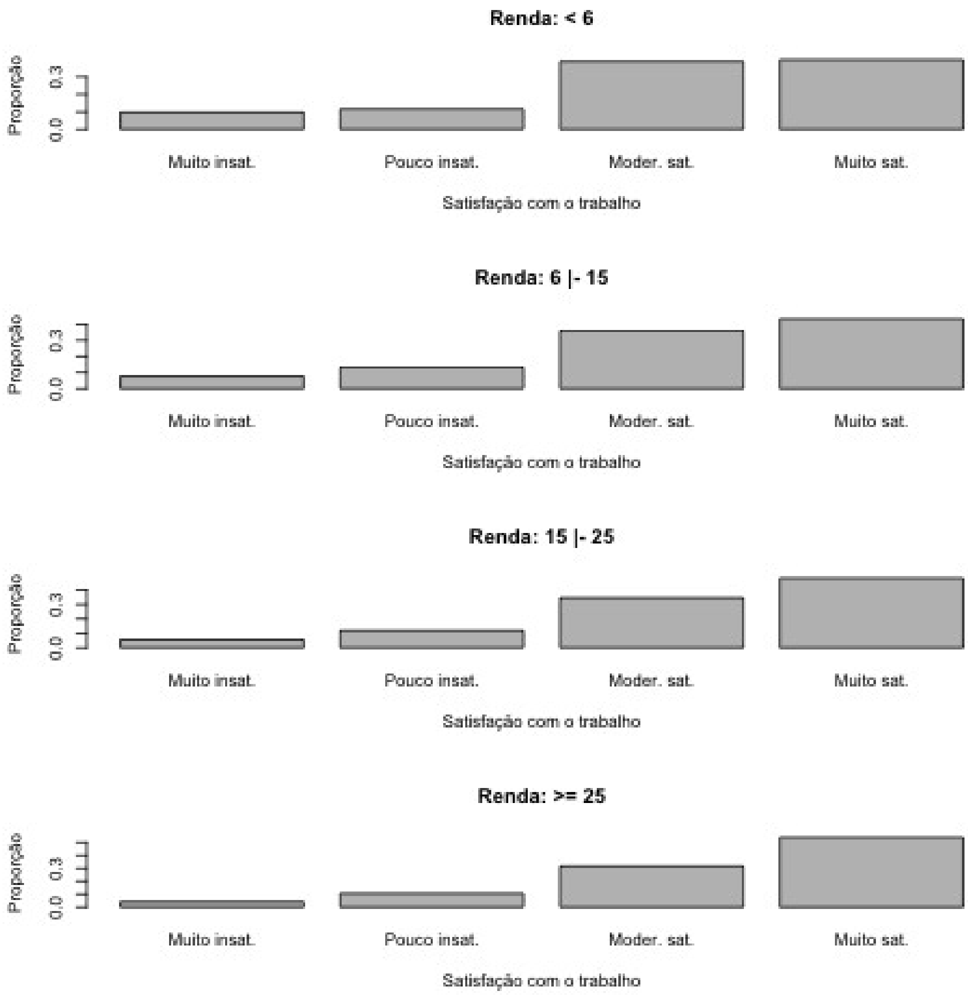
```
## Cross-classification of job satisfaction by income
## Table 2.4, p. 21, in Agresti, A. (1990, Categorical Data Analysis,
##    Wiley: New York)

# Y: job satisfaction
# X: income
tc <- matrix(c(20, 24, 80, 82, 22, 38, 104, 125, 13, 28, 81, 113,
               7, 18, 54, 92), ncol = 4, byrow = TRUE)
rownames(tc) <- c("< 6", "6 |- 15", "15 |- 25", ">= 25")
colnames(tc) <- c("Muito insat.", "Pouco insat.", "Moder. sat.",
                  "Muito sat.")

# Tabela com os totais
addmargins(tc)

         Muito insat. Pouco insat. Moder. sat. Muito sat. Sum
< 6                20           24          80         82 206
6 |- 15            22           38         104        125 289
15 |- 25           13           28          81        113 235
>= 25               7           18          54         92 171
Sum                62          108         319        412 901

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

         Muito insat. Pouco insat. Moder. sat. Muito sat.
< 6             0.097         0.12        0.39       0.40
6 |- 15         0.076         0.13        0.36       0.43
15 |- 25        0.055         0.12        0.34       0.48
>= 25           0.041         0.11        0.32       0.54

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

par(mfrow = c(I, 1))
for (i in 1:I) {
  barplot(dcondy[i,], xlab = "Satisfação com o trabalho",
     ylab = "Proporção", main = paste("Renda:", rownames(tc)[i]))
}
```

Nota1. Com base no gráfico da página seguinte, explique (ao seu cliente) que existe uma tendência fraca de encontrarmos mais satisfação com o trabalho em indivíduos com renda mais alta.



Procure justificar que os números de pares concordantes (C) e discordantes (D) são calculados com as expressões abaixo.

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

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

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

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

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

```
Número de pares concordantes (C): 109520
Número de pares discordantes (D): 84915
gama: 0.1265461
```

O cálculo de C e D é simplificado no código abaixo.

```
# Número de pares concordantes
nC <- 0L
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])
   }
}

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

Nota 2. Escreva um código em R para o cálculo do número de pares empatados.

Nota 3. Verifique o resultado da função `GKgamma` do pacote `vcdExtra` em R.

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