

Introducción al Analisis Multivariado

Daniela Pico Arredondo, Juan Sebastián Falcón

13/3/2021

Introducción al analisis multivariado - Tarea #1

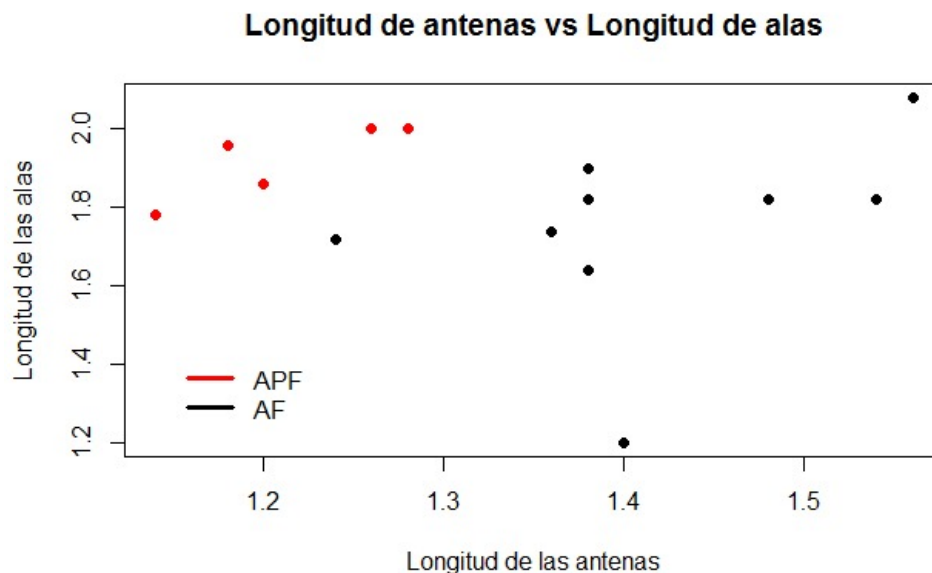
Profesor: Raúl Alberto Pérez Agaméz

Los Biologos Grojan y Wirth (1981) descubrieron dos nuevas especies de insectos Ameroheleafasciata (AF) y Apseudofasciata (APF). Puesto que las especies son similares en apariencia, resulta útil para el biólogo estar en capacidad de clasificar un espécimen como AF o APF basado en características externas que son fáciles de medir. Entre alguna de las características que distinguen los AF de los APF, los biólogos reportan medidas de la longitud de las antenas (X) y la longitud de las alas (Y), ambas medidas en milímetros, de nueve insectos AF y de seis insectos APF. Los datos son los siguientes:

ESPECIE	APF	APF	APF	APF	APF
X	1.14	1.20	1.18	1.26	1.28
Y	1.78	1.86	1.96	2.00	2.00

ESPECIE	AF	AF	AF	AF	AF	AF	AF	AF	AF
X	1.38	1.40	1.24	1.36	1.38	1.48	1.54	1.38	1.56
Y	1.64	1.20	1.72	1.74	1.82	1.82	1.82	1.90	2.08

1. Construya un gráfico de dispersión X e Y . Comente acerca de la apariencia de los datos.



- se observa que:

Debido a que no hay suficientes observaciones, la nube de puntos no ofrece mucha información respecto al comportamiento de cada grupo, sin embargo, se observa una leve tendencia lineal positiva para ambos grupos, esto quiere decir que por cada aumento en la longitud de las antenas hay un pequeño aumento en la longitud de las alas de cada especie, además se ve un outlier en el grupo AF.

2. Para cada grupo de especies (Grupo-1 y Grupo-2), calcule el vector de medias muestrales, la matriz de varianzas-covarianzas muestrales, la matriz de correlaciones muestrales, la varianza total y la varianza generalizada de los datos.

Vector de Medias para la especie AF

	x
X	1.413
Y	1.749

Con esto se obtiene que el vector de medias para las variables Longitud de antenas y Longitud de alas para la categoría AF es: $\bar{X}=(1,413, 1,749)$

Vector de Medias para la especie APF

	x
X	1.212
Y	1.920

Con esto se obtiene que el vector de medias para las variables Longitud de antenas y Longitud de alas para la categoría APF es: $\bar{Y}=(1,212, 1,920)$

Matriz de varianzas-Covarianzas Muestrales para AF

	X	Y
X	0.010	0.009
Y	0.009	0.058

Matriz de Varianzas-Covarianzas Muestrales APF

	X	Y
X	0.003	0.005
Y	0.005	0.009

Se obtienen dos matrices correspondientes a la especie AF y APF respectivamente, se observa que efectivamente es simétrica y que en la diagonal se encuentran las varianzas de cada variable.

Matriz de Correlaciones Muestrales para AF

	X	Y
X	1.000	0.375
Y	0.375	1.000

Matriz de Correlaciones Muestrales para APF

Se observa en la matriz de correlaciones para la especie AF que las variables longitud de alas y longitud de antenas tienen una baja correlación, concretamente del 0.375, lo que indica una baja dependencia lineal entre estas variables para dicha especie. Por otro lado, en la matriz de correlaciones para la especie APF se

	X	Y
X	1.000	0.841
Y	0.841	1.000

observa un coeficiente de correlación del 0.841, lo cual indica una alta dependencia lineal positiva entre las variables longitud de antenas y longitud de alas.

Varianza Total para AF

x
0.06751

Varianza Total para APF

x
0.01272

Se obtiene la varianza total sumando los elementos de la diagonal de cada matriz de varianzas covarianzas, esta no es una medida de variabilidad muy marcada puesto que no tiene en cuenta dependencia entre las variables.

Varianza Generalizada de los datos para AF

x
0.00049

Varianza Generalizada de los datos de APF

x
1e-05

Luego se obtiene la varianza generalizada para analizar una medida mas precisa de la variabilidad. Sabemos que a medida que el determinante se acerca mas a cero indica una mayor dependencia lineal entre las variables, como se puede observar al comparar ambas varianzas de cada especie y obtener una menor varianza en la especie APF donde se encontro anteriormente una dependencia lineal positiva

3. Para cada grupo de especies (Grupo-1 y Grupo-2), calcule la distancia euclidiana y la distancia de Mahalanobis entre cada especie y los respectivos vectores de medias muestrales.

Distancia de MahalanobiS para AF

Table 1: Distancia Mahalanobis

0.238	5.827	3.4	0.322	0.321	0.456	1.676	0.777	2.983
-------	-------	-----	-------	-------	-------	-------	-------	-------

Distancia de Mahalanobis para APF

Table 2: Distancia Mahalanobis

2.089	0.717	2.958	0.747	1.489
-------	-------	-------	-------	-------

Distancia euclidiana para AF y APF respectivamente

Table 3: Distancia Euclidean

1	2	3	4	5	6	7	8	9
0.000	0.440	0.161	0.102	0.180	0.206	0.241	0.260	0.475
0.440	0.000	0.544	0.541	0.620	0.625	0.636	0.700	0.894
0.161	0.544	0.000	0.122	0.172	0.260	0.316	0.228	0.482
0.102	0.541	0.122	0.000	0.082	0.144	0.197	0.161	0.394
0.180	0.620	0.172	0.082	0.000	0.100	0.160	0.080	0.316
0.206	0.625	0.260	0.144	0.100	0.000	0.060	0.128	0.272
0.241	0.636	0.316	0.197	0.160	0.060	0.000	0.179	0.261
0.260	0.700	0.228	0.161	0.080	0.128	0.179	0.000	0.255
0.475	0.894	0.482	0.394	0.316	0.272	0.261	0.255	0.000

Table 4: Distancia Euclidean

1	2	3	4	5
0.000	0.100	0.184	0.251	0.261
0.100	0.000	0.102	0.152	0.161
0.184	0.102	0.000	0.089	0.108
0.251	0.152	0.089	0.000	0.020
0.261	0.161	0.108	0.020	0.000

4. ¿Considera razonable usar la distancia de Mahalanobis en cada uno de los dos grupos? ¿Es posible usar distancia euclidean en este caso? ¿porqué si o porque no?

De las especies se obtuvo que existe una alta correlación (son dependientes) en el grupo APF, aunque en la especie AF no existe una correlación (son independientes), Ambas especies estan sujetas a fluctuaciones aleatorias ya que son medidas através de magnitudes (longitud de antenas y longitud de alas), esto indica que es preferible ponderarlas deacuerdo a su variabilidad, Por lo tanto se recomienda usar la distancia de Mahalanobis ya que esta tiene en cuenta la variabilidad y la dependencia entre las variables.