"la proporción de la varianza de X_i explicada por la j-ésima componente principal Z_j".

<u>ejercicio</u>: pruebe la ecuación (cor npc).

La figura F muestra gráficas de las componentes principales calculadas a partir de los datos estandarizados. Para esta figura se usa el caracter "o" para los billetes verdaderos y el caracter "+" para los billetes falsos.

El vector de valores propios de $\hat{R}$ es

$$l^{\hat{R}} = (2.946, 1.278, 0.869, 0.450, 0.269, 0.189)'$$

El panel inferior izquierdo en la figura F muestra una gráfica de las parejas ordenadas $(i, l_i^{\hat{R}})$; $i=1,2,\dots,6$.

Podemos calcular la proporción de la varianza explicada por la componente i, al dividir $l_i^{\hat{R}}$ por $\sum_{j=1}^{p} l_j^{\hat{R}} = p$

$$p^{\hat{R}} \equiv \left(\frac{2.946}{6}, \frac{1.278}{6}, \frac{0.869}{6}, \frac{0.450}{6}, \frac{0.269}{6}, \frac{0.189}{6} \right)'$$

$$= (0.491, 0.213, 0.145, 0.075, 0.045, 0.032).$$

Si $\psi_q^{\hat{R}} = \dfrac{\sum_{j=1}^{q} l_j^{\hat{R}}}{\sum_{j=1}^{6} l_j^{\hat{R}}}$ es la proporción de la variabilidad explicada por las primeras q componentes (normalizadas), entonces el vector $\psi^{\hat{R}} = (\psi_1^{\hat{R}}, \psi_2^{\hat{R}}, \ldots, \psi_6^{\hat{R}})$

es $\psi^{\hat{R}} = (0.491, 0.704, 0.849, 0.924, 0.969, 1.0)$

Entonces, la primera componente principal (normalizada) explica 49% de la variabilidad y las primeras dos componentes explican 70.4% de la varibilidad en los datos

La matriz $\mathcal{G}_{\hat{R}}$ de vectores propios está dada por

$$
G_{\vec{R}} = \begin{pmatrix}
-0.007 & 0.815 & -0.018 & 0.575 & -0.059 & -0.031 \\
0.468 & 0.342 & 0.103 & -0.395 & 0.639 & 0.298 \\
0.487 & 0.252 & 0.123 & -0.430 & -0.614 & -0.349 \\
0.407 & -0.266 & 0.584 & 0.404 & -0.215 & 0.462 \\
0.368 & -0.091 & -0.788 & 0.110 & -0.220 & 0.419 \\
-0.493 & 0.274 & 0.114 & -0.392 & -0.340 & 0.632
\end{pmatrix}
$$

$$
Z_1 = -0.007x_1 + 0.468x_2 + 0.487x_3 + 0.407x_4 + 0.368x_5 - 0.493x_6
$$

$$
Z_2 = 0.815x_1 + 0.342x_2 + 0.252x_3 - 0.266x_4 - 0.091x_5 + 0.274x_6
$$

Notemos que los pesos asignados a cada variable en estas combinaciones lineales son ahora de "magnitud similar".

La figura 6 muestra las parejas ordenadas $(r_{x_i z_1}, r_{x_i z_2})$; $i=1,2,\dots,p$, los valores de estas correlaciones están en la siguiente tabla

variable	$r_{x_i z_1}$	$r_{x_i z_2}$	$r_{x_i z_1}^2 + r_{x_i z_2}^2$
X_1 longitud	-0.012	0.922	0.850
X_2 ancho (izq.)	0.803	0.387	0.794
X_3 ancho (der.)	0.835	0.285	0.78
X_4 long. borde inferior	0.698	-0.301	0.58
X_5 long. borde superior	0.631	-0.104	0.41
X_6 long. diag	-0.847	0.309	0.81

Observando la ecuación para Z_1 (los magnitudes y signos de los pesos), así como la columna $r_{x_i z_1}$ de la tabla, notamos que Z_1 queda descrita por la diferencia entre: la suma de los anchos (izquierdo y derecho) y la longitud de la diagonal de los billetes. Usando la ecuación para Z_2 y la columna $r_{x_i z_2}$, notamos que Z_2 está dominada por la longitud de los billetes.

Ahora nos concentremos en la figura F, tomando los anteriores observaciones en cuenta. Vemos que para los billetes verdaderos (con caracter "0") se tienen longitudes de su diagonal que son más grandes y valores de longitud del ancho (izquierdo y derecho) más pequeños(1). Para los billetes falsos, se tienen valores de la longitud de la diagonal más pequeños y valores de longitud del ancho (izquierdo y derecho) más grandes(2)

Con respecto a la longitud de los billetes X_1, esta tiende a ser más grande para los billetes verdaderos(3)

(1) $Z_1 < 0$ (2) $Z_1 > 0$ (3) $Z_2 > 0$