



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TAREA 9

EJERCICIOS

1. Se han ensayado cuatro diferentes métodos de enseñar a realizar sumas y multiplicaciones “con llevadas” (sobre números de dos dígitos). Cada uno de los cuatro métodos se ha utilizado sobre un colectivo de 20 niños diferentes. Los datos de los 80 niños están disponibles en el fichero `aritm.dat`. Para cada niño se da la “nota” en sumas y en multiplicaciones. ¿Hay evidencia de que algunos métodos den mejor resultado que otros?
2. La Tabla 1 recoge datos¹ correspondientes a cuatro medidas de cráneos hallados en necrópolis de diferentes épocas en Egipto, de acuerdo con la siguiente clave:

Variable	Descripción
MB	Maxima anchura del cráneo
BH	Altura basibregmática del cráneo.
BL	Altura basalveolar del cráneo
NH	Altura nasal
Año	Año aproximado de formación del cráneo

Se trata de contrastar si las medidas medias difieren significativamente de un periodo a otro, o si por el contrario puede aceptarse que son las mismas. El interés de esto radica en que, si hubieran cambiado, habría cierta evidencia de migraciones de otros pueblos en el área, con la consiguiente mezcla racial².

- a) Describe los datos. Señala las peculiaridades que observes.
- b) ¿Parece razonable suponer que la matriz de covarianzas de las observaciones es común en todos los periodos?
- c) Haz un análisis multivariante de varianza (MANOVA). Muestra tus resultados y coméntalos.

¹En `StatLib`, originalmente procedentes de [10]. Están en el fichero `craneos.dat`.

²Esta es la “historia” que obtuve junto con los datos. Comentarios en el apartado de ayudas y sugerencias más abajo.

Cuadro 1: Medidas de cráneos hallados en Egipto en diferentes periodos.

MB	BH	BL	NH	Año	MB	BH	BL	NH	Año	MB	BH	BL	NH	Año
131	138	89	49	-4000	137	136	106	49	-3300	132	133	90	53	-200
125	131	92	48	-4000	126	131	100	48	-3300	134	134	97	54	-200
131	132	99	50	-4000	135	136	97	52	-3300	135	135	99	50	-200
119	132	96	44	-4000	129	126	91	50	-3300	133	136	95	52	-200
136	143	100	54	-4000	134	139	101	49	-3300	136	130	99	55	-200
138	137	89	56	-4000	131	134	90	53	-3300	134	137	93	52	-200
139	130	108	48	-4000	132	130	104	50	-3300	131	141	99	55	-200
125	136	93	48	-4000	130	132	93	52	-3300	129	135	95	47	-200
131	134	102	51	-4000	135	132	98	54	-3300	136	128	93	54	-200
134	134	99	51	-4000	130	128	101	51	-3300	131	125	88	48	-200
129	138	95	50	-4000	137	141	96	52	-1850	139	130	94	53	-200
134	121	95	53	-4000	129	133	93	47	-1850	144	124	86	50	-200
126	129	109	51	-4000	132	138	87	48	-1850	141	131	97	53	-200
132	136	100	50	-4000	130	134	106	50	-1850	130	131	98	53	-200
141	140	100	51	-4000	134	134	96	45	-1850	133	128	92	51	-200
131	134	97	54	-4000	140	133	98	50	-1850	138	126	97	54	-200
135	137	103	50	-4000	138	138	95	47	-1850	131	142	95	53	-200
132	133	93	53	-4000	136	145	99	55	-1850	136	138	94	55	-200
139	136	96	50	-4000	136	131	92	46	-1850	132	136	92	52	-200
132	131	101	49	-4000	126	136	95	56	-1850	135	130	100	51	-200
126	133	102	51	-4000	137	129	100	53	-1850	137	123	91	50	150
135	135	103	47	-4000	137	139	97	50	-1850	136	131	95	49	150
134	124	93	53	-4000	136	126	101	50	-1850	128	126	91	57	150
128	134	103	50	-4000	137	133	90	49	-1850	130	134	92	52	150
130	130	104	49	-4000	129	142	104	47	-1850	138	127	86	47	150
138	135	100	55	-4000	135	138	102	55	-1850	126	138	101	52	150
128	132	93	53	-4000	129	135	92	50	-1850	136	138	97	58	150
127	129	106	48	-4000	134	125	90	60	-1850	126	126	92	45	150
131	136	114	54	-4000	138	134	96	51	-1850	132	132	99	55	150
124	138	101	46	-4000	136	135	94	53	-1850	139	135	92	54	150
124	138	101	48	-3300	132	130	91	52	-1850	143	120	95	51	150
133	134	97	48	-3300	133	131	100	50	-1850	141	136	101	54	150
138	134	98	45	-3300	138	137	94	51	-1850	135	135	95	56	150
148	129	104	51	-3300	130	127	99	45	-1850	137	134	93	53	150
126	124	95	45	-3300	136	133	91	49	-1850	142	135	96	52	150
135	136	98	52	-3300	134	123	95	52	-1850	139	134	95	47	150
132	145	100	54	-3300	136	137	101	54	-1850	138	125	99	51	150
133	130	102	48	-3300	133	131	96	49	-1850	137	135	96	54	150
131	134	96	50	-3300	138	133	100	55	-1850	133	125	92	50	150
133	125	94	46	-3300	138	133	91	46	-1850	145	129	89	47	150
133	136	103	53	-3300	137	134	107	54	-200	138	136	92	46	150
131	139	98	51	-3300	141	128	95	53	-200	131	129	97	44	150
131	136	99	56	-3300	141	130	87	49	-200	143	126	88	54	150
138	134	98	49	-3300	135	131	99	51	-200	134	124	91	55	150
130	136	104	53	-3300	133	120	91	46	-200	132	127	97	52	150
131	128	98	45	-3300	131	135	90	50	-200	137	125	85	57	150
138	129	107	53	-3300	140	137	94	60	-200	129	128	81	52	150
123	131	101	51	-3300	139	130	90	48	-200	140	135	103	48	150
130	129	105	47	-3300	140	134	90	51	-200	147	129	87	48	150
134	130	93	54	-3300	138	140	100	52	-200	136	133	97	51	150

3. Muestra que en un modelo MANOVA con dos tratamientos y diseño equilibrado las “sumas de cuadrados” correspondientes a los dos efectos (H_A y H_B en la notación que hemos venido empleando) se distribuyen realmente como matrices de Wishart independientes. (Ayuda: Puedes imitar la estrategia del Teorema 1.3 de los apuntes. Puedes también consultar bibliografía y seguir cualquier vía que tengas por conveniente.)

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

1. No desprecies el análisis descriptivo variable a variable y las herramientas gráficas de que dispones (te interesará casi con seguridad emplear funciones como `boxplot`, `stripchart`, `pairs` o `coplot`).
2. Ten en cuenta que los resultados de un análisis MANOVA presuponen normalidad multivariante e igualdad de las matrices de covarianzas dentro de cada grupo. Conoces formas de contrastar ambas cosas.
3. Tienes tablas permitiendo obtener valores críticos de la distribución lambda de Wilks en [9]. (Observa que las tablas referidas dan valores críticos de una transformación del estadístico $|\hat{\Sigma}|/|\hat{\Sigma}_H|$.) No obstante, `summary(resultado)`, en que `resultado` es un objeto de tipo MANOVA, te da todo lo preciso sin necesidad de consultar tablas.

4. Uno puede pensar en términos diferentes a los de la historia obtenida de `Statlib`. Una mezcla racial puede, en efecto, alterar las medidas corporales de la población, y entre ellas las medidas del cráneo. Pero la mejora o empeoramiento de condiciones de vida también podría concebiblemente dar lugar a lo mismo. Si el conocimiento que tenemos del periodo bajo estudio permite descartar este tipo de influencia, entonces un contraste de igualdad de medias sobre los datos brutos parecería lo indicado.

Si, por el contrario, pensáramos que, por ejemplo, las medidas corporales de los sujetos (y las de sus cráneos entre ellas) variaron a lo largo del tiempo, parecería que la diferencia de medias no sería necesariamente evidencia de mezcla racial. Una mezcla racial podría más bien manifestarse en un cambio *de la forma* de las cabezas. Si esta conjetura fuera correcta, podríamos preferir analizar las medidas de los cráneos normalizadas de forma conveniente, de modo que reflejaran no el tamaño absoluto de los cráneos sino sus proporciones.

Puedes hacer el análisis que te apetezca: sobre los datos brutos o normalizados. El objetivo es que te ejercites en realizar un contraste MANOVA y en interpretar su resultado. En la vida real tendrías que hacer todas las preguntas necesarias a tu cliente, asegurarte de qué es lo que realmente le interesa, y traducir sus conjeturas a hipótesis sobre los parámetros en un modelo adecuado. Es lo habitual que tu cliente formule el problema en forma vaga o inconsistente. Una parte no pequeña del cometido de un estadístico consiste en hacer entender a quien le consulta cuál es su problema: una vez bien formulado está siempre más cerca de su solución.

5. Muchos textos de análisis Multivariante te servirán de referencia para entender e interpretar los resultados de MANOVA. Al margen de las notas de clase, puedes ver [7], p. 376, [3], Cap. 20 y 21, [8], Sec. 10.7 o [6], Cap. 11, entre otros muchos.

Referencias

- [1] 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.

-
- [2] J.M. Chambers and T.J. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole, Pacific Grove, Ca., 1992.
- [3] C.M. Cuadras. *Métodos de Análisis Multivariante*. Eunibar, Barcelona, 1981.
- [4] P. Dalgaard. *Introductory statistics with R*. Statistics and Computing. Springer-Verlag, 2002. Signatura: 519.682 DAL.
- [5] W.R. Dillon and M. Goldstein. *Multivariate Analysis: Methods and Applications*. Wiley, New York, 1984.
- [6] D.E. Johnson. *Métodos multivariados aplicados al análisis de datos*. Thomson, 1998.
- [7] W.J. Krzanowski. *Principles of Multivariate Analysis: A User's Perspective*. Oxford, 1988. Signatura: 519.23 KRZ.
- [8] D. Peña. *Análisis de Datos Multivariantes*. McGraw-Hill, 2002.
- [9] G.A.F. Seber. *Multivariate Observations*. Wiley, New York, 1984.
- [10] A. Thomson and R. Randall-Maciver. *Ancient Races of the Thebaid*. Oxford University Press, 1905.
- [11] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-PLUS*. Springer-Verlag, New York, third edition, 1999.