(Tarea 2) Estadística Aplicada III

David Isaac López Romero
CU: 173993

1) Tenemos las siguientes variables con las siguientes varianzas:

Variable	Varianza
Falta de agua	4.35
Electricidad	3.25
Número de Habitantes	3.54
Número de Escuelas	6.58
Número de Hospitales	4.26
Pavimentación	4.65
Tamaño de la localidad	5.23

Adicionalmente, tenemos los coeficientes de los primeros dos componentes:

Variable	Coef. Primer Componente	Coef. Segundo Componente
Falta de agua	−0.37876	−0.23703
Electricidad	0.389513	0.182809
Número de Habitantes	0.093304	0.700643
Número de Escuelas	0.171092	0.321193
Número de Hospitales	0.255164	0.550858
Pavimentación	0.325639	0.302078
Tamaño de la Localidad	0.076763	0.648485

1a) Suponiendo que el primer eigenvalor es 20 y el segundo 6, encuentra las covarianzas y correlaciones de todas las variables con los dos primeros componentes.

Tenemos que

$X_1 =$ Falta de agua $\qquad \sigma_1^2 = 4.35$

$X_2 =$ Electricidad $\qquad \sigma_2^2 = 3.25$

$X_3 =$ Número de Habitantes $\qquad \sigma_3^2 = 3.54$

$X_4 =$ Número de Escuelas $\qquad \sigma_4^2 = 6.58$

$X_5 =$ Número de Hospitales $\qquad \sigma_5^2 = 4.26$

$X_6 =$ Pavimentación $\qquad \sigma_6^2 = 4.65$

$X_7 =$ Tamaño de la Localidad $\qquad \sigma_7^2 = 5.23$

Denotamos a

$$a_1 = \begin{pmatrix} -0{,}37876 \\ 0{,}384513 \\ 0{,}093304 \\ 0.171042 \\ 0{,}255164 \\ 0.325639 \\ 0{,}076763 \end{pmatrix} \quad y \quad a_2 = \begin{pmatrix} -0{,}23703 \\ 0{,}182804 \\ 0{,}700643 \\ 0{,}321143 \\ 0.558858 \\ 0{,}382078 \\ 0{,}648495 \end{pmatrix}$$

Usando las fórmulas $Cov(X_i, Y_j) = \lambda_j a_{ji}$, $Corr(X_i, Y_j) = \frac{\sqrt{\lambda_j}\, a_{ji}}{\sigma_i}$

Primer componente

Covarianza	Correlación
$Cov(X_1, Y_1) = -7{,}5752$	$Corr(X_1, Y_1) = -0{,}812146$
$Cov(X_2, Y_1) = 7{,}79026$	$Corr(X_2, Y_1) = 0{,}9662628$
$Cov(X_3, Y_1) = 1{,}86608$	$Corr(X_3, Y_1) = 0{,}2217755$
$Cov(X_4, Y_1) = 3{,}42184$	$Corr(X_4, Y_1) = 0{,}2982852$
$Cov(X_5, Y_1) = 5{,}10328$	$Corr(X_5, Y_1) = 0{,}552878$
$Cov(X_6, Y_1) = 6{,}51278$	$Corr(X_6, Y_1) = 0{,}6753438$
$Cov(X_7, Y_1) = 1{,}53526$	$Corr(X_7, Y_1) = 0{,}150122$

Segundo Componente

Covarianza	Correlación
$Cov(X_1, Y_2) = -1{,}42218$	$Corr(X_1, Y_2) = -0{,}2703776$
$Cov(X_2, Y_2) = 1{,}096854$	$Corr(X_2, Y_2) = 0{,}2483885$
$Cov(X_3, Y_2) = 4{,}203858$	$Corr(X_3, Y_2) = 0{,}9121595$
$Cov(X_4, Y_2) = 1{,}927158$	$Corr(X_4, Y_2) = 0{,}3067106$
$Cov(X_5, Y_2) = 3{,}353148$	$Corr(X_5, Y_2) = 0{,}66 32 42424$
$Cov(X_6, Y_2) = 2{,}292468$	$Corr(X_6, Y_2) = 0{,}43401144$
$Cov(X_7, Y_2) = 3{,}89091$	$Corr(X_7, Y_2) = 0.694587$

1b) En base de las correlaciones, describe qué es el componente principal 1 y componente principal 2. (3)

En el componente uno, las variables que tienen mayor correlación con la componente, en términos absolutos, son

1) Falta de agua
2) Electricidad
3) Pavimentación
4) Número de hospitales

Así, el componente 1 se refiere a servicios básicos.

En el componente dos, las variables con mayor correlación con la componente, en valor absoluto, son

1) Número de habitantes
2) Número de hospitales
3) Tamaño de la localidad

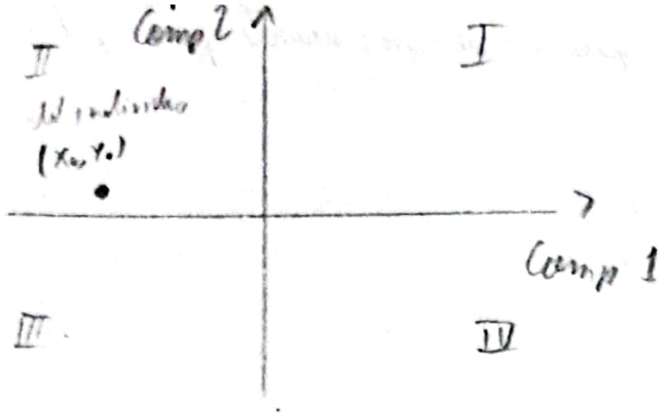
Así, el componente 2 se refiere a Factores vivienda.

1c) i) Grafica en un plano cartesiano los primeros dos componentes principales y describe cada uno de los cuadrantes. ii) Si tenemos un individuo con los siguientes datos en las variables.

Sol ii)

Falta de agua	5
Electricidad	-1
Número de Habitantes	1.2
Número de Escuelas	1
Número de Hospitales	0
Pavimentación	-2
Tamaño de la Localidad	1.5

Las variables tienen media cero, es por eso, que hay valores negativos, encuentra los valores de los primeros dos componentes principales y úbicalo en el plano cartesiano.

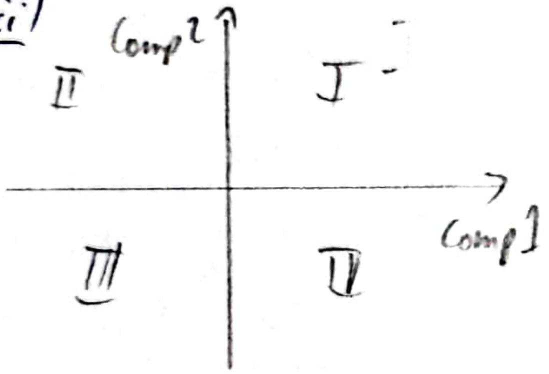


Comp 1 = -2.5363897

Comp 2 = $2.5771 \times 10^{-3} = 0.0025771$

El individuo tiene mejores condiciones de vivienda (componente 2) pero muy malos servicios básicos (I)

Sol ci)



I) Buenos servicios básicos
Buenos factores vivienda

II) Malos servicios básicos
Buenos factores vivienda

III) Malos servicios básicos
Malos factores vivienda

IV) Buenos servicios básicos
Malos factores vivienda

1d) ¿Qué porcentaje explican cada uno de los componentes principales y los primeros dos acumulados? Varianza total = 31.86

	λ_j	λ_j/Var tot	% explicado	acumulado	% acumulado
Comp. 1	20	0.62775	62.775%	20	62.775%
Comp 2	6	0.18832	18.832%	26	81.607%

2) Tenemos 3 variables estandarizadas con la siguiente matriz de varianzas y covarianzas.

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

Encuentra los componentes principales y menciona cuánto explica cada uno porcentualmente y acumulado.

Vemos que las columnas 2 y 3 se generan a partir de la primera por lo que el rango de la matriz es uno. Solo tiene un eigenvalor distinto de cero. Vemos que ese eigenvalor es 3 pues

$$\left(\begin{array}{ccc|c} 1-3 & 1 & -1 & 0 \\ 1 & 1-3 & -1 & 0 \\ -1 & -1 & 1-3 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} -2 & 1 & -1 & 0 \\ 1 & -2 & -1 & 0 \\ -1 & -1 & -2 & 0 \end{array}\right) \underset{R_2+R_3}{\overset{2R_2+R_1}{\sim}} \left(\begin{array}{ccc|c} 0 & -3 & -3 & 0 \\ 1 & -2 & -1 & 0 \\ 0 & -3 & -3 & 0 \end{array}\right)$$

$$\underset{\substack{-R_2+R_3 \\ -\frac{1}{3}R_2}}{\overset{R_1 \leftrightarrow R_2}{\sim}} \left(\begin{array}{ccc|c} 1 & -2 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \overset{2R_2+R_1}{\sim} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \therefore \begin{array}{l} x = -t \\ y = -t \, , \ t \in \mathbb{R} \\ z = t \end{array}$$

Normalizando tenemos que

$$x^2+y^2+z^2=1 \Leftrightarrow (-t)^2+(-t)^2+t^2=1 \Leftrightarrow 3t^2=1 \Leftrightarrow t=\pm\frac{1}{\sqrt{3}}$$

$$\text{Así,} \quad \left. x=\mp\frac{1}{\sqrt{3}}, \ y=\mp\frac{1}{\sqrt{3}}, \ z=\pm\frac{1}{\sqrt{3}} \right\} = \text{Com } 1$$

Entonces, el componente 1 con eigenvalor 3 explica el 100% de la variabilidad.

Los otros dos componentes no aportan nada, explican el 0% por ser base del espacio nulo de la matriz. Esto es que

Comp 2 : $x = \frac{1}{\sqrt{2}}$, $y = 0$, $z = \frac{1}{\sqrt{2}}$

Comp 3 : $x = -\frac{1}{\sqrt{6}}$, $y = \frac{2}{\sqrt{6}}$, $z = \frac{1}{\sqrt{6}}$

% Explicado Comp 1 = 100%
||
% Explicado Acumulado

Así el acumulado es [illegible]

3) Escribe las siguientes funciones de densidad en forma de familia exponencial, encuentra su esperanza y varianza.

a) $f_x(x) = (1-p)^{x-1} p$

$$f(x) = (1-p)^{x-1} p = e^{\ln\{(1-p)^{x-1} p\}} = e^{(x-1)\ln(1-p) + \ln(p)} \cdots$$

$$= e^{x\ln(1-p) - \ln(1-p) + \ln(p)} = e^{x\ln(1-p) - \ln\left(\frac{1-p}{p}\right)}$$

donde $h(x) \equiv 1$, $\eta(p) = \ln(1-p)$, $t(x) = x$, $A(p) = \ln\left(\frac{1-p}{p}\right)$

$(= \mathbb{E}(T(x)))$

$$\mathbb{E}(X) = \frac{A'(p)}{\eta'(p)} = \frac{\left[\ln\left(\frac{1-p}{p}\right)\right]'}{[\ln(1-p)]'} = \frac{\frac{1}{\left(\frac{1-p}{p}\right)}\left\{\frac{-1p - 1(1-p)}{p^2}\right\}}{\frac{1}{1-p}(-1)}$$

$$= \frac{\frac{\cancel{p}}{1-p}\left\{\frac{-p-1+p}{p^{\cancel{2}}}\right\}}{-\left(\frac{1}{1-p}\right)} = -\frac{1}{p} = \frac{1}{p}$$

$$\underset{\text{Var}(T(x))}{\text{Var}(x)} = \frac{A''(p) - \eta''(p)\,\mathbb{E}(T(x))}{[\eta'(p)]^2} = \frac{\left[\ln\left(\frac{1-p}{p}\right)\right]'' - [\ln(1-p)]''\frac{1}{p}}{([\ln(1-p)]')^2}$$

$$= \frac{\left[\frac{1}{\left(\frac{1-p}{p}\right)}\left\{\frac{-\cancel{1}p - 1(1-p)}{p^{\cancel{2}}}\right\}\right]' - \left(\frac{1}{1-p}(-1)\right)'\frac{1}{p}}{\left(\frac{1}{1-p}(-1)\right)^2} = \frac{\left(\frac{1}{(p-1)p}\right)' - \left(\frac{1}{p-1}\right)'\frac{1}{p}}{\left(\frac{1}{(1-p)^2}\right)}$$

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

$$= \frac{1-2p}{p^2} + \frac{1}{p} = \frac{1-p}{p^2}$$

b) $f_X(x) = \binom{x+r-1}{x} (1-p)^{x-r} p^r$

$f_X(x) = \binom{x+r-1}{x} (1-p)^{x-r} p^r = \exp\{\ln\binom{x+r-1}{x} + (x+r)\ln(1-p) + r\ln(p)\}$

$= \exp\{\ln\left(\frac{(x+r-1)!}{x!\,(r-1)!}\right) + (x+r)\ln(1-p) + r\ln(p)\}$

$= \exp\{x\ln(1-p) + r\ln(p(1-p)) + \ln((x+r-1)!) - \ln(x!) - \ln((r-1)!)\}$

$= \frac{1}{x!}\exp\{x\ln((1-p)(x+r-1)!) - \{\ln\left(\frac{1}{p^r(1-p)^r}\right) + \ln\left(\frac{1}{(r-1)!}\right)\}\}$

donde $h(x) = \frac{1}{x!}$, $\eta(\theta) = \ln(1-p)(x+r-1)!$, $t(x) = x$,

$A(\theta) = \ln\left(\frac{1}{p^r(1-p)^r}\right) + \ln\left(\frac{1}{(r-1)!}\right)$

$\mathbb{E}(X) = \frac{r(1-p)}{p}$, $Var(X) = \frac{r(1-p)}{p^2}$

c) $f_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$

$f_X(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} = \exp\{\ln(\beta^\alpha) - \beta x\} \frac{x^{\alpha-1}}{\Gamma(\alpha)}$

$= \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{\{-\beta x - \ln\left(\frac{1}{\beta^\alpha}\right)\}}$

donde $h(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}$, $\eta(\beta) = -\beta$, $t(x) = x$, $A(\beta) = \ln\left(\frac{1}{\beta^\alpha}\right)$

$\mathbb{E}(X) = \frac{A'(\beta)}{\eta'(\beta)} = \frac{-\frac{1}{\beta^\alpha}\alpha\beta^{\alpha-1}}{-1} = \frac{\alpha}{\beta}$

$Var(X) = \frac{-\frac{\alpha}{\beta^2} - 0}{\cdot} = \frac{\alpha}{\beta^2}$ //

4) Escribe la siguiente densidad como miembro de la familia exponencial. Halla esperanza y varianza de $\ln(x)$.

$$f_x(x) = \frac{1}{\sqrt{2\pi}\,x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

$$f_x(x) = \frac{1}{\sqrt{2\pi}\,x} e^{-\frac{\ln^2(x)}{2\sigma^2}} e^{\frac{2\mu\ln(x)}{2\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi}\,x} e\left\{\ln^2(x)\frac{1}{2\sigma^2}\right\} e\left\{\frac{\mu\ln(x)}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\}$$

$$h(x) = \frac{1}{\sqrt{2\pi}\,x} e^{-\frac{\ln^2(x)}{2\sigma^2}}, \quad \eta(\mu) = \frac{\mu}{\sigma^2}, \quad t(x) = \ln(x), \quad A(\mu) = \frac{\mu^2}{2\sigma^2}$$

$$\mathbb{E}(\ln(x)) = \frac{A'(\mu)}{\eta'(\mu)} = \frac{\left(\frac{2\mu}{2\sigma^2}\right)}{\left(\frac{1}{\sigma^2}\right)} = \mu$$

$$Var(\ln(x)) = \frac{A''(\mu) - \eta''(\mu)\,\mathbb{E}(\ln(x))}{(\eta'(\mu))^2}$$

$$= \frac{\frac{1}{\sigma^2} - 0(\mu)}{\left(\frac{1}{\sigma^2}\right)^2} = \frac{\left(\frac{1}{\sigma^2}\right)}{\left(\frac{1}{\sigma^4}\right)} = \sigma^2$$

5) Modelo logístico para un candidato:

$$\ln\left(\frac{p_i}{1-p_i}\right) = 0.02 + 0.43\, Ing_i + 1.02\, M_i + 0.42\, Med_i + 0.92\, Sup_i$$

Ing_i = Ingreso, $M_i = \begin{cases} 1 & \text{mujer} \\ 0 & \text{c.o.c} \end{cases}$, $Med_i = \begin{cases} 1 & \text{clase media} \\ 0 & \text{c.o.c.} \end{cases}$, $Sup_i = \begin{cases} 1 & \text{clase alta} \\ 0 & \text{c.o.c.} \end{cases}$

a) Interpreta los parámetros.

p_i = probabilidad base con un ingreso cero

Entonces para lo que tenemos usamos el momio

$$\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = e^{0.02 + 0.43\,Ingi + 1.02\,Mi + 0.42\,Medi + 0.42\,Supi}$$

- Si aumenta el ingreso $\dfrac{\left(\frac{\hat{p}_i'}{1-\hat{p}_i'}\right)}{\left(\frac{p_i}{1-p_i}\right)} = ⊛$

$$\frac{\left(\frac{\hat{p}_i'}{1-\hat{p}_i'}\right)}{\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)} = e^{0.43} \Rightarrow (⊛ - 1) \times 100\% = 53.72\%$$

Por cada peso de incremento en ingreso, el momio sube 53.72%

- Si es mujer

$$\frac{\left(\frac{\hat{p}_i'}{1-\hat{p}_i'}\right)}{\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)} = e^{1.02} \Rightarrow (⊛ - 1)\, 100\% = 177.31\%$$

Si es mujer, aumenta en 177.31 % el momio

- Si clase media

$$\frac{\left(\frac{\hat{p}_i'}{1-\hat{p}_i'}\right)}{\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)} = e^{0.42} \Rightarrow (⊛ - 1)\, 100\% = 52.196\%$$

Si es clase media el momio sube en 52.196%

- Si clase alta

$$\frac{\left(\frac{\hat{p}_i'}{1-\hat{p}_i'}\right)}{\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)} = e^{0.42} \Rightarrow (⊛ - 1)\, 100\% = 150.93\%$$

Si es clase alta el momio aumenta 150.93%

5b) Si intervalos al 95% son.

Exponencial del coeficiente		
Ordenada al origen	0.97	1.02
Ingreso	0.93	2.53
Mujer	2.27	3.38
Nivel Medio	1.34	2.71
Nivel Alto (Superior)	1.85	3.38

¿Qué parámetros son significativos al 95%

Vemos que $1 \in IC$ (ordenada al origen)

$1 \in IC$ (Ingreso)

entonces los parámetros ordenada al origen e ingreso no son significativos

Como $1 \notin IC$ (Mujer)

$1 \notin IC$ (Nivel Medio)

$1 \notin IC$ (Nivel Superior)

los parámetros mujer, nivel medio y nivel superior sí son significativos.