

### UNIDAD 3: Práctica 09-Análisis de una variable bidimensional categórica.

Ejemplo: Se selecciona aleatoriamente una muestra de 18 personas adultas, para estudiar si existe relación entre su estado civil y su ocupación.

#### REALICE UN ANÁLISIS ESTADÍSTICO DE LOS DATOS.

##### 1) Activa tu directorio de trabajo.

```
getwd()

## [1] "C:/Users/user/OneDrive/Paquete R/Practicas-S3"

setwd("C:/Users/user/OneDrive/Paquete R/Practicas-S3")
```

##### 2) Limpia de objetos el área de trabajo (Workspace).

```
ls()

## character(0)

rm(list=ls(all=TRUE))
ls()

## character(0)
```

##### 3) Crea un nuevo Script y llámale "Script09-DatosBivariados1".

##### 4) Crea en Excel una hoja de datos con dos columnas o variables

Recuerda que al guardar la hoja, el tipo de archivo es de extensión .csv(delimitado por comas).

Llámale al archivo: HojaCat

Otra forma de crear la hoja de datos es la siguiente (Vea la Práctica 04):

Primero crear las dos variables categóricas en un editor de texto como NotePad o WordPad, colocando nombre a cada columna, y llamándole "HojaCat.txt".

Luego puede leer o recuperar este archivo con la función read.table()

```
HojaCat <- read.table("HojaCat.txt", header=TRUE)
HojaCat

##      Estado  Ocupacion
## 1    casado desocupado
## 2    soltero  estudia
## 3    soltero  trabaja
## 4    casado  estudia
## 5 acompañado trabaja
## 6    soltero desocupado
## 7    casado  trabaja
## 8    casado  estudia
## 9 acompañado desocupado
```

```
## 10 acompañado estudia
## 11 casado trabaja
## 12 soltero estudia
## 13 acompañado desocupado
## 14 casado desocupado
## 15 soltero estudia
## 16 soltero trabaja
## 17 casado desocupado
## 18 soltero trabaja
```

##### 5) Recupera desde el entorno de R la hoja de datos de Excel.

```
HojaCat<-read.csv("HojaCat.csv", strip.white=TRUE)
HojaCat
```

```
## Estado.Ocupacion
## 1 casado;desocupado
## 2 soltero;estudia
## 3 soltero;trabaja
## 4 casado;estudia
## 5 acompañado;trabaja
## 6 soltero;desocupado
## 7 casado;trabaja
## 8 casado;estudia
## 9 acompañado;desocupado
## 10 acompañado;estudia
## 11 casado;trabaja
## 12 soltero;estudia
## 13 acompañado;desocupado
## 14 casado;desocupado
## 15 soltero;estudia
## 16 soltero;trabaja
## 17 casado;desocupado
## 18 soltero;trabaja
```

##### 6) Conecta la hoja de datos a la segunda ruta o lista de búsqueda.

```
attach(HojaCat, pos=2) #pos especifica la posición donde buscar la conexión
search()
```

```
## [1] ".GlobalEnv" "HojaCat" "package:knitr"
## [4] "package:stats" "package:graphics" "package:grDevices"
## [7] "package:utils" "package:datasets" "package:methods"
## [10] "Autoloads" "package:base"
```

##### 7) Crea una tabla de contingencia o de doble entrada

```
tablaCont<-table(HojaCat)
tablaCont

## HojaCat
## acompañado;desocupado acompañado;estudia acompañado;trabaja
## 2 1 1
## casado;desocupado casado;estudia casado;trabaja
## 3 2 2
```

```
##      soltero;desocupado      soltero;estudia      soltero;trabaja
##              1              3              3

length(HojaCat)

## [1] 1
```

*Note que esta instrucción no devuelve el número de elementos, sino más bien el número de variables o columnas consideradas en el conjunto de datos.*

Encuentra la suma de cada fila de la tabla de contingencia

Distribución marginal de X=Estado civil

```
suma.filas<-apply(tablaCont, 1, sum)# El 1 indica que son totales por fila
suma.filas

## acompañado;desocupado      acompañado;estudia      acompañado;trabaja
##              2              1              1
##      casado;desocupado      casado;estudia      casado;trabaja
##              3              2              2
##      soltero;desocupado      soltero;estudia      soltero;trabaja
##              1              3              3
```

Encuentra la suma de cada fila de la tabla de contingencia

distribución marginal de Y=Ocupación

```
suma.columnas<-apply(tablaCont,2,sum)# 2 indica que son totales por columna

## Error in apply(tablaCont, 2, sum):  'MARGIN' does not match dim(X)

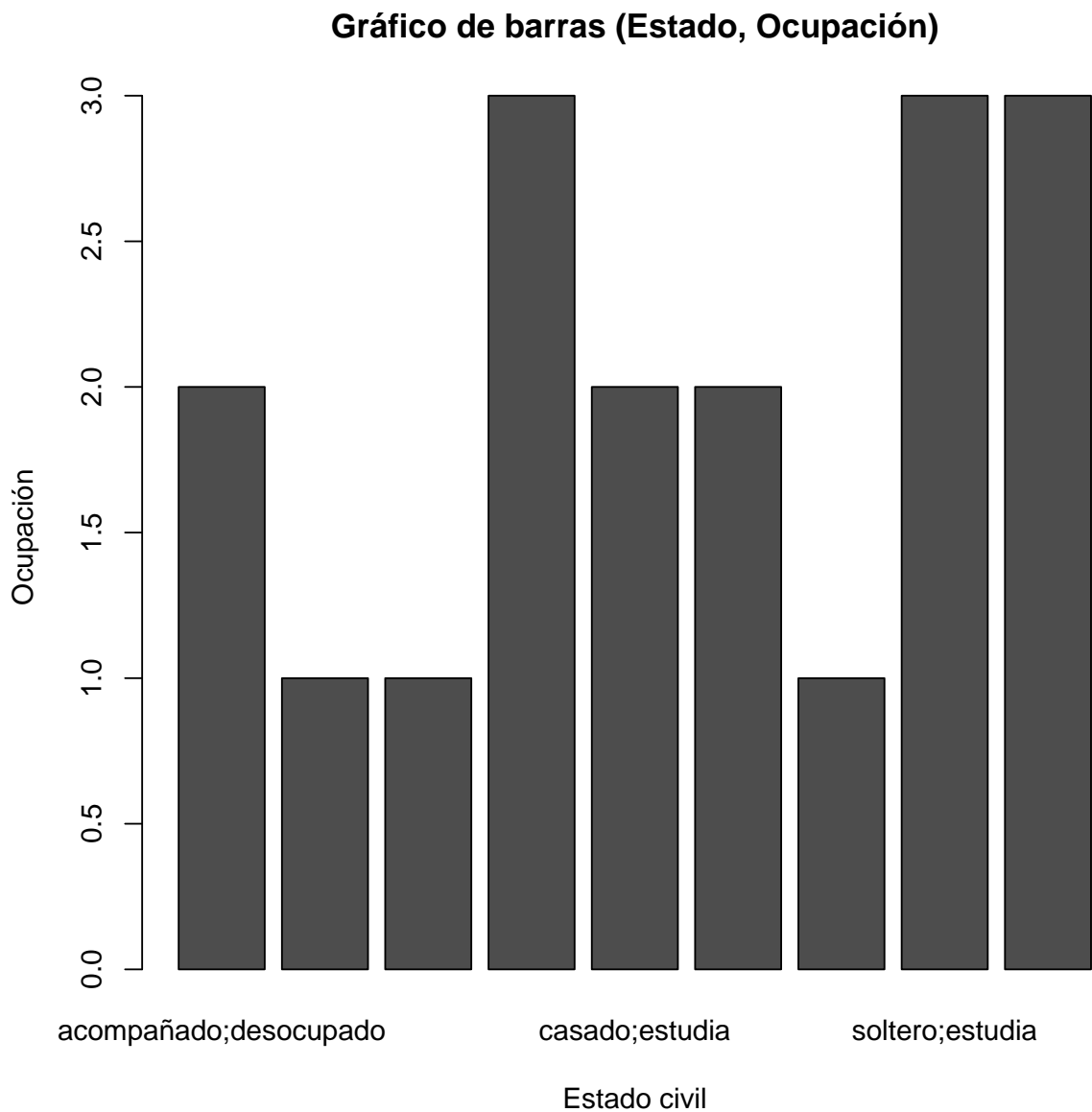
suma.columnas

## Error in eval(expr, envir, enclos):  objeto 'suma.columnas' no encontrado
```

**Gráficos de barras para tabla de contingencia.**

Barras apiladas

```
barplot(t(tablaCont), main="Gráfico de barras (Estado, Ocupación)", xlab="Estado civil",ylab="Ocupación")
```

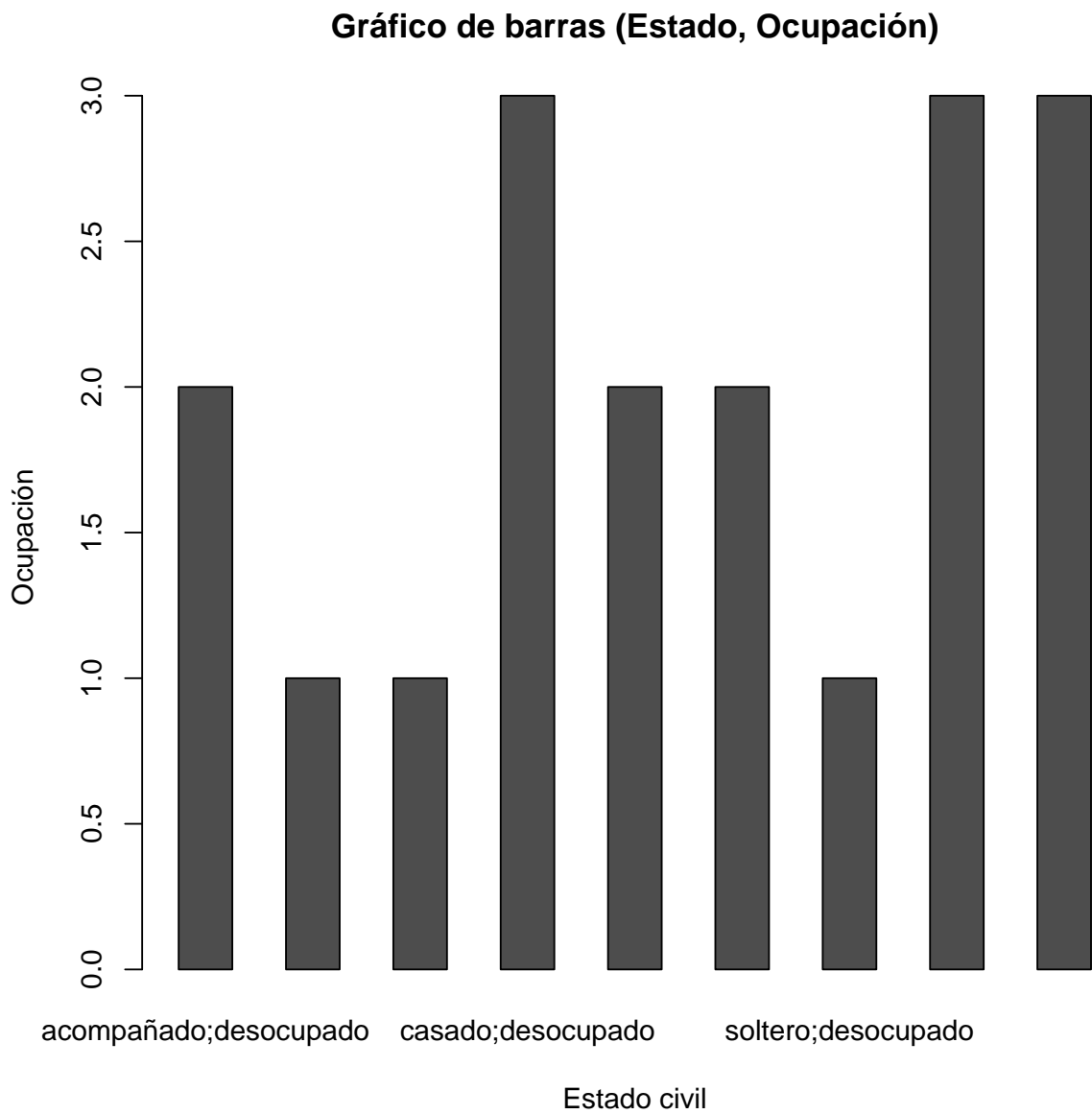


*Note que  $t(\text{tablaCont})$  indica que las barras representan el Estado civil de los encuestados y que éstas se subdividen en cada una de las diferentes ocupaciones consideradas.*

*En caso de usar únicamente  $\text{tablaCont}$ ; las barras representarán las diferentes ocupaciones y éstas estarán subdividas en cada uno de los estados civiles.*

Barras agrupadas

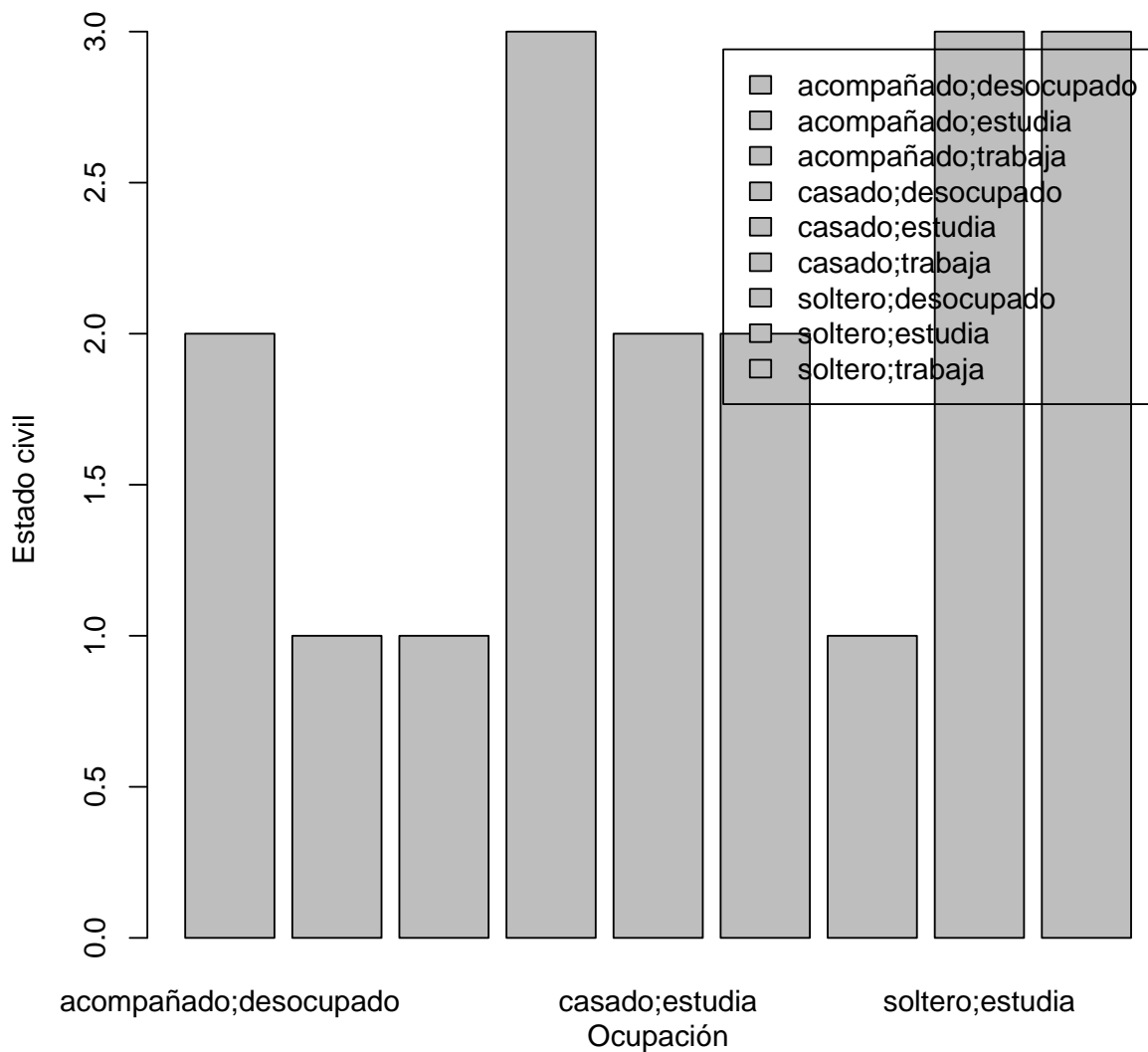
```
barplot(t(tablaCont), main="Gráfico de barras (Estado, Ocupación)", xlab="Estado civil", ylab="Ocupación")
```



*Note que la instrucción `beside = TRUE`, indica que por cada una de las diferentes ocupaciones se creará una barra para cada estado civil. Note que al usar `beside = FALSE` se obtiene el mismo gráfico de la instrucción anterior.*

```
barplot(tablaCont, main="Gráfico de barras (Ocupación, Estado)", xlab="Ocupación\n", ylab="Estado civil")
```

**Gráfico de barras (Ocupación, Estado)**



#### 8) Calcula tablas de proporciones o de probabilidades

Guardar las todas las opciones iniciales y modificar número de decimales

```
op<-options()
options(digits=3) #sólo imprime 3 lugares decimales
options('digits')

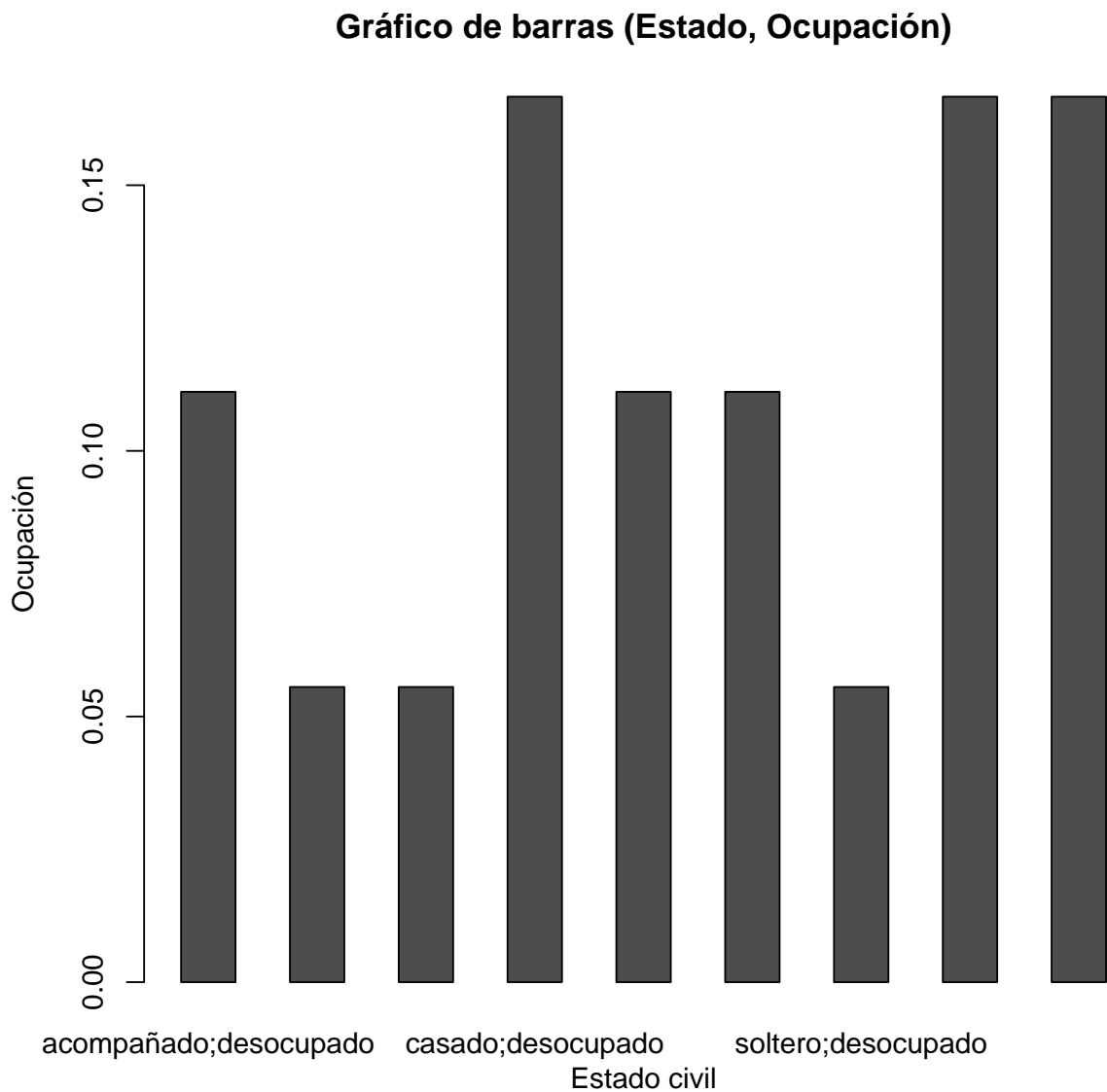
## $digits
## [1] 3
```

Proporciones basadas en el total de la muestra, la suma de filas y columnas suman 1.

```
propTotal<-prop.table(tablaCont)
propTotal
```

```
## HojaCat
## acompañado;desocupado    acompañado;estudia    acompañado;trabaja
##           0.1111           0.0556           0.0556
##      casado;desocupado      casado;estudia      casado;trabaja
##           0.1667           0.1111           0.1111
##      soltero;desocupado     soltero;estudia     soltero;trabaja
##           0.0556           0.1667           0.1667

barplot(t(propTotal), main="Gráfico de barras (Estado, Ocupación)", xlab="Estado civil\n", ylab="Ocupación")
```



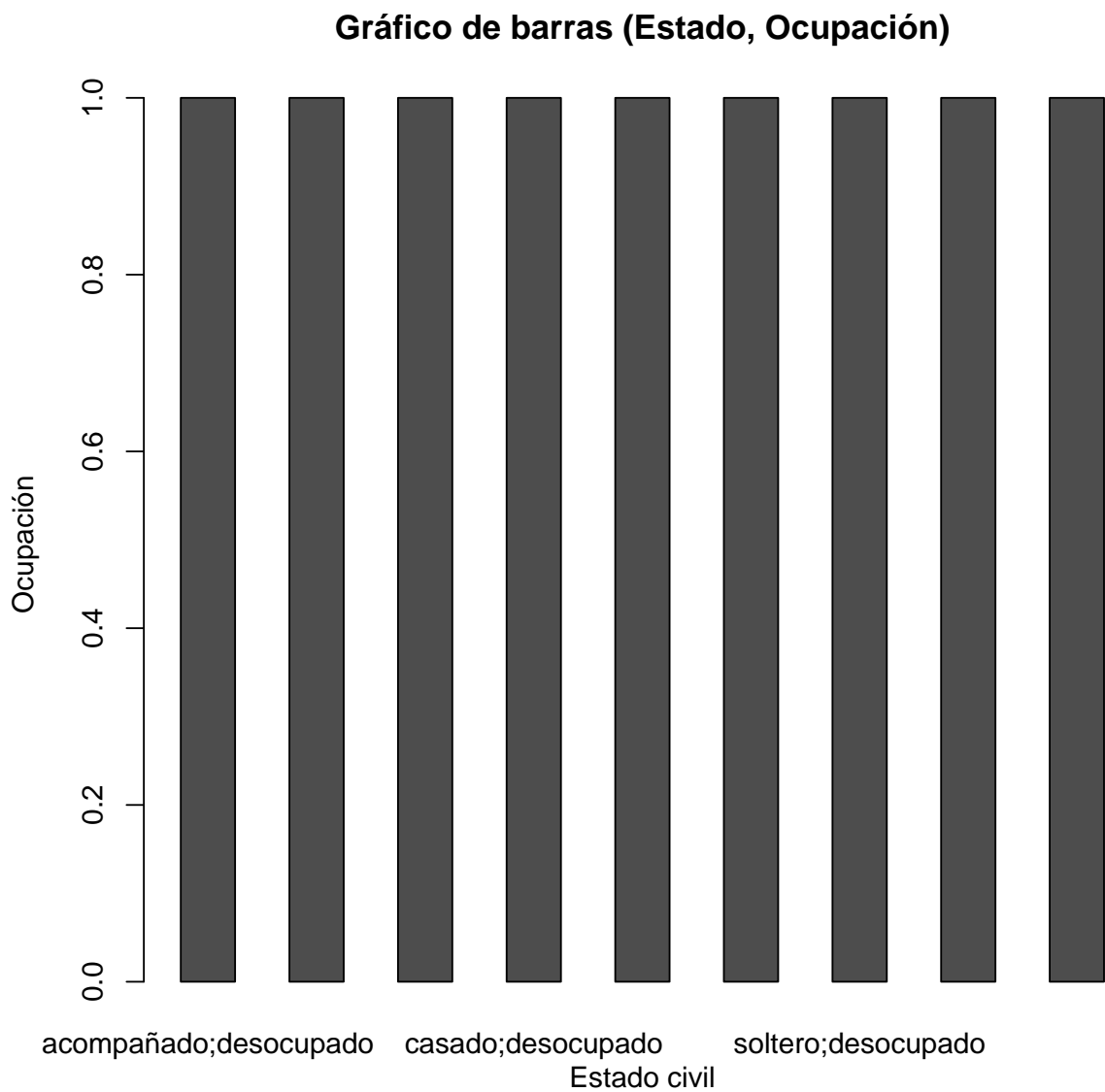
Proporciones basadas en el total por fila, cada fila suma 1.

```
propFila <- prop.table(tablaCont, 1)
propFila
## HojaCat
```

##	acompañado;desocupado	acompañado;estudia	acompañado;trabaja
##	1	1	1
##	casado;desocupado	casado;estudia	casado;trabaja
##	1	1	1
##	soltero;desocupado	soltero;estudia	soltero;trabaja
##	1	1	1

Total por fila se indica en 1

```
barplot(t(propFila), main="Gráfico de barras (Estado, Ocupación)", xlab="Estado civil\n", ylab="Ocupación")
```



Proporciones basadas en el total por columna, cada columna suma 1.



```
propColum<-prop.table(tablaCont, 2)

## Error in apply(x, margin, sum): 'MARGIN' does not match dim(X)

propColum

## Error in eval(expr, envir, enclos): objeto 'propColum' no encontrado
```

Total por columna se indica en 2

```
barplot(propColum, main="Gráfico de barras (Ocupación, Estado)", xlab="Ocupación\n", ylab="Estado civil")

## Error in barplot(propColum, main = "Gráfico de barras (Ocupación, Estado)", : objeto 'propColum' no encontrado
```

**9) Otra forma de elaborar los gráficos de barras para el vector bidimensional categórico.**  
Gráfico de barras no apiladas y colocación de leyenda

```
barplot(table(Ocupacion, Estado), main="Gráfico de barras (Estado, Ocupación)", xlab="Estado civil", ylab="Ocupación")

## Error in table(Ocupacion, Estado): objeto 'Ocupacion' no encontrado

barplot(table(Estado, Ocupacion), main="Gráfico de barras (Ocupación, Estado)", xlab="Ocupación", ylab="Estado civil")

## Error in table(Estado, Ocupacion): objeto 'Estado' no encontrado

barplot(table(Estado, Ocupacion), main="Gráfico de barras (Ocupación, Estado)", xlab="Ocupación", ylab="Estado civil")

## Error in table(Estado, Ocupacion): objeto 'Estado' no encontrado
```

*Note que se puede definir a conveniencia la leyenda que se desea incorporar en el gráfico con la instrucción legend.text*

**10) Realizar la prueba o contraste Chi-cuadrado de independencia**

```
prueba<-chisq.test(tablaCont)

## Warning in chisq.test(tablaCont): Chi-squared approximation may be incorrect

prueba

##
## Chi-squared test for given probabilities
##
## data:  tablaCont
## X-squared = 3, df = 8, p-value = 0.9
```

*Tenga en cuenta que las frecuencias esperadas deben ser todas mayores a 5*

Frecuencias absolutas esperadas para la prueba Chi-cuadrada

```
prueba$expected #  $f_{ij} = f_{i.}/No. \text{ column}$ 

## acompañado;desocupado    acompañado;estudia    acompañado;trabaja
##                2                2                2
## casado;desocupado          casado;estudia          casado;trabaja
##                2                2                2
## soltero;desocupado         soltero;estudia         soltero;trabaja
##                2                2                2
```