



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TAREA 2

EJERCICIOS

1. Considera los datos en el fichero `sons.dat`, en el lugar habitual, que se reproducen a continuación. Dan para una muestra de 25 familias las longitudes y anchuras de cabeza del primer y segundo hijos, respectivamente. Haciendo uso de los mismos, responde a las siguientes cuestiones:
 - a) Haz un gráfico Q-Q para cada una de las cuatro variables para examinar su ajuste a la distribución normal. Completa tu impresión subjetiva mediante un contraste de normalidad (se sugiere el de D'Agostino; la tabla de valores críticos del estadístico D en [5], está reproducida en [8], Tabla A.4, p. 492).
 - b) Estima el vector de medias, matriz de covarianzas y matriz de correlación.
 - c) (*optativo*) Haz un diagnóstico de normalidad multivariante de las cuatro variables consideradas conjuntamente (por ejemplo, el de Gnanadesikan y Kettenring). Emplea la tabulación de Barnett y Lewis para decidir si alguna observación se separa anormalmente de las demás respecto a lo que cabría esperar bajo el supuesto de normalidad multivariante.
 - d) Contrasta el paralelismo entre los subvectores de medias del primer y segundo hijo. Expresa en palabras la conjetura que este contraste podría ayudar a establecer o refutar.
 - e) Contrasta la igualdad de medias del primer y segundo hijo de cada familia (Ayuda: *NO* se trata de dos muestras independientes: son 25 parejas de hermanos.)

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

1. Al margen de tus apuntes de clase puedes mirar las secciones relevantes de cualquiera de los muchos manuales a tu disposición en Biblioteca (clasificados en 519.237, segunda planta). En particular puedes consultar [4], [10], [8]. También [1] y [7].

En algunos de estos textos encontrarás también reproducidas las tablas en [5] y [2].
2. Las funciones `qqnorm` y `qqplot` aparecen documentadas en [3] y en la documentación "on line". Adicionalmente dispones de la función `shapiro.test`, que realiza una adaptación del contraste de Shapiro-Wilk apta para tamaños muestrales bastante grandes (ver [9] y las referencias que cita).

Cuadro 1: Medidas de cabeza del primer y segundo hijos de una muestra de 25 familias.

Primer hijo		Segundo hijo		Primer hijo		Segundo hijo	
X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
191	155	179	145	174	150	185	152
195	149	201	152	190	159	195	157
181	148	185	149	188	151	187	158
183	153	188	149	163	137	161	130
176	144	171	142	195	155	183	158
208	157	192	152	186	153	173	148
189	150	190	149	181	145	182	146
197	159	189	152	175	140	165	137
188	152	197	159	192	154	185	152
192	150	187	151	174	143	178	147
179	158	186	148	176	139	176	143
183	147	174	147	197	167	200	158
—	—	—	—	190	163	187	150

Fuente: Tomados de [8] (Tabla 3.9, p. 91), que a su vez los toma de [6].

Referencias

- [1] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 1984 edition, 1978. Signatura: 519.237 AND.
- [2] V. Barnett and T. Lewis. *Outliers in Statistical Data*. Wiley, New York, 1978.
- [3] R.A. Becker, J.M. Chambers, and A.R. Wilks. *The New S Language. A Programming Environment for Data Analysis and Graphics*. Wadsworth & Brooks/Cole, Pacific Grove, California, 1988.
- [4] C.M. Cuadras. *Métodos de Análisis Multivariante*. Eunibar, Barcelona, 1981.
- [5] R.B. D'Agostino. Small sample probability points for the D test of normality. *Biometrika*, 59:219–221, 1972.
- [6] G.P. Frets. Heredity of head form in man. *Genetica*, 3:193–384, 1921.
- [7] A. Kshirsagar. *Multivariate Analysis*. Marcel Dekker, 1978.
- [8] A.C. Rencher. *Methods of Multivariate Analysis*. Wiley, 1995.
- [9] P. Royston. A remark on algorithm as 181: The W test for normality. *Journal of the Royal Statistical Soc., Ser. B*, 44:547–551, 1995.
- [10] G.A.F. Seber. *Multivariate Observations*. Wiley, New York, 1984.