r_{X_iY_2})$ estan más cercanas a la frontera del círculo unitario $(r^2_{X_iY_1} + r^2_{X_iY_2} \approx 1)$, siguiendo el razonamiento antes descrito concluimos que $Y_1$ y $Y_2$ están muy correlacionadas(1) con $X_4$, $X_5$ y $X_6$, pero no lo están al mismo grado con $X_1$, $X_2$ y $X_3$.

De la figura E, la correlación $r_{X_6,Y_1}$ es negativa, la correlación $r_{X_6,Y_2}$ es negativa y ambos son grandes, pero la correlación $r_{X_4,Y_1}$ es positiva y la correlación $r_{X_4,Y_2}$ es negativa, siendo ambas grandes. Esta información coincide con lo que teníamos

(1) La varianza de $X_4$, $X_5$ y $X_6$ esta bien explicada por $Y_1$ y $Y_2$.

observado antes: el coeficiente de $x_6$ en en $y_1$ es negativo y su magnitud grande (página 41), asimismo el coeficiente de $x_4$ es positivo (ecuación para $y_1$, página 41) y su magnitud es grande. En otras palabras $y_1$ es básicamente determinado por la diferencia entre $x_4$ y $x_6$.

En forma análoga $r_{x_6,y_2}$ y $r_{x_4,y_2}$ son negativos y grandes[1] (en comparación con $r_{x_1,y_2}$, $r_{x_2,y_2}$ y $r_{x_3,y_2}$), así en la ecuación para $y_2$ en la página 41, los coeficientes de $X_4$ y $X_6$ son negativos y de magnitud mayor a los coeficientes de $X_1$, $X_2$ y $X_3$. Como $r_{x_5,y_2}$ es positivo y su magnitud (en comparación con la de $r_{x_1,y_2}$, $r_{x_2,y_2}$ y $r_{x_3,y_2}$) es grande

(1) Su magnitud es grande

observamos que en la ecuación para $y_2$ en la pagina 41 el coeficiente para $X_5$ es positivo y su magnitud es mayor que la de los coeficientes de $X_1, X_2$ y $X_3$.

Se confirma entonces que $Y_2$ queda explicada por la diferencia entre el ancho del borde superior del billete y la suma de: el ancho del borde inferior y la longitud de la diagonal en los billetes.

| | | $r_{X_iY_1}$ | $r_{X_iY_2}$ | $r^2_{X_iY_1}+r^2_{X_iY_2}$ |
|---|---|---|---|---|
| longitud | $X_1$ | $-0.201$ | $0.028$ | $0.041$ |
| ancho izquierda | $X_2$ | $0.538$ | $0.191$ | $0.326$ |
| ancho derecho | $X_3$ | $0.597$ | $0.159$ | $0.381$ |
| borde inf | $X_4$ | $0.921$ | $-0.377$ | $0.991$ |
| borde sup. | $X_5$ | $0.435$ | $0.794$ | $0.820$ |
| diagonal | $X_6$ | $-0.870$ | $-0.410$ | $0.926$ |

El porcentaje de la varianza de $X_1$, $X_2$ y $X_3$ explicado por $Y_1$ y $Y_2$ es pequeño (las correlaciones $r_{X_iY_1}$ y $r_{X_iY_2}$ $i=1,2,3$ no tienen magnitud grande. De aquí que los coeficientes de $x_1$, $x_2$ y $x_3$ en las ecuaciones para $y_1$ y $y_2$ de la página 41 no son grandes (su magnitud no es grande) comparados con los coeficientes de $x_4$, $x_5$ y $x_6$. Estas observaciones explican la figura G (panel superior a la izquierda), nosotros ya habíamos observado que $X_6$ muestra dos subconjuntos (verdaderos ó falsos) en los billetes pero además en la figura D[1] se hace aparente que también $X_5$ es importamte. Pero la discusión de arriba nos dice que $y_1$ y $y_2$ dependen de $X_4$, $X_5$ y $X_6$ (no tanto de $X_1$, $X_2$ y $X_3$)

(1) Página 19 de las notas sobre métodos gráficos. Diagramas de dispersión.

## LA DISTRIBUCION WISHART

La distribucion Wishart juega un papel importante en la estimación de matrices de covarianza

**DEFINICIÓN:**
Sea $V$ una matriz de dimensiones $p \times p$, simétrica y positivo definida. Considerando $V$ aleatoria, decimos que $V$ tiene la distribución Wishart (no-singular, p-dimensional) con matriz de escala $\Sigma$ y $n$ grados de libertad ($p \leq n$), si la distribución conjunta de los elementos (distintos) de $V$ es continua y con densidad dada por:

$$(Wi) \cdots \quad f(V; \Sigma, n) = \frac{c\,|V|^{(n-p-1)/2}}{|\Sigma|^{n/2}} e^{-\frac{1}{2}\operatorname{tr}\{\Sigma^{-1} V\}} \times \mathbb{1}_B(V),$$

donde: $\Sigma > 0$, $B = \{A \in \mathbb{R}^{p \times p} : A > 0\}$,

$$\mathbb{1}_B(V) = \begin{cases} 1 & V \in B, \\ 0 & V \notin B. \end{cases}$$

y $c$ es la constante

definida por

$$C = \left\{ 2^{np/2} \pi^{\frac{1}{4}(p(p-1))} \prod_{j=1}^{p} \Gamma\left(\frac{n+1-j}{2}\right)\right\}^{-1}.$$

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx .$$

NOTACION

$\mathbb{V} \sim \text{Wishart}_p(\Sigma, n)$

Teorema: Sean $\mathbb{X}_1, \ldots, \mathbb{X}_n$ vectores independientes de dimensión $p \times 1$ tales que $\mathbb{X}_i \sim N_p(\mathbb{0}, \Sigma)$ $\forall i = 1, 2, \ldots, n$, con $\Sigma > 0$. Sea $\mathcal{X}$ la matriz de dimensiones $n \times p$ dada por

$$\mathcal{X} = \begin{pmatrix} \mathbb{X}_1' \\ \mathbb{X}_2' \\ \vdots \\ \mathbb{X}_n' \end{pmatrix} \quad ; \quad p \leq n .$$

Sea $\mathbb{V} = \mathcal{X}'\mathcal{X}$. Entonces $\mathbb{V} > 0$ y $\mathbb{V} \sim \text{Wishart}_p(\Sigma, n)$.

La demostración se puede encontrar en Cramer, H. (1946) "Mathematical Methods of Statistics" p. 405, Princeton University Press.

ejemplo: Supóngase $p=1$, entonces tenemos v.a.

$$X_1, X_2, \ldots, X_n \underset{i.i.d.}{\sim} N_1(0, \sigma^2)$$

$$\underset{n\times 1}{\mathbb{X}} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad \underset{1\times 1}{\mathbb{X}'\mathbb{X}} = \sum_{i=1}^{n} X_i^2$$

y de acuerdo al Teorema si $V = \sum_{i=1}^{n} X_i^2$, su densidad

es
$$f_V(v; \sigma, n) = \underbrace{\frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})}}_{c} \frac{v^{\frac{n-2}{2}}}{\sigma^{n/2}} e^{-\frac{1}{2} v/\sigma} \mathbb{1}_{(0,\infty)}(v)$$

$$= \frac{\left(\frac{1}{2\sigma}\right)^{n/2}}{\Gamma(n/2)} v^{n/2-1} e^{-v/2\sigma} \cdot \mathbb{1}_{(0,\infty)}(v)$$

es decir $V$ tiene distribución

$$\text{Gamma}\left(\alpha = \frac{n}{2};\ \beta = \frac{1}{2\sigma}\right) \left(= \text{Wishart}_1(\sigma, n)\right)$$

"La distribución $\text{Wishart}_p(\Sigma, n)$ representa una generalización de la distribución Gamma a $p$ dimensiones".

De hecho si $X \sim \chi^2_{(n)}$ entonces $V$ tiene la misma distribución que $\sigma^2 X$, entonces podemos re-escribir el párrafo anterior como

"La distribución $Wishart_p(\Sigma, n)$ representa una generalización a $p$ dimensiones de distribuciones proporcionales a la distribución $\chi^2_{(n)}$".
$\left( p=1 \Rightarrow \frac{V}{\sigma^2} \sim \chi^2_{(n)} \right)$

NOTA Aunque hablamos de la distribución de una matriz en la definición (Wi), nos referimos a la distribución conjunta de los elementos de la matriz simétrica $V$

$$(\beta) \cdots \quad V = \begin{pmatrix} v_{11} & v_{21} & \cdots & v_{p1} \\ v_{21} & v_{22} & \cdots & v_{p2} \\ \vdots & \vdots & & \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix},$$

pero como $V$ es simétrica, en realidad se esta definiendo una distribución para el vector $p(p+1)/2$-dimensional

$$(\alpha) \cdots \quad (v_{11}, v_{21}, \ldots, v_{p1}, v_{22}, v_{32}, \ldots, v_{p2}, \ldots, v_{pp})',$$

el conjunto B en la definición (Wi), es el conjunto de matrices <u>simétricas</u> de dimensión $p \times p$ tales que son positivo definidas, pero debido a esta observación, podemos escribir $\mathcal{B}$ como el conjunto de vectores $p(p+1)/2$ dimensionales (como en $(\alpha)$) que hacen que $V$ en $(\beta)$ sea positivo definida.

El teorema establece que si las observaciones $x_1, \dots, x_n$ provienen de una distribución $N_p(0, \Sigma)$, $\mathrm{COV}(X_i, X_j) = \mathbb{E}(X_i \cdot X_j)$ ya que $\mu = 0$, entonces

$$n\hat{\Sigma} = \mathcal{X}'\mathcal{X}$$

es una observación ó realización de la v.a. matricial

$$\mathcal{X}'\mathcal{X} \sim \mathrm{Wishart}_p(\Sigma, n)$$

Teorema: Sean $\mathbb{X}_1, \dots, \mathbb{X}_n$ vectores de dimensión $p \times 1$, independientes y con distribución $N_p(\mu, \Sigma)$; $\Sigma > 0$. Sean $\overline{\mathbb{X}}_{(n)}$ el vector

$$\overline{\mathbb{X}}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{X}_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} X_{i2} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ip} \end{pmatrix} = \begin{pmatrix} \overline{X}_{\cdot 1} \\ \overline{X}_{\cdot 2} \\ \vdots \\ \overline{X}_{\cdot p} \end{pmatrix}$$

y
$$\begin{aligned}
\mathbb{V} &= \sum_{i=1}^{n}\left(\mathbb{X}_i - \overline{\mathbb{X}}_{(n)}\right)\left(\mathbb{X}_i - \overline{\mathbb{X}}_{(n)}\right)' \\
&= \sum_{i=1}^{n} \mathbb{X}_i \mathbb{X}_i' - n\, \overline{\mathbb{X}}_{(n)} \overline{\mathbb{X}}_{(n)}' \\
&\overset{(1)}{=} \mathcal{X}'\mathcal{X} - \frac{1}{n}\mathcal{X}'\mathbb{1}_n\mathbb{1}_n'\mathcal{X} \\
&= \mathcal{X}'\left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\mathcal{X} \\
&= \mathcal{X}'\mathcal{H}\mathcal{X}.
\end{aligned}$$

Entonces si $p+1 \leq n$,

$$\mathbb{V} > 0 \quad \text{y} \quad \mathbb{V} \sim \text{Wishart}_p(\Sigma, n-1)^{(2)}.$$

Como consecuencia de este resultado, si $\hat{\Sigma}$ es

(1) $$\overline{\mathbb{X}}_{(n)} = \frac{1}{n}\mathcal{X}'\mathbb{1}_n = \frac{1}{n}\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & & & \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

(2) Véase Press, J. (2005). "Applied Multivariate Analysis". Dover

la matriz de covarianzas estimada a partir de los datos $x_1, \dots, x_n$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}_{(n)})(x_i - \bar{x}_{(n)})'$$

$$= \frac{1}{n}\mathcal{X}'\mathcal{H}\mathcal{X},$$

entonces $n\hat{\Sigma}$ es una realización de la v.a. matricial $\mathcal{X}'\mathcal{H}\mathcal{X} \sim \text{Wishart}_p(\Sigma, n-1)$.

Recordemos que para n pequeña suele utilizarse el estimador insesgado (corrección de Bessel) para $\Sigma$

$$\hat{\Sigma}^{(c)} = \frac{n}{n-1}\hat{\Sigma} = \frac{1}{n-1}\mathcal{X}'\mathcal{H}\mathcal{X}$$

De la discusión anterior

$(n-1)\hat{\Sigma}^{(c)}$ es una realización de

$$\mathcal{X}'\mathcal{H}\mathcal{X} \sim \text{Wishart}_p(\Sigma, n-1).$$

Por ejemplo si $p=1$ $\hat{\Sigma}^{(c)} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ $(1\times1)$

$$(n-1)\hat{\Sigma}^{(c)} \sim \text{Wishart}_1(\Sigma, n-1)$$

ó bien $\frac{(n-1)\hat{\Sigma}^{(c)}}{\Sigma} \sim \chi^2_{(n-1)}$

Teorema (Transformaciones de Estadísticas asintóticamente Normales) TEAN

Si $\sqrt{m}\left(\mathbb{T}_m - \mu\right) \xrightarrow[m\to\infty]{d} N_p(\mathbb{0}, \Sigma)$

y si $\forall\, i = 1, 2, \dots, s$ $\quad f_i : \mathbb{R}^p \longrightarrow \mathbb{R}$ definen

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_s \end{pmatrix} : \mathbb{R}^p \longrightarrow \mathbb{R}^s$$

una transformación diferenciable en $\mu \in \mathbb{R}^p$, entonces $f(\mathbb{T}_m)$ es asintóticamente Normal con media $f(\mu)$ y matriz de covarianzas $\mathbb{D}'\Sigma\mathbb{D}$:

$$\sqrt{m}\left(f(\mathbb{T}_m) - f(\mu)\right) \xrightarrow[m\to\infty]{d} N_s\left(\mathbb{0}, \mathbb{D}'\Sigma\mathbb{D}\right),$$

donde

$$\mathbb{D}_{ij} = \frac{\partial f_j}{\partial T_i}(\mathbb{T})\Bigg|_{\mathbb{T}=\mu} \qquad \begin{array}{l} i = 1, 2, \dots, p \\ j = 1, 2, \dots, s \end{array}$$

$\mathbb{D}$ es la matriz, de dimensiones $p \times s$, de derivadas parciales

Teorema (distribución asintótica de los valores propios muestrales $l_1, \dots, l_p$ de $\hat{\boldsymbol{\Sigma}}$) DAVP

Sea $\boldsymbol{\Sigma} > 0$ con valores propios diferentes

Sea $n\mathbf{U} \sim \text{Wishart}_p(\boldsymbol{\Sigma}, m = n-1)$ y

supóngase que las descomposiciones de Jordan para $\boldsymbol{\Sigma}$ y $\mathbf{U}$ son $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}'$ y $\mathbf{U} = \mathbf{G}\mathbf{L}\mathbf{G}'$. Entonces

$$\sqrt{n-1}\,(\boldsymbol{l} - \boldsymbol{\lambda}) \xrightarrow[n\to\infty]{d} N_p(\mathbf{0}, 2\boldsymbol{\Delta}^2)$$

donde $\boldsymbol{l} = (l_1, \dots, l_p)'$ y $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)'$ son las diagonales de $\mathbf{L}$ y $\boldsymbol{\Delta}$.

Sea $s=1$ y $f = f_1 : \mathbb{R}^p \to \mathbb{R}$

dada por

$$f(T_1, \dots, T_p) = \frac{T_1 + \cdots + T_q}{T_1 + \cdots + T_p}, \quad (q \leq p).$$

si $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{pmatrix}$, $f(\lambda_1, \dots, \lambda_p) = f(\lambda)$

$= \psi_q$ y para

$l = \begin{pmatrix} l_1 \\ \vdots \\ l_p \end{pmatrix}$, $f(l) = \hat{\psi}_q$.

Para cada $i = 1, 2, \dots, q$

$$\frac{\partial f(T)}{\partial T_i} = \frac{(T_1 + \cdots + T_p) \cdot 1 - (T_1 + \cdots + T_q) \cdot 1}{(T_1 + \cdots + T_p)^2}$$

$$= \frac{1 - \dfrac{T_1 + \cdots + T_q}{T_1 + \cdots + T_p}}{(T_1 + \cdots + T_p)}.$$

y para cada $i = q+1, q+2, \dots, p$

$$\frac{\partial f(T)}{\partial T_i} = \frac{(T_1 + \cdots + T_p) \cdot 0 - (T_1 + \cdots + T_q) \cdot 1}{(T_1 + \cdots + T_p)^2}$$

$$= \frac{-(T_1 + \cdots + T_q)}{(T_1 + \cdots + T_p)^2}.$$

nota si $i=1,2,\dots,q$ y como $\mu=\lambda$

$$\left.\frac{\partial f(\pi)}{\partial \pi_i}\right|_{\pi=\lambda} = \frac{1-\psi_q}{tr(\Sigma)}.$$

Por otra parte, si $i=q+1,\dots,p$

$$\left.\frac{\partial f(\pi)}{\partial \pi_i}\right|_{\pi=\lambda} = \frac{-\psi_q}{tr(\Sigma)}.$$

Considerando que $\Sigma = 2\Lambda^2 = 2\begin{pmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \lambda_p^2 \end{pmatrix}$

$$D' = \left.\left(\frac{\partial f(\pi)}{\partial \pi_1}, \dots, \frac{\partial f(\pi)}{\partial \pi_p}\right)\right|_{\pi=\lambda}$$

$$= \Big(\underset{1}{\frac{1-\psi_q}{tr(\Sigma)}}, \dots, \underset{q}{\frac{1-\psi_q}{tr(\Sigma)}}, \underset{q+1}{-\frac{\psi_q}{tr(\Sigma)}}, \dots, \underset{p}{-\frac{\psi_q}{tr(\Sigma)}}\Big)$$

$$= \frac{1}{tr(\Sigma)}\left(1-\psi_q, \dots, 1-\psi_q, -\psi_q, \dots, -\psi_q\right).$$

De forma que $D'\Sigma D =$

$$\frac{2}{(tr(\Sigma))^2}\left(1-\psi_q, \dots, 1-\psi_q, -\psi_q, \dots, -\psi_q\right)\begin{pmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \lambda_p^2 \end{pmatrix}\begin{pmatrix} 1-\psi_q \\ \vdots \\ 1-\psi_q \\ -\psi_q \\ \vdots \\ -\psi_q \end{pmatrix}$$

$$D' \Sigma D = \left[(1-\psi_q)^2 \lambda_1^2 + \cdots + (1-\psi_q)^2 \lambda_q^2 + \psi_q^2 \lambda_{q+1}^2 + \cdots + \psi_q^2 \lambda_p^2\right] \times \frac{2}{(tr(\Sigma))^2}$$

$$= \frac{2}{(tr(\Sigma))^2} \times \left[ (1-\psi_q)^2 (\lambda_1^2 + \cdots + \lambda_q^2) + \psi_q^2 (\lambda_{q+1}^2 + \cdots + \lambda_p^2) \right]$$

$$= \frac{2}{(tr(\Sigma))^2} \times \left[ \psi_q^2 (\lambda_1^2 + \cdots + \lambda_p^2) - 2\psi_q (\lambda_1^2 + \cdots + \lambda_q^2) + (\lambda_1^2 + \cdots + \lambda_q^2) \right].$$

Sea $\beta = \dfrac{\lambda_1^2 + \cdots + \lambda_q^2}{\lambda_1^2 + \cdots + \lambda_p^2}$, $\quad \lambda_1^2 + \cdots + \lambda_q^2 = \beta(\lambda_1^2 + \cdots + \lambda_p^2)$,

entonces

$$D' \Sigma D = \frac{2(\lambda_1^2 + \cdots + \lambda_p^2)}{(tr(\Sigma))^2} \times \left[ \psi_q^2 - 2\psi_q \beta + \beta \right].$$

Así, para encontrar la distribución asintótica de $\hat{\psi}_q = f(l_1, \ldots, l_p)$, usamos los teoremas DAUP y TEAN para obtener el siguiente

TEOREMA (Distribución Asintótica de $\hat{\psi}_q$)

Para $q \leq p$, Sean $\psi_q = \dfrac{\lambda_1 + \cdots + \lambda_q}{\lambda_1 + \cdots + \lambda_p}$ y $\hat{\psi}_q = \dfrac{l_1 + \cdots + l_q}{l_1 + \cdots + l_p}$

La proporción de la variabilidad de los datos explicada por las primeras q componentes principales y su estimación muestral. Entonces, cuando el tamaño de muestra n crece a $+\infty$.

$$\sqrt{n-1}\left(\hat{\psi}_q - \psi_q\right) \xrightarrow{d} N_1\left(0, \omega^2\right), \quad \cdots (u)$$

donde

$$\omega^2 = \frac{2(\lambda_1^2 + \cdots + \lambda_p^2)}{\left(tr\left(\Sigma\right)\right)^2} \times \left\{ \psi_q^2 - 2\psi_q \beta + \beta \right\}$$

$$y \quad \beta = \frac{\lambda_1^2 + \cdots + \lambda_q^2}{\lambda_1^2 + \cdots + \lambda_p^2}.$$

Con este resultado, podemos construir un intervalo de confianza a nivel $(1-\alpha)\times 100\%$ para el verdadero valor de la proporción $\psi_q$.

Usando $(u)$; si n es suficientemente grande

$$\mathbb{P}\left( \left| \frac{\hat{\psi}_q - \psi_q}{\omega/\sqrt{n-1}} \right| \leq Z_{1-\alpha/2} \right) \overset{\text{aprox}}{\approx} 1-\alpha, \quad \cdots (v)$$

donde $\alpha \in (0,1)$ es un tamaño de error (nivel de significancia) y $Z_{1-\alpha/2}$ es el cuantil de una distribución $N(0,1)$ correspondiente a probabilidad $1-\alpha/2$:

$$\int_{-\infty}^{z_{1-\alpha/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx = 1-\alpha/2 .$$

$$1-\alpha \overset{\text{aprox}}{\approx} \mathbb{P}\left(\left|\frac{\hat{\psi}_q-\psi_q}{w/\sqrt{n-1}}\right| \leq z_{1-\alpha/2}\right)$$

$$= \mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\psi}_q-\psi_q}{w/\sqrt{n-1}} \leq z_{1-\alpha/2}\right)$$

$$= \mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\psi_q-\hat{\psi}_q}{w/\sqrt{n-1}} \leq z_{1-\alpha/2}\right)$$

$$= \mathbb{P}\left(\hat{\psi}_q - w/\sqrt{n-1}\times z_{1-\alpha/2} \leq \psi_q \leq \hat{\psi}_q + w/\sqrt{n-1}\times z_{1-\alpha/2}\right).$$

Entonces, a nivel $(1-\alpha)\times 100\%$ de confianza $\psi_q$ debe estar en el intervalo

$$\left(\hat{\psi}_q - w/\sqrt{n-1}\times z_{1-\alpha/2} ,\ \hat{\psi}_q + w/\sqrt{n-1}\times z_{1-\alpha/2}\right).$$

cuantil de una distribución $N(0,1)$ correspondiente a probabilidad $1-\alpha/2$:

$$\int_{-\infty}^{Z_{1-\alpha/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = 1-\alpha/2 .$$

$$1-\alpha \overset{\text{aprox}}{\approx} \mathbb{P}\left( \left| \frac{\hat{\psi}_q - \psi_q}{w/\sqrt{n-1}} \right| \leq Z_{1-\alpha/2} \right)$$

$$= \mathbb{P}\left( -Z_{1-\alpha/2} \leq \frac{\hat{\psi}_q - \psi_q}{w/\sqrt{n-1}} \leq Z_{1-\alpha/2} \right)$$

$$= \mathbb{P}\left( -Z_{1-\alpha/2} \leq \frac{\psi_q - \hat{\psi}_q}{w/\sqrt{n-1}} \leq Z_{1-\alpha/2} \right)$$

$$= \mathbb{P}\left( \hat{\psi}_q - w/\sqrt{n-1} \times Z_{1-\alpha/2} \leq \psi_q \leq \hat{\psi}_q + w/\sqrt{n-1} \times Z_{1-\alpha/2} \right).$$

Entonces, a nivel $(1-\alpha)\times 100\%$ de confianza $\psi_q$ debe estar en el intervalo

$$\left( \hat{\psi}_q - w/\sqrt{n-1} \times Z_{1-\alpha/2} \,,\, \hat{\psi}_q + w/\sqrt{n-1} \times Z_{1-\alpha/2} \right).$$

Para los datos de billetes del banco suizo, sabemos que $\hat{\psi}_1 = \dfrac{l_1}{l_1+l_2+\cdots+l_6} = 0.67$ .

Supóngase que quisieramos probar la hipótesis (h).... $\begin{cases} H_0 : \psi_1 = 0.75 \\ \text{vs.} \\ H_1 : \text{No } H_0 \end{cases}$

Una forma de llevar a cabo la prueba de hipótesis es usando intervalos de confianza

$$\hat{\beta} = \frac{l_1^2}{l_1^2 + \cdots + l_6^2} = \frac{(2.985)^2}{(2.985)^2 + (0.931)^2 + \cdots + (0.035)^2}$$

$$= 0.902 .$$

$$\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_6^2 = tr(\Lambda^2) = tr(\Lambda^2 \Gamma' \Gamma)$$
$$= tr(\Gamma \Lambda^2 \Gamma')$$
$$= tr(\Gamma \Lambda \Gamma' \Gamma \Lambda \Gamma') = tr(\Sigma \Sigma)$$
$$= tr(\Sigma^2)$$

Se puede estimar con $tr(\hat{\Sigma}^2) = \sum_{j=1}^{6} l_j^2$

$$= 9.883$$

$tr(\Sigma)$ se puede estimar con $tr(\hat{\Sigma})$

$$= tr(\mathcal{G} \mathcal{L} \mathcal{G}') = tr(\mathcal{L} \mathcal{G}' \mathcal{G}) = tr(\mathcal{L})$$

$$= l_1 + \cdots + l_6 = 4.472$$

para $q=1$

De manera que $w$ en la ecuación $(u)$, página 63 se puede estimar por

$$\hat{w}^2 = \frac{2\,tr(\hat{\Sigma}^2)}{[tr(\hat{\Sigma})]^2}\left(\hat{\psi}_1^2 - 2\hat{\psi}_1\hat{\beta} + \hat{\beta}\right)$$

$$= \frac{2(9.883)}{(4.472)^2}\left[(0.668)^2 - 2(0.668)(0.902) + 0.902\right]$$

$$= 0.142 .$$

Entonces, a nivel $0.95 \times 100\%$ $(\alpha = 0.05)$ de significancia un intervalo (aproximado) para $\psi_1$ es (véase página 64)

$$\left(0.668 - \sqrt{\frac{0.142}{199}} \times 1.96\,,\ 0.668 + \sqrt{\frac{0.142}{199}} \times 1.96\right)$$

$$= (0.615, 0.720),$$

el cual no contiene el valor de $0.75$, por tanto rechazamos Ho en (h).

# ANALISIS DE COMPONENTES PRINCIPALES NORMALIZADO

Como se mencionó antes, las componentes principales son sensibles a cambios de escala en las variables, de esta forma al re-escalar alguna de las variables bajo estudio los resultados pueden cambiar.

Por otra parte, uno de los objetivos en el análisis de componentes principales es encontrar las direcciones en $\mathbb{R}^p$, para las cuales los datos presentan mayor variabilidad. Sin embargo si los datos originales presentan heterogeneidad con respecto a sus varianzas, por ejemplo si la escala de la variable $X_i$ es en kilogramos y la escala de $X_j$ este en dólares, entonces "la dirección en $\mathbb{R}^p$ para la cual los datos presentan mayor variabilidad" podría originarse por la heterogeneidad en las escalas y no debido a la naturaleza de los datos.

Para evitar estos problemas se utiliza una estandarización de las variables en el vector $\mathcal{X}_G$, a saber