

PRÁCTICA 5: Distribuciones bidimensionales o multidimensionales.

DESARROLLO DE LA PRÁCTICA

Tablas de contingencia.

Si se dispone de dos o más variables estadísticas necesitamos de una tabla de frecuencias (absolutas, relativas, porcentajes) de doble o superior entrada para recoger las frecuencias conjuntas observadas de las variables. Cuando ambas variables son cualitativas o numéricas con pocas modalidades procede la creación de las mencionadas tablas (tablas de contingencia). Si tuviésemos variables continuas, previamente, sería necesario llevar a cabo una recodificación de las variables mediante intervalos de clase o esperar a la práctica número 6. Para realizar las citadas tablas, disponemos de la librería de la instalación básica “base”, y las librerías “DescTools” y “vcd”.

Los siguientes comandos, del paquete base:

```
tabla1 <- table(sal, cafe)
print(tabla1)
prop.table(tabla1)
prop.table(tabla1, 1)
prop.table(tabla1, 2)
```

nos permiten crear una tabla de contingencia entre Sal y Café, con frecuencias absolutas.

sal	cafe			
	no toma	poco	moderado	mucho
poca	34	19	13	9
normal	26	33	26	21
mucha	4	4	5	6

Después la citada tabla en frecuencias relativas,

sal	cafe			
	no toma	poco	moderado	mucho
poca	0.170	0.095	0.065	0.045
normal	0.130	0.165	0.130	0.105
mucha	0.020	0.020	0.025	0.030

luego las distribuciones condicionadas de Café para cada nivel de Sal

sal	cafe			
	no toma	poco	moderado	mucho
poca	0.453	0.253	0.173	0.120
normal	0.245	0.311	0.245	0.198
mucha	0.211	0.211	0.263	0.316

y finalmente las distribuciones condicionadas de Sal en cada nivel de Café

sal	cafe			
	no toma	poco	moderado	mucho
poca	0.531	0.339	0.295	0.250
normal	0.406	0.589	0.591	0.583
mucha	0.062	0.071	0.114	0.167

El paquete DescTools, nos permite obtener las cuatro tablas anteriores de manera conjunta

```
library(DescTools)
Desc(table(sal,cafe),plotit=FALSE)
```

Si deseamos una tabla de contingencia con tres variables, podemos intentar los comandos:

```
tabla3<-table(sal,estudios,genero)
print(tabla3)
fa <- xtabs(~sal+estudios+genero, data=HIPER200)
ftable(fa)
```

La primera opción nos proporciona dos tablas de contingencia entre Sal y Estudios para cada nivel de Genero (en frecuencias absolutas). La segunda opción, con el paquete DescTools, nos proporciona la tabla siguiente

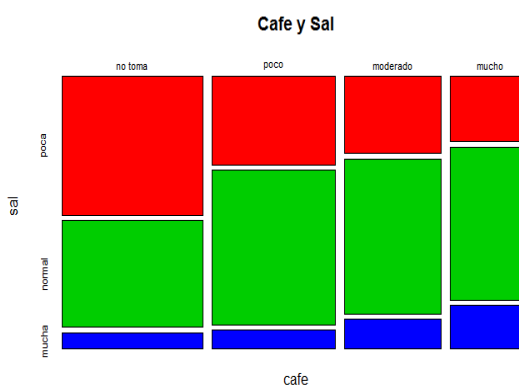
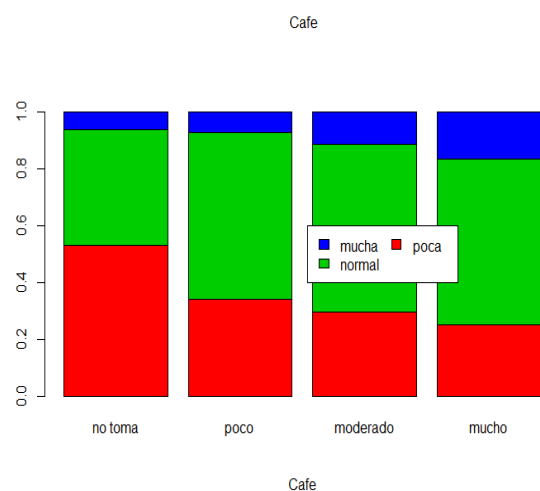
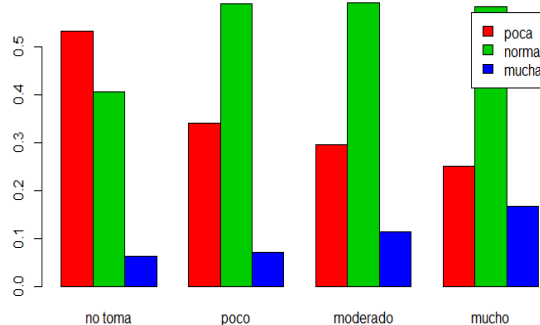
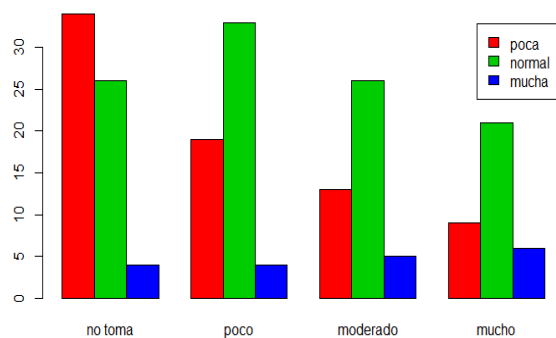
		genero	
		masculi no	femeni no
poca	estudi os		
	si n estudi os	16	16
	estudi os pri mari os	13	22
	estudi os secundari os	3	1
normal	estudi os superi ores	2	2
	si n estudi os	13	18
	estudi os pri mari os	26	28
	estudi os secundari os	9	7
much a	estudi os superi ores	5	0
	si n estudi os	2	2
	estudi os pri mari os	6	3
	estudi os secundari os	3	1
	estudi os superi ores	2	0

Para la realización de graficas con dos o más variables disponemos de las librerías ya citadas “graphics”, “lattice” y “ggplot2”, pero también los programas utilizados “DescTools” y “vcd”, entre otros disponen de gráficos dentro de sus respectivos paquetes. Realizaremos determinados gráficos, que pueden ser considerados como básicos, dentro de la amplia variedad de gráficos con dos o más variables.

Los siguientes comandos

```
t1a<-as.matrix(table(sal,cafe))
rownames(t1a)<- c("poca", "normal", "much a")
colnames(t1a)<- c("no toma", "poco", "moderado", "mucho")
barplot(prop.table(t1a,2),beside=TRUE,col=c(2,3,4),legend =rownames(t1a),sub="Cafe")
```

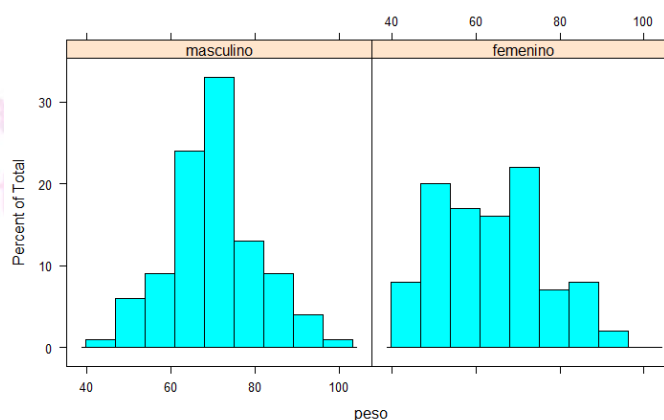
nos permite crear una matriz con la tabla de contingencia de las variables Sal y Café, darle nombres a las filas y columnas de la tabla creada, para luego representar un diagrama de rectángulos adosado. Se pide comentar las siguientes gráficas:



Si disponemos de una variable cuantitativa continua y un factor, podemos realizar histogramas de la variable continua por cada nivel del factor utilizado. Los siguientes comandos

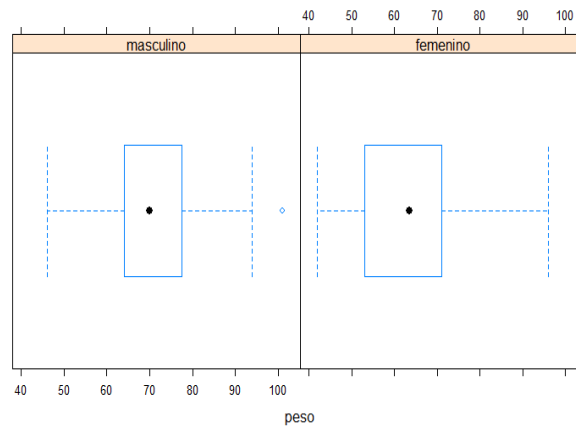
```
library(lattice)
histogram(~peso | genero)
```

obtienen los histogramas de la variable Peso agrupados por la variable Genero.



Otra opción para las variables anteriores, puede ser la realización de un diagrama de cajas (Boxplot) de la variable Peso para cada nivel de Genero

```
library(DescTools)
Desc(peso~sal,HIPER150, digits=2, plotit=TRUE)
```



También se puede calcular estadísticos de posición y dispersión de una variable cuantitativa para cada nivel de una variable cualitativa. Una posible opción se puede realizar con los comandos

```
library(DescTools)
Desc(peso~sal,HIPER150, digits=2, plotit=TRUE)
```

EJERCICIOS:

1. Para el conjunto de datos del fichero “HIPER200.RData”, se solicitan las siguientes cuestiones:
 - a) Crear una tabla de frecuencias conjuntas entre las variables Café y Actividad Física
 - b) La distribución de Actividad Física condicionada a la toma de Café.
 - c) Representaciones graficas apropiadas para estas dos variables.
 - d) ¿Son variables independientes.
 - e) Tomar la variable Peso agrupada en intervalos y la variable Actividad Física. Construir su tabla de frecuencias conjuntas.
 - f) Determinar la distribución del Peso para cada nivel de Actividad Física.
 - g) Representaciones graficas apropiadas para estas dos variables.
 - h) ¿Son variables independientes?
2. Para el conjunto de datos del fichero “EMPLEADOS.RData”, se solicitan las siguientes cuestiones:
 - a) Crear una tabla de frecuencias conjuntas entre las variables Género y Categoría laboral
 - b) La distribución de Categoría laboral condicionada al Género.
 - c) Representaciones graficas apropiadas para estas dos variables.
 - d) ¿Son variables independientes.
 - e) Tomar la variable Salario recodifica en ≥ 12000 , 12000-24000 y > 24000 , y la variable Tiempo en la empresa en año. Construir su tabla de frecuencias conjuntas.
 - f) Determinar la distribución del Salario para cada Año de antigüedad en la empresa.
 - g) Representaciones graficas apropiadas para estas dos variables.
 - h) ¿Son variables independientes?