

```
## Religious identification now (Y) and at age 16 (X)
## Table 2.6, p. 26, in Agresti, A. (1990, Categorical Data Analysis,
## Wiley: New York)
## Source: 1978 USA General Social Survey
```

```
relig <- matrix(c(918, 27, 1, 70, 30, 351, 0, 37, 1, 1, 28, 1, 29,
                 5, 0, 25), ncol = 4, byrow = TRUE)
rownames(relig) <- c("Protestante", "Católica", "Judaísmo",
                    "Nenhuma ou outra")
colnames(relig) <- c("Protestante", "Católica", "Judaísmo",
                    "Nenhuma ou outra")
```

```
# Tabela com os totais
addmargins(relig)
```

	Protestante	Católica	Judaísmo	Nenhuma ou outra	Sum
Protestante	918	27	1	70	1016
Católica	30	351	0	37	418
Judaísmo	1	1	28	1	31
Nenhuma ou outra	29	5	0	25	59
Sum	978	384	29	133	1524

```
# Distribuição conjunta
dconj <- prop.table(relig)
print(dconj, digits = 2)
```

	Protestante	Católica	Judaísmo	Nenhuma ou outra
Protestante	0.60236	0.01772	0.00066	0.04593
Católica	0.01969	0.23031	0.00000	0.02428
Judaísmo	0.00066	0.00066	0.01837	0.00066
Nenhuma ou outra	0.01903	0.00328	0.00000	0.01640

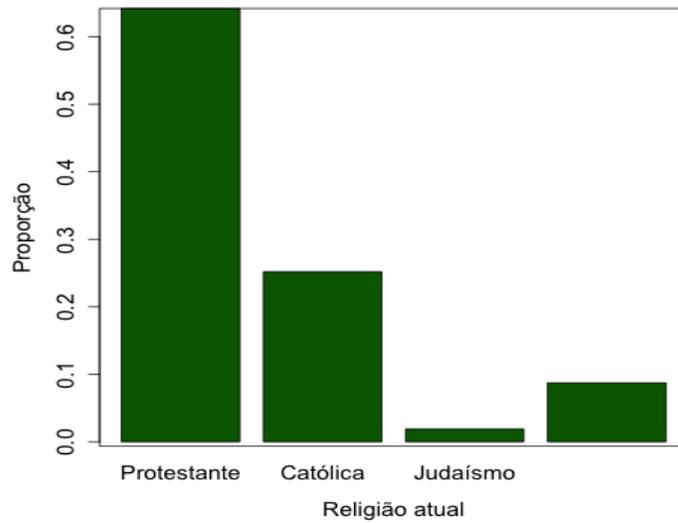
```
# Distribuição marginal de Y (religião atual)
dmargy <- margin.table(dconj, margin = 2)
print(dmargy, digits = 2)
```

Protestante	Católica	Judaísmo	Nenhuma ou outra
0.642	0.252	0.019	0.087

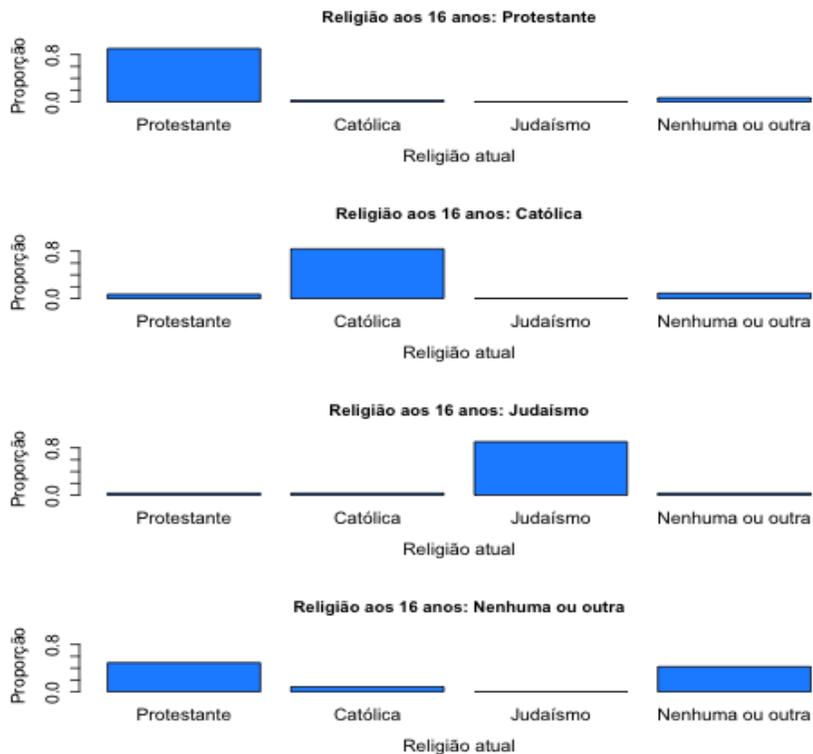
```
barplot(dmargy, xlab = "Religião atual", ylab = "Proporção",
        col = "darkgreen", cex = 1.3, cex.axis = 1.3, cex.lab = 1.3)
box()
```

```
# Distribuição condicional de Y | X (religião aos 16 anos)
dcondy <- prop.table(relig, margin = 1)
print(dcondy, digits = 2)
```

	Protestante	Católica	Judaísmo	Nenhuma ou outra
Protestante	0.904	0.027	0.00098	0.069
Católica	0.072	0.840	0.00000	0.089
Judaísmo	0.032	0.032	0.90323	0.032
Nenhuma ou outra	0.492	0.085	0.00000	0.424



```
# Gráficos das distribuições condicionais
I <- length(rownames(relig))
maxp <- max(dcondy)
par(mfrow = c(I, 1))
for (i in 1:I) {
  barplot(dcondy[i, ], xlab = "Religião atual", ylab = "Proporção",
    main = paste("Religião aos 16 anos:", rownames(relig)[i]),
    col = "dodgerblue", cex = 1.3, cex.axis = 1.3, cex.lab = 1.3,
    ylim = c(0, maxp))
}
```



No código acima, o eixo vertical (Proporção) é apresentado na mesma escala em todas as distribuições condicionais.

```
## Cálculo de tau^
# Distribuição marginal de X
dmargx <- margin.table(dconj, margin = 1)

vy <- 1 - sum(dmargy^2)
vyx <- 1 - sum(dmargx * rowSums(dcondy^2))
tauc <- 1 - vyx / vy

cat("\n Estimativas:")
cat("\n Prob. de classificação incorreta:", vy)
cat("\n Média da prob. cond. de classificação incorreta:", vyx)
cat("\n tau:", tauc)
```

Estimativas:

```
Prob. de classificação incorreta: 0.5167133
Média da prob. cond. de classificação incorreta: 0.221903
tau: 0.571
```

Nota 1. Como interpretar a estimativa de τ (tau) acima?

Nota 2. Escreva uma função para o gráfico das distribuições condicionais e utilizando a função `apply`, substitua a execução em laço (`for`).

Nota 3. Apresente outra forma de calcular `vyx` acima.

Nota 4. Refaça o exemplo com outros programas estatísticos (SAS, SPSS, etc).