

```
## Happiness and Relative Family Income
## Table 2.3, p 22, in Agresti, A. (2010, Analysis of Ordinal
## Categorical Data, 2nd ed., Wiley: Hoboken)

# Source: 2006 USA General Social Survey
# Y: job satisfaction
# X: income
tc <- matrix(c(208, 527, 185, 131, 835, 454, 49, 294, 272),
             ncol = 3, byrow = TRUE)
rownames(tc) <- c("Below average", "Average", "Above average")
colnames(tc) <- c("Not too happy", "Pretty happy", "Very happy" )
```

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

	Not too happy	Pretty happy	Very happy	Sum
Below average	208	527	185	920
Average	131	835	454	1420
Above average	49	294	272	615
Sum	388	1656	911	2955

```
# Distribuição condicional de Y | X
dcondy <- prop.table(tc, margin = 1)
print(dcondy, digits = 2)
```

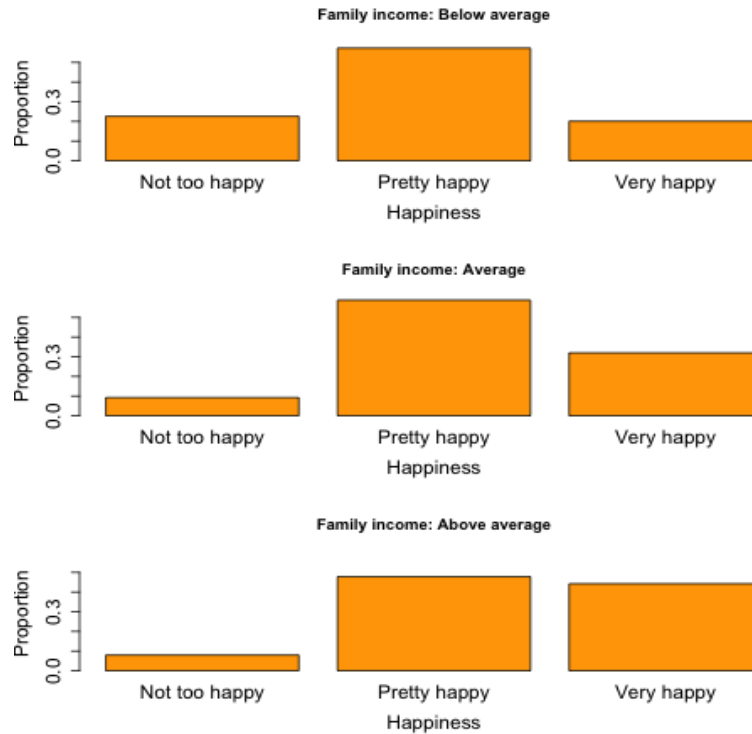
	Not too happy	Pretty happy	Very happy
Below average	0.226	0.57	0.20
Average	0.092	0.59	0.32
Above average	0.080	0.48	0.44

```
# Gráficos das distribuições condicionais
I <- nrow(tc)
J <- ncol(tc)
```

```
maxp <- max(dcondy)
par(mfrow = c(I, 1))
for (i in 1:I) {
  barplot(dcondy[i, ], xlab = "Happiness", ylab = "Proportion",
          main = paste("Family income:", rownames(tc)[i]), cex = 1.5,
          ylim = c(0, maxp), col = "orange", cex.lab = 1.5, cex.axis = 1.5)
}
```

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

Nota1. Com base no gráfico da página seguinte, procure explicar (ao seu cliente) que existe uma tendência de encontrarmos mais felicidade em famílias com renda mais alta.



Nota 2. Procure justificar que os números de pares concordantes (C) e discordantes (D) são calculados com as expressões abaixo, em que n_{ij} representa o número de observações com $X = i$ e $Y = j$.

$$C = \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} n_{ij} \sum_{k=i+1}^I \sum_{l=j+1}^J n_{kl}.$$

$$D = \sum_{i=1}^{I-1} \sum_{j=2}^J n_{ij} \sum_{k=i+1}^I \sum_{l=1}^{j-1} n_{kl}.$$

```
# Número de pares concordantes
nC <- 0
for (i in 1:(I - 1)) {
  for (j in 1:(J - 1)) {
    nC <- nC + tc[i, j] * sum(tc[(i + 1): I, (j + 1): J])
  }
}

# Número de pares discordantes
nD <- 0
for (i in 1:(I - 1)) {
  for (j in 2:J) {
    nD <- nD + tc[i, j] * sum(tc[(i + 1): I, 1: (j - 1)])
  }
}
```

```
cat("\n Número de pares concordantes (C):", nC)  
cat("\n Número de pares discordantes (D):", nD)  
cat("\n gama:", round((nC - nD) / (nC + nD), digits = 3))
```

Número de pares concordantes: 1069708

Número de pares discordantes: 533662

gama: 0.334

Nota 3. Modifique os dados de modo que a estimativa de γ seja (i) -1 , (ii) 0 e (iii) 1 .

Nota 4. Escreva um código em R para o cálculo do número de pares empatados.

Nota 5. Refaça o exemplo utilizando a função `GKgamma` do pacote `vcdExtra` em R.

Nota 6. Refaça o exemplo com outros programas estatísticos (SAS, SPSS, etc).