



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TAREA 7

EJERCICIOS

Los ejercicios siguientes tienen por objeto suministrar práctica en la aplicación de modelos logarítmico lineales al análisis de tablas de contingencia.

El Centro de Investigaciones sobre la Realidad Social (CIRES) es un organismo patrocinado por varias instituciones financieras cuyo objetivo es investigar sobre diversos aspectos de la realidad social de España. Ha publicado algunos trabajos examinando muy diversos aspectos. Mejor aún, ha hecho públicos los resultados de algunas encuestas efectuadas hasta el nivel del microdato —es decir, no sólo resultados agregados, sino los datos resultantes de grabar las encuestas—. Ello constituye una información muy rica que permite a cualquier investigador reproducir los resultados del CIRES u obtener otros sobre cuestiones de su interés.

El fichero `cires.dge` en el lugar de costumbre recoge las respuestas de 7200 personas en toda España a diversas preguntas. El formulario de la encuesta se incluye como apéndice para comodidad de referencia: la consulta de [1] y [2] puede también ser de utilidad¹.

1. Selecciona una tabla de dos dimensiones y analízala mediante un análisis de correspondencias. Da cuenta de tus hallazgos. Entre las ayudas tienes sugerencias sobre cosas que te puede interesar estudiar.
2. Selecciona una tabla con, al menos, tres dimensiones de la que la anterior sea un margen y analízala ajustando modelos logarítmico-lineales. Compara los resultados. Explica las ventajas e inconvenientes de las dos aproximaciones. Si los dos análisis arrojan resultados aparentemente contradictorios, explica por qué.

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

Bibliografía. Los libros [4], [5], [3] o [6] van todos mucho más allá de lo que vemos en el curso. Están en Biblioteca.

¹En Biblioteca con signatura topográfica 316.3.

Funciones y consejos para trabajar en R.

- Los datos, en el lugar de costumbre, puedes leerlos mediante un `dget`. Personas trabajando con ordenadores un poco limitados pueden tener problemas, pues se trata de una `dataframe` grande. Si fuera el caso, puedes extraer un subconjunto de las columnas de dicha `dataframe` en un ordenador capaz (en la Facultad, pidiéndolo a tu profesor o un compañero) y trabajar con los datos reducidos.
- Te resultará de utilidad para construir las tablas de tu elección la función `table`. Aplicada a varias columnas de una `dataframe` produce una tabla de contingencia de la dimensión adecuada. Por ejemplo, si quisieras obtener una tabulación cruzada de las variables en las columnas `d1`, `d2`, `d3` harías algo como:

```
attach(cires)
mitabla <- table(d1,d2,d3)
```

La invocación de `attach` tiene por objeto indicar a R que las variables `d1`, `d2`, `d3` son columnas de la `dataframe` `cires`.

- El análisis propiamente dicho puede efectuarse mediante la función `loglin`.
- Una tabla de contingencia construida por `table` es una matriz k -dimensional. Puedes referirte a sus filas y columnas con índices, si lo deseas. En particular, puede que la tabla que construyas tenga una fila, columna, etc. completamente llena de ceros (lo que impediría el trabajo de funciones como `loglin`: ¿ves por qué?). Imagina, por ejemplo, que es la segunda fila de una tabla tridimensional. La puedes eliminar así:

```
mitabla <- mitabla[-2, , ]
```

Si quisieras eliminar la segunda fila y la tercera columna:

```
mitabla <- mitabla[-2, -3, ]
```

- Hay excelentes herramientas de visualización para algunos tipos de tablas de contingencia, que pueden arrojar bastante luz complementaria (o previa) a la que proporcionen los modelos logarítmico-lineales ajustados. Puedes mirar, por ejemplo, `mosaicplot` o `fourfoldplot` (ambas en el paquete `vcd`, enteramente dedicado a la visualización de datos categóricos).

¿Que tablas analizar? ¡Hay millones de posibilidades! Escoge cualquiera que te interese. Por ejemplo, ¿está relacionado el nivel educativo de los padres con el de los hijos? (parece obvio que ha de estarlo). Pero, ¿de igual manera para todas las edades, para ambos sexos, en unas regiones que en otras, para diferentes niveles de ingresos. . . ?

¿Qué relación hay, si es que alguna, entre el grado de felicidad declarado por los sujetos y diferentes circunstancias personales (edad, sexo, lugar de residencia, educación, ingresos, profesión. . .)?

El sentimiento de pertenencia a comunidades distintas de España (como sentimiento de identidad vasca, gallega, catalana. . .) ¿cómo se relaciona con el nivel cultural, los ingresos, la edad, etc.?

Y cualquier otra cosa que se te ocurra.

Otros datos. Los hay. La fundación CIRES los presenta en diskettes en formato SPSS, que no es totalmente trivial traducir a R (ni a nada más). Si quieres, puedes pedirlos a tu profesor.

Varios. Lee detenidamente la documentación relativa a los datos que se adjunta. En particular, parece que la muestra ha sido estratificada por comunidades autónomas y tamaño de municipio. ¿Qué consecuencias tiene ello en la especificación de los modelos, en la interpretación de los resultados?

Referencias

- [1] *La realidad social en España 1993–95*. Fundación BBV, Fundación BBK y Fundación Caja Madrid, 1995.
- [2] *La realidad social en España 1995–96*. Fundación BBV, Fundación BBK y Fundación Caja Madrid, 1997.
- [3] A. Agresti. *Categorical Data Analysis*. Wiley, 1990.
- [4] Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland. *Discrete Multivariate Analysis. Theory and Practice*. MIT Press, Cambridge, Mass., 1975.
- [5] S.E. Fienberg. *The Analysis of Cross-Classified Categorical Data*. MIT Press, Cambridge, Mass., 1980.
- [6] S.J. Haberman. *Analysis of Qualitative Data*. Academic Press, 1979.