

Universidad de Chile

Facultad de Ciencias Físicas y Matemáticas

Departamento de Ingeniería Matemática

Auxiliar #11 MA34B

Profesor: Julio Deride, Auxiliar: Gonzalo Contador.

P1. Se realizó un estudio de alfabetización digital en 10 países, en los cuales se recolectaron los datos de: Área(km²), población(millones), tasa de desempleo y cantidad de computadores por persona. Los resultados se presentan resumidos en la tabla

	<i>Media</i>	<i>D.est.</i>
<i>Area</i>	1368597	2763092,6
<i>Pob</i>	53,5	83,49
<i>Desempleo</i>	17,96	18,25
<i>Computadores</i>	1,5	2,19

Los valores y vectores propios de la matriz de correlaciones son

	λ_1	λ_2	λ_3	λ_4
<i>valor</i>	2,753	0,874	0,353	0,02
<i>componente</i>	1	2	3	4
<i>Area</i>	0,517	0,393	-0,595	-0,473
<i>Pob</i>	0,584	0,296	0,047	0,775
<i>Desempleo</i>	-0,318	0,881	0,346	-0,05
<i>Computadores</i>	0,539	-0,113	0,794	-0,415

a) ¿Como se obtienen los valores y vectores propios presentados anteriormente a partir de la muestra? Obtenga las primeras dos componentes principales a partir de ellos y explicita el porcentaje de la varianza total asociado a ellas.

b) Dibuje el círculo de correlaciones e interprete la correlación entre las cuatro variables estudiadas a partir de ellas. ¿Es una buena aproximación?

c) Se quiere agregar la variable "Índice de Estabilidad Política" al modelo, cuya correlación con las primeras componentes principales es de 0,08 y 0,24 respectivamente. Si la desviación estándar observada para este dato fue de 1,922, determine su posición en el círculo de correlaciones. ¿Permite una buena interpretación el círculo de su correlación con las otras variables? ¿Se puede subestimar la correlación real? ¿Se puede sobreestimar?

d) Determine las contribuciones de cada variable original sobre cada variable original para la primera componente principal.

e) Se tiene la información de 4 nuevos países. Sobre el plano determinado por las dos primeras componentes principales, determine su ubicación e interprete

el cuadrante obtenido en cada uno de ellos.

<i>Pais</i>	<i>Area</i>	<i>Pob</i>	<i>Desempleo</i>	<i>PC</i>
<i>EEUU</i>	9629091	278,1	4	5,4
<i>Namibia</i>	825418	1,8	35	0,03
<i>Chile</i>	756950	15,3	8,1	0,56
<i>Francia</i>	547930	59,6	9	0,89

f) ¿Cuál es el mínimo porcentaje de la varianza representada por las dos primeras componentes principales sobre n variables? ¿Qué sucede si ese porcentaje se alcanza? ¿Es de interés el ACP en tal caso?