

# SOC3070 Análisis de Datos Categóricos

## Tarea corta 4

Ponderación: 4% de la nota final del curso. Entrega: Desde el momento de entrega, los estudiantes tienen 1 semana exacta de plazo para completar esta tarea.

### Problema:

En su artículo “Models for comparing mobility tables: toward parsimony and substance”, publicado en 1987 en *American Sociological Review*, Kazuo Yamaguchi (1987) presentó una tabla de contingencia 3-way en la que se clasifican las categorías ocupacionales de hijos y padres en tres países: USA, UK y Japón. Estos datos se han convertido en un clásico en modelos log-multiplicative con layer effects para comparar tablas de movilidad social. En esta oportunidad, sin embargo, usarán estos datos para estimar proporciones y hacer inferencia estadística sobre dichas proporciones.

```
print(ctable)
```

```
## , , Country = US
##
##      Father
## Son   UpNM LoNM  UpM  LoM Farm
## UpNM 1275 1055 1043 1159 666
## LoNM  364  597  587  791 496
## UpM   274  394 1045 1323 1031
## LoM   272  443  951 2046 1632
## Farm   17   31   47   52  646
##
## , , Country = UK
##
##      Father
## Son   UpNM LoNM  UpM  LoM Farm
## UpNM  474  300  438  601   76
## LoNM  129  218  254  388   56
## UpM    87  171  669  932  125
## LoM   124  220  703 1789  295
## Farm   11    8   16   37  191
##
## , , Country = Japan
##
##      Father
## Son   UpNM LoNM  UpM  LoM Farm
## UpNM  127   86   43   35  109
## LoNM  101  207   73   51  206
## UpM   24   64  122   62  184
## LoM   30   61   60   66  253
## Farm  12   13   13   11  325
```

En la tabla, UpNM refiere a profesionales, gerentes y funcionarios; LoNM son propietarios, vendedores y oficinistas; UpM son trabajadores cualificados; LoM son trabajadores no agrícolas semicualificados y no cualificados; y Farm son agricultores y obreros agrícolas.

Como se puede observar, la tabla tiene tres dimensiones: ocupación del hijo (filas), ocupación del padre (columnas) y país ("layer"). Para acceder a la sub-tabla 2-way correspondiente a cada país usa los índices de la tabla. Por ejemplo, `ctable[,1]` corresponde a la tabla para USA (layer=1). `ctable[1,1]` corresponde a la primera fila de la tabla para USA, y `ctable[,2,3]` corresponde a la columna 2 de la tabla para Japón (layer=3)

## Ejercicios:

- 1) Estima las distribuciones marginales de ocupación del hijo y ocupación del padre en cada país. Comenta brevemente las implicaciones sustantivas de los resultados obtenidos.
- 2) Asume que los recuentos de **cada columna** en la tabla provienen de una distribución multinomial, donde la probabilidad asociada a cada categoría puede ser estimada via MLE como sigue  $\hat{p}_{ij} = n_{ij}/n_{i+}$ . Por ejemplo, en UK la probabilidad de que una persona cuyo padre era agricultor/obrero agrícola (Farm) también sea agricultor u obrero agrícola (Farm) se estima como: 191/743 o, equivalentemente, 0.023/0.089.

Nos interesa estudiar los fenómenos *rags-to-riches* y *riches-to-riches* en los tres países. Para ello, debes estimar las siguientes 6 probabilidades:

$$\mathbb{P}(\text{Son} = \text{UpNM} \mid \text{Father} = \text{LoM}, \text{Country})$$

$$\mathbb{P}(\text{Son} = \text{UpNM} \mid \text{Father} = \text{UpNM}, \text{Country})$$

- 3) Usando la misma fórmula para intervalos de confianza de una proporción Binomial, calcula un intervalo de confianza al 99% para las proporciones estimadas en 2). Comenta brevemente las implicaciones sustantivas de los resultados obtenidos.

**Sugerencia:** Dado que vas a repetir la mismas operaciones varias veces, te sugiero la posibilidad de usar loops y/o funciones en R.

## Respuestas:

1)

```
for (k in 1:3) {
  country = dimnames(ctable)$Country[k]
  print("=====")
  print(paste0("Distribución marginal ocupación del hijo en ",country))
  print(apply(prop.table(ctable[, ,k]),1,sum))
  print("")
  print(paste0("Distribución marginal ocupación del padre en ",country))
  print(apply(prop.table(ctable[, ,k]),2,sum) )
  print("")
}
```

```
## [1] "=====
## [1] "Distribución marginal ocupación del hijo en US"
##      UpNM      LoNM      UpM      LoM      Farm
## 0.28502495 0.15545320 0.22300817 0.29303065 0.04348303
## [1] ""
## [1] "Distribución marginal ocupación del padre en US"
##      UpNM      LoNM      UpM      LoM      Farm
## 0.1207435 0.1381806 0.2014037 0.2945112 0.2451609
## [1] ""
## [1] "=====
## [1] "Distribución marginal ocupación del hijo en UK"
##      UpNM      LoNM      UpM      LoM      Farm
## 0.2272618 0.1257218 0.2386910 0.3766843 0.0316410
## [1] ""
## [1] "Distribución marginal ocupación del padre en UK"
##      UpNM      LoNM      UpM      LoM      Farm
## 0.09925409 0.11032243 0.25024062 0.45079403 0.08938884
## [1] ""
## [1] "=====
## [1] "Distribución marginal ocupación del hijo en Japan"
##      UpNM      LoNM      UpM      LoM      Farm
## 0.1710864 0.2728828 0.1950385 0.2010265 0.1599658
## [1] ""
## [1] "Distribución marginal ocupación del padre en Japan"
##      UpNM      LoNM      UpM      LoM      Farm
## 0.1257485 0.1843456 0.1330197 0.0962361 0.4606501
## [1] ""
```

2) y 3)

```
est_prop <- function(fila,columna,pais) {
  ctable[filas,columna,pais]/sum(ctable[,columna,pais])
}

conf_int <- function(fila,columna,pais,alpha) {
  z_halfalpha = qnorm(c(alpha/2, 1 - alpha/2))
  phat = est_prop(fila,columna,pais)
  n = sum(ctable[,columna,pais])
  se = sqrt((phat*(1-phat)/n))
  ci = phat + z_halfalpha*se
}

for (j in c(1,4)) {
  for (k in 1:3) {
    father_occ = dimnames(ctable)$Father[j]
    country = dimnames(ctable)$Country[k]
    p = est_prop(fila,columna,pais)
    ci = conf_int(fila,columna,pais,alpha=0.01)
    print("=====")
    print(paste0("Estimate P(Son=UpNM | Father=",father_occ,", Country=",country,"))
    print(round(p,2))
    print("")
```

```

print(paste0("99% CI Estimate P(Son=UpNM | Father=",father_occ," , Country=",country,""))
print(round(ci,2))
print("")
}
}

```

```

## [1] "=====
## [1] "Estimate P(Son=UpNM | Father=UpNM, Country=US)"
## [1] 0.58
## [1] ""
## [1] "99% CI Estimate P(Son=UpNM | Father=UpNM, Country=US)"
## [1] 0.55 0.61
## [1] ""
## [1] "=====
## [1] "Estimate P(Son=UpNM | Father=UpNM, Country=UK)"
## [1] 0.57
## [1] ""
## [1] "99% CI Estimate P(Son=UpNM | Father=UpNM, Country=UK)"
## [1] 0.53 0.62
## [1] ""
## [1] "=====
## [1] "Estimate P(Son=UpNM | Father=UpNM, Country=Japan)"
## [1] 0.43
## [1] ""
## [1] "99% CI Estimate P(Son=UpNM | Father=UpNM, Country=Japan)"
## [1] 0.36 0.51
## [1] ""
## [1] "=====
## [1] "Estimate P(Son=UpNM | Father=LoM, Country=US)"
## [1] 0.22
## [1] ""
## [1] "99% CI Estimate P(Son=UpNM | Father=LoM, Country=US)"
## [1] 0.20 0.23
## [1] ""
## [1] "=====
## [1] "Estimate P(Son=UpNM | Father=LoM, Country=UK)"
## [1] 0.16
## [1] ""
## [1] "99% CI Estimate P(Son=UpNM | Father=LoM, Country=UK)"
## [1] 0.14 0.18
## [1] ""
## [1] "=====
## [1] "Estimate P(Son=UpNM | Father=LoM, Country=Japan)"
## [1] 0.16
## [1] ""
## [1] "99% CI Estimate P(Son=UpNM | Father=LoM, Country=Japan)"
## [1] 0.09 0.22
## [1] ""

```