

Rincón de la Bioestadística

Representación gráfica de datos multivariantes

Gabriel Cavada Ch.^{1,2}

¹Facultad de Medicina, Universidad de los Andes.

²División de Bioestadística, Escuela de Salud Pública, Universidad de Chile.

Graphical representation of multivariate data

Generalmente la información que se recolecta es “multivariante”, es decir, cada sujeto o unidad de observación está caracterizada por más de un atributo que ha sido medido en dicho sujeto, esto es lo que se llama en lenguaje técnico que cada sujeto es un “vector p-variante”, así el i-ésimo sujeto es representado por un objeto p-dimensional de la siguiente forma:

$$\vec{X} = (X_1, X_2, \dots, X_p)$$

Donde cada X_i representa una de las características del sujeto.

Por ejemplo, si estamos interesados en caracterizar una muestra de oficiales de aviación, que a su vez está compuesta por pilotos e ingenieros y cada sujetos está caracterizado por 6 variables a saber:

clase	0: ingenieros 1: pilotos
y1	inteligencia
y2	relación formal
y3	dinamismo
y4	agudeza
y5	coordinación censorsa motriz
y6	perseveración

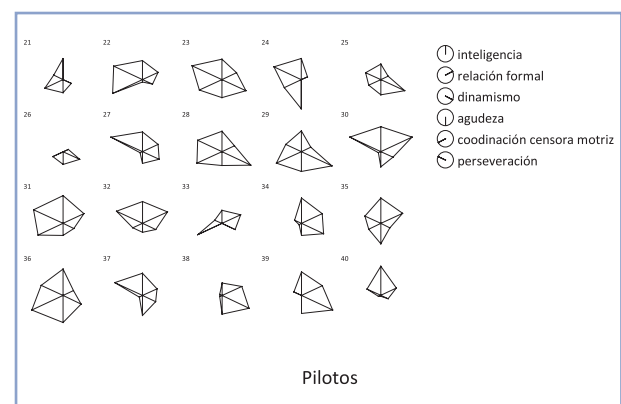
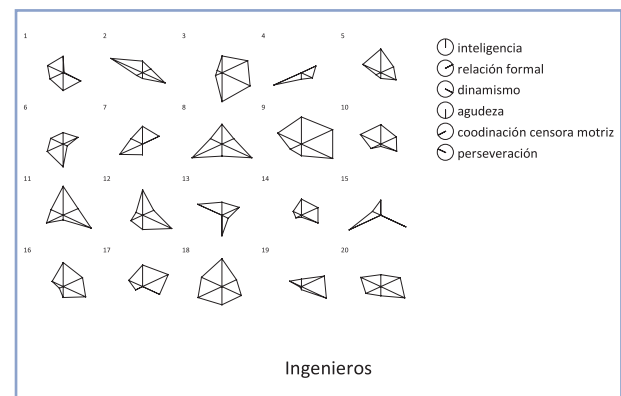
Cuyos datos se pueden observar al final del texto.

Cada sujeto está caracterizado por 6 variables y la clase a que pertenece, esto es, cada sujeto es representado en un vector 6-dimensional.

Un desafío interesante es hacer una representación gráfica en que simultáneamente se expresen las 6 variables, con el fin de establecer la taxonomía entre ingenieros y pilotos así ¿Cómo será la confluencia de ellas?

Una primera representación es hacer un gráfico de estrellas, que consiste en representar cada variable sobre cada uno de los radios que se pueden dibujar en un hexágono regular (dado que son 6 las componentes del vector de caracterización), sin embargo, cada radio será prolongado o disminuido de acuerdo a lo medido en cada variable y el fin de estas prolongaciones unidas por una recta, así se obtienen polígonos de 6 lados para cada sujeto que parecen estrellas, ahora la

idea es observar si las estrellas de los ingenieros son distintas a la de los pilotos. Este modo de representación permite hacer los siguientes gráficos:

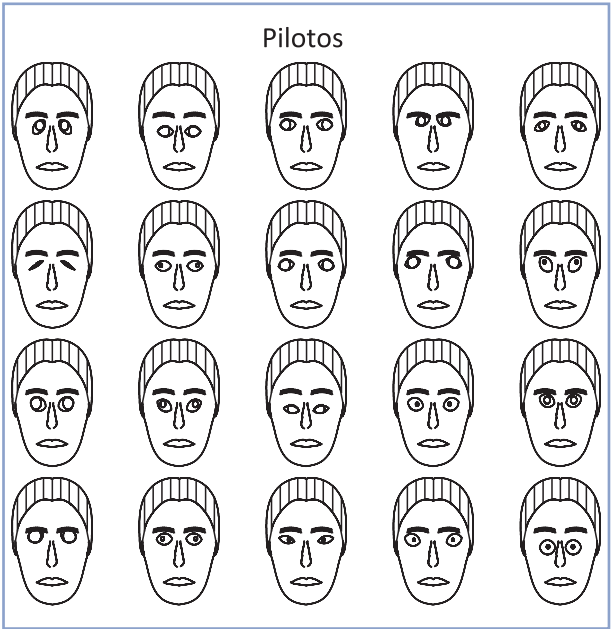
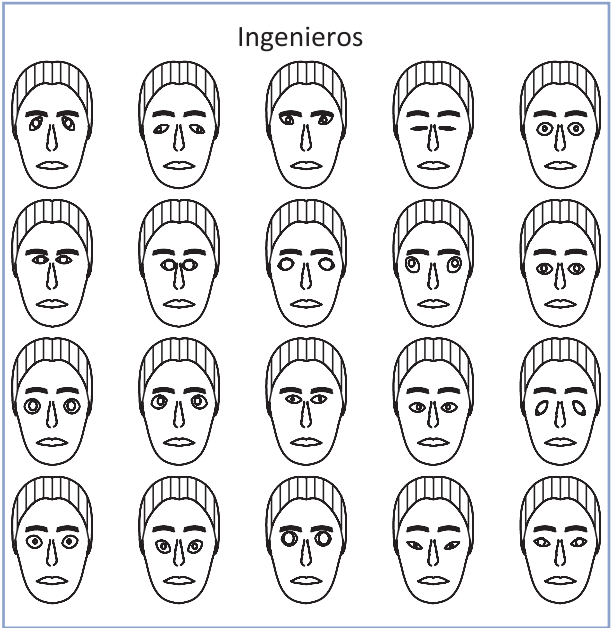


Con cierto ojo entrenado podrían observarse las diferencias más notables entre ambos grupos de estrellas, o incluso observar sujetos atípicos dentro de cada grupo.

Otra propuesta para representar información multivariante, fue propuesta por Chernoff en el año 1980, que consiste en asociar cada variable a distancias definidas en un rostro humano, tales como: tamaño de los ojos, ángulo de los ojos, tamaño de las pupilas, ángulo de las cejas y todas las características morfológicas identificables en el rostro huma-

Rincón de la Bioestadística

no. Este tipo de representación sólo fue posible difundirla en los últimos tiempos en concordancia con el desarrollo de los programas computacionales. Los datos de ingenieros y pilotos se ven como sigue:



El principal objetivo de este artículo fue mostrar técnicas de representación gráfica poco conocidas en bioestadística o medicina, que sin embargo, con un poco de innovación y creatividad podrían ser bastante atractivas.

Datos

clase	y1	y2	y3	y4	y5	y6
0	121	22	74	223	54	254
0	108	30	80	175	40	300
0	122	49	87	266	41	223
0	77	37	66	178	80	209
0	140	35	71	175	38	261
0	108	37	57	241	59	245
0	124	39	52	194	72	242
0	130	34	89	200	85	242
0	149	55	91	198	50	277
0	129	38	72	162	47	268
0	154	37	87	170	60	244
0	145	33	88	208	51	228
0	112	40	60	232	29	279
0	120	39	73	159	39	233
0	118	21	83	152	88	233
0	141	42	80	195	36	241
0	135	49	73	152	42	249
0	151	37	76	223	74	268
0	97	46	83	164	31	243
0	109	42	82	188	57	267
1	132	17	77	232	50	249
1	123	32	79	192	64	315
1	129	31	96	250	55	319
1	131	23	67	291	48	310
1	110	24	96	239	42	268
1	47	22	87	231	40	217
1	125	32	87	227	30	324
1	129	29	102	234	58	300
1	130	26	100	256	58	270
1	147	47	82	240	30	322
1	159	37	80	227	58	317
1	135	41	83	216	39	306
1	100	35	83	183	57	242
1	149	37	94	227	30	240
1	149	38	78	258	42	271
1	153	27	89	283	66	291
1	136	31	83	257	31	311
1	97	36	100	252	30	225
1	141	37	110	250	27	243
1	164	32	76	187	30	264