eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

TAREA 9

EJERCICIOS

1. Tres diferentes dietas se ensayan sobre dos razas de cerdos, midiéndose el número de días necesarios en cada caso (a partir del nacimiento) para alcanzar el mismo peso. Se obtienen las observaciones siguientes:

	DIETA A	DIETA B	DIETA B
RAZA I	183	169	173
	185	172	175
	192	175	179
RAZA II	201	169	175
	198	173	175
	212	166	180

a) Escribe la matriz X de regresores asociada al diseño experimental en la tabla proporcionada.

b) Compila la tabla de Análisis de Varianza. Especifica los grados de libertad de cada forma cuadrática. Especifica los supuestos del modelo.

c) Contrasta la hipótesis de que las dos razas de cerdos son igualmente rentables.

d) Contrasta la hipótesis de que las tres dietas son igualmente efectivas (es decir, provocan el engorde a la misma velocidad).

e) ¿Hay alguna raza de cerdo que responda a alguna cierta dieta de modo significativamente diferente a lo que cabría esperar de la consideración separada de los efectos “cerdo” y “dieta” ?

f) Imagina que hubieras tomado una sola observación de cada combinación cerdo-dieta. ¿Podrías efectuar el contraste de interacción entre ambos factores?

g) ¿Qué dietas son significativamente distintas al nivel de significación del 5 %?

2. Consideremos un experimento para medir la pérdida de calor que se tiene cuando se usan 2 tipos de cristal térmico (Opaco y Traslúcido) para ventanas en dos grosores (Grueso y Delgado). Se hacen pruebas para un temperatura interior constante y 5 temperaturas exteriores prefijadas Se prueban tres cristales de cada tipo para cada temperatura exterior y se mide la correspondiente pérdida de calor expresada porcentualmente. Los datos obtenidos (en la *data frame* `cristal.dge`: lee mediante un `dget`) son:

	Tipo de cristal			
	Grueso		Delgado	
Temperatura exterior (°C)	Opaco	Traslúcido	Opaco	Traslúcido
15	2.9	7.2	10.0	12.1
	3.6	8.0	11.2	13.3
	4.0	7.7	10.6	12.6
10	5.4	10.3	12.6	14.7
	6.3	9.8	13.1	14.5
	5.8	9.4	12.7	15.2
5	7.2	10.7	13.1	15.2
	6.7	11.3	14.3	15.8
	6.2	10.4	13.7	15.9
0	7.7	11.7	14.3	17.2
	7.4	12.1	14.7	16.8
	7.9	11.4	15.0	16.1
–5	9.8	13.3	16.8	18.5
	9.6	13.7	16.7	18.9
	9.7	14.2	16.5	18.5

a) Construye la tabla ANOVA para este experimento equilibrado con tres tratamientos, utilizando el modelo adecuado (es decir, aditivo o no aditivo). Si decides usar el modelo aditivo, justifica esta decisión.

b) Contrasta la hipótesis de que los cuatro tipos de cristal tienen igual pérdida de calor.

c) Contrasta la hipótesis de que la pérdida de calor es la misma para las distintos temperaturas.

d) ¿Hay alguna observación anómala (= con residuo studentizado muy grande)?

3. Un contenedor de huevos batidos como los que se emplean en hostelería y hospitales, se agita bien, y su contenido se divide en dos mitades, G y H. Se toman a continuación doce muestras formadas cada una por dos ampollas, una rellenada del contenido de la mitad G y la otra del contenido de la mitad H. El contenido de las dos ampollas que forman cada muestra es por tanto idéntico en composición.

   Las doce muestras así tomadas se remiten a seis laboratorios, dos muestras a cada laboratorio. En cada laboratorio, se entrega una ampolla de cada muestra a uno de dos analistas, para que evalúen su contenido en grasas, y cada analista realiza el análisis de cada ampolla dos veces.

   Si el análisis se pudiera efectuar sin error, los resultados por todos los analistas de todos los laboratorios sobre todas las ampollas de todas las muestras serían idénticos. Pero el análisis tiene un margen de error, por lo que los resultados normalmente no coincidirán. Se trata no obstante de ver si hay alguna pauta sistemática: ¿proporciona algún laboratorio resultados sesgados al alza (o a la baja)? ¿alguno de los técnicos lo hace? Haz el análisis que estimes oportuno, detalla los supuestos y describe el resultado.

   Los datos se reproducen en el Cuadro 1, y proceden en última instancia de [1]. Puedes encontrarlos en forma de *dataframe* en el fichero `eggs.frame`, y en ASCII en el fichero `eggs.dat`.

Cuadro 1: Resultados de análisis de contenido de grasa en huevos en diferentes laboratorios y por diferentes personas

Grasa	Laboratorio	Analista	Ampolla	Grasa	Laboratorio	Analista	Ampolla
.62	I	1	G	.18	IV	1	G
.55	I	1	G	.47	IV	1	G
.34	I	1	H	.53	IV	1	H
.24	I	1	H	.32	IV	1	H
.80	I	2	G	.40	IV	2	G
.68	I	2	G	.37	IV	2	G
.76	I	2	H	.31	IV	2	H
.65	I	2	H	.43	IV	2	H
.30	II	1	G	.35	V	1	G
.40	II	1	G	.39	V	1	G
.33	II	1	H	.37	V	1	H
.43	II	1	H	.33	V	1	H
.39	II	2	G	.42	V	2	G
.40	II	2	G	.36	V	2	G
.29	II	2	H	.20	V	2	H
.18	II	2	H	.41	V	2	H
.46	III	1	G	.37	VI	1	G
.38	III	1	G	.43	VI	1	G
.27	III	1	H	.28	VI	1	H
.37	III	1	H	.36	VI	1	H
.37	III	2	G	.18	VI	2	G
.42	III	2	G	.20	VI	2	G
.45	III	2	H	.26	VI	2	H
.54	III	2	H	.06	VI	2	H

AYUDAS, SUGERENCIAS, COMENTARIOS

a) Haz el primer problema a mano (salvo, quizá, para encontrar los residuos *studentizados*); no lleva mucho tiempo. Puedes luego, si lo deseas, comprobar los resultados con ayuda de R.

b) Mira la documentación de la función `aov`. Junto con otras funciones relacionadas, permite una estimación muy cómoda de todo tipo de modelos. Presta particular atención a la notación que se ha de emplear: `A:B` denota interacción entre `A` y `B`, mientras que `A*B` denota la misma interacción *y todos los efectos inferiores en la jerarquía*. Para especificar un modelo completo puedes por tanto indicar sólo la interacción de orden superior.

4. En el caso de modelos anidados, familiarízate con la notación `B %in% A` (tratamiento `B` anidado en el `A`) y `A/B` (forma corta de escribir `A + B %in% A`).

5. Observa: el modelo ANOVA está identificado sólo mediante restricciones. En clase hemos utilizado restricciones de suma cero sobre los parámetros correspondientes a cada efecto, pero otras son posibles. Si empleas R tal cual, obtendrás resultados que no se corresponden con los que puedes obtener a mano, empleando las fórmulas en las notas de clase. Por omisión de otra especificación, R regresa sobre combinaciones lineales ("contrastes") de las columnas, en el número preciso para no introducir dependencias lineales exactas.

   Así por ejemplo, si el modelo fuera

$$\boldsymbol{Y} = \alpha_0\mathbf{1} + \sum_{i=1}^{4}\alpha_i\boldsymbol{v}_i + \boldsymbol{\epsilon}, \tag{1}$$

R crearía tres columnas combinaciones lineales de $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_4$ generando el mismo espacio que $\sum_{i=1}^{4} \alpha_i \boldsymbol{v}_i$ con los α_i sujetos a la condición $\sum_{i=1}^{4} \alpha_i = 0$.

Si quieres obtener valores de los parámetros idénticos a los que obtienes a mano, basta que especifiques (por una sóla vez al comienzo de la sesión)

```
options(contrasts=c("contr.sum","contr.poly").
```

Puedes mirar el tipo de contrastes que especifican `contr.sum`, `contr.treatment`, etc. mediante la ayuda *on line*. Mira también el ejemplo realizado en clase (colgado en Moodle).

6. Por omisión, R emplea contrastes `contr.treatment` para variables cualitativas no ordenadas: el primer nivel se emplea como caso de referencia. Te puede resultar de utilidad la función `relevel` de R, que permite reordenar los niveles y decidir cuál queremos que forme parte del caso de referencia. No es de interés para los ejercicios más arriba, pero puede serlo en otras ocasiones.

   Observa que, cualquiera que sea la parametrización que empleas, los resultados de los contrastes son idénticos, y la suma de cuadrados explicada por cada efecto también.

7. Tienes una descripción muy completa de todo tipo de modelos ANOVA en [8]. Encontrarás, en particular, una exposición muy detallada del modo de hacer diseño de experimentos. También puedes encontrar de utilidad la primera parte de [4], [6], y el clásico [5]. Los libros [2] y [9] te ayudarán en las cuestiones relacionadas con R.

8. En el problema 1, apartado g, puedes servirte de la función `TukeyHSD` examinada en clase.

Referencias

[1] Bliss. *Statistics in Biology*. Wiley, 1967.

[2] J. J. Faraway. *Linear Models with R*. Chapman & Hall/CRC, 2005. Signatura: 519.233 FAR.

[3] R. H. Myers. *Classical and Modern Regression with Applications*. PWS-KENT Pub. Co., Boston, 1990.

[4] D. Peña. *Regresión y Diseño de Experimentos*. Alianza Editorial, 2002.

[5] H. Scheffé. *The Analysis of Variance*. Wiley, New York, 1959.

[6] G. A. F. Seber. *Linear Regression Analysis*. Wiley, New York, 1977.

[7] J. H. Stapleton. *Linear Statistical Models*. Wiley, New York, 1995.

[8] A. Fdez. Trocóniz. *Modelos Lineales*. Serv. Editorial UPV/EHU, Bilbao, 1987.

[9] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York, third edition, 1999.