

## Práctica 6

# Estadística descriptiva bidimensional: representación gráfica

En esta práctica aprenderemos a usar R Commander para representar gráficamente dos variables de forma simultánea, en particular, variables de tipo cualitativo.

### Contenido de la práctica

---

<b>6.1. Representación gráfica bivalente de variables cualitativas . . . . .</b>	<b>41</b>
6.1.1. Diagramas de barras apiladas . . . . .	41
6.1.2. Diagramas de barras agrupadas . . . . .	43
6.1.3. Diagramas de barras apiladas condicionadas . . . . .	44
6.1.4. Diagramas de barras agrupadas condicionadas . . . . .	45
<b>6.2. Ejercicios propuestos . . . . .</b>	<b>46</b>

---

### 6.1. Representación gráfica bivalente de variables cualitativas

Dadas dos variables cualitativas  $X$  e  $Y$  es posible representar sus distribuciones de frecuencias (conjunta, marginales o condicionadas) utilizando distintos tipos de diagramas de barras. Estos diagramas también pueden realizarse si alguna de las variables es cuantitativa (o las dos), siempre que dicha variable tome pocos valores distintos (puede ser el caso de algunas variables discretas) o esté previamente agrupada en clases o intervalos (suele ser el caso de las variables continuas).

#### 6.1.1. Diagramas de barras apiladas

El diagrama de barras apiladas permite representar, al mismo tiempo, las frecuencias (absolutas o relativas) de la distribución conjunta de dos variables  $X$  e  $Y$  y de la distribución marginal de alguna de ellas ( $X$  o  $Y$ ). Para realizar un diagrama de este tipo utilizamos la opción:

*Gráficas > Gráfica de barras...*

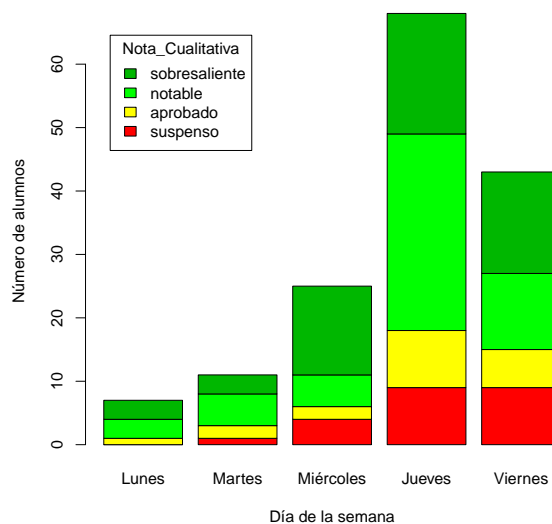
En la pestaña *Datos* se elige una de las variables (aquella para la que se desee representar, también, su distribución marginal). Después, se pulsa el botón *Gráfica por grupos...* y se selecciona la otra variable de interés. Por otro lado, en la pestaña *Opciones*:

- Marca la opción *Dividido (apilado)* que aparece en *Estilo del grupo de barras*.
- Marca la opción *Total* que aparece en *Porcentajes para barras agrupadas*.
- Desmarca la opción *Mostrar el conteo o los porcentajes en barras* que aparece en *Dibujar etiquetas*.

Observa que hay otras opciones que puedes modificar: frecuencias absolutas o porcentajes, posición de la leyenda, colores, títulos, etc.

**Ejercicio 39:** El fichero *test1.xlsx* contiene las calificaciones del primer test de evaluación realizado en la asignatura de Estadística del grado de Relaciones Laborales y Recursos Humanos durante el curso 2021/2022. Además de las calificaciones, se indica el sexo de cada alumno, el tiempo que ha tardado en realizar el test (minutos) y el día de la semana en el que lo comenzó.

1. Crea una variable, llamada *Nota\_Cualitativa*, que contenga la calificación cualitativa de los distintos alumnos (suspense, aprobado, notable y sobresaliente).
2. Las variables *Nota\_Cualitativa* y *Día* son cualitativas ordinales. Indícale a R Commander cuál es el orden de los valores que toman dichas variables, para que así aparezcan ordenadas.
3. Realiza un diagrama de barras apiladas con las frecuencias absolutas, donde haya una barra para cada día de la semana que a su vez se subdivida en varias barras más pequeñas que indique la nota obtenida por los alumnos.
4. Modifica los colores utilizados, los títulos de los ejes y la posición de la leyenda para que tu dibujo se asemeje a este:



5. A la vista del gráfico, responde:
  - a) ¿Qué día hubo más suspensos?
  - b) ¿Qué día se hicieron menos tests?
  - c) ¿Qué día hubo más notas por debajo del 7 (suspensos + aprobados)?

### 6.1.2. Diagramas de barras agrupadas

El diagrama de barras agrupadas permite representar las frecuencias de la distribución conjunta de dos variables  $X$  e  $Y$ . A diferencia del diagrama de barras apiladas, que permite acumular visualmente las frecuencias de barras contiguas, este diagrama facilita la comparación entre las distintas frecuencias conjuntas. Para realizar un diagrama de este tipo se utiliza la opción:

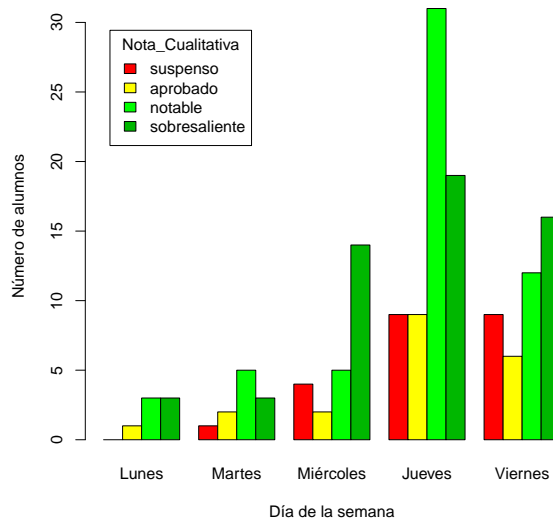
*Gráficas > Gráfica de barras...*

En la pestaña *Datos* se elige una de las variables y la otra se selecciona pulsando el botón *Gráfica por grupos...* Por otro lado, en la pestaña *Opciones*:

- Marca la opción *Lado a lado (paralelo)* que aparece en *Estilo del grupo de barras*.
- Marca la opción *Total* que aparece en *Porcentajes para barras agrupadas*.

Observa que hay otras opciones que puedes modificar: frecuencias absolutas o porcentajes, posición de la leyenda, colores, títulos, etc.

**Ejercicio 40:** Realiza un diagrama de barras agrupadas para las variables *Día* y *Nota\_Cualitativa*, de manera que las barras de un mismo día aparezcan contiguas, es decir:



A la vista del gráfico, responde:

1. ¿Qué día hubo más sobresalientes, el miércoles o el viernes? ¿Cómo se aprecia esta información en el gráfico con barras apiladas?
2. ¿Qué día hubo más notas por debajo del 7 (suspensos + aprobados)? ¿Cómo se aprecia esta información en el gráfico con barras apiladas?

### 6.1.3. Diagramas de barras apiladas condicionadas

Un diagrama de barras apiladas condicionadas permite representar las frecuencias relativas de una distribución condicionada: o bien de  $X$  condicionada a  $Y$ , o bien de  $Y$  condicionada a  $X$ . Para realizar un diagrama de este tipo se utiliza la opción:

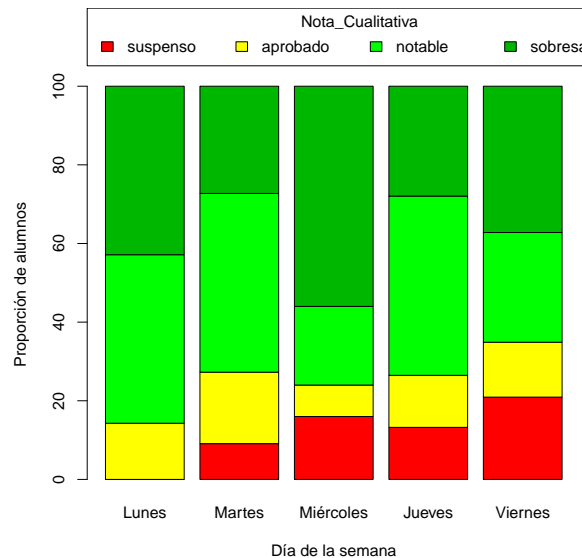
*Gráficas > Gráfica de barras...*

En la pestaña *Datos* se elige una de las variables: aquella a la que se quiere condicionar. Después, pulsando el botón *Gráfica por grupos...* se selecciona la otra variable a representar. Por otro lado, en la pestaña *Opciones*:

- Marca la opción *Porcentajes* que aparece en *Escala de los ejes*.
- Marca la opción *Dividido (apilada)* que aparece en *Estilo del grupo de barras*.
- Marca la opción *Condicional* que aparece en *Porcentajes para barras agrupadas*.
- Marca la opción *Encima del gráfico* que aparece en *Posición del texto*.
- Desmarca la opción *Mostrar el conteo o los porcentajes en barras* que aparece en *Dibujar etiquetas*.

Observa que hay otras opciones que puedes modificar: colores, títulos, etc.

**Ejercicio 41:** Realiza un diagrama de barras apiladas condicionadas para las variables *Día* y *Nota\_Cualitativa*, de manera que represente las frecuencias relativas en porcentaje de la variable *Nota\_Cualitativa* condicionada a la variable *Día*, es decir:



A la vista del gráfico, responde:

1. ¿Qué día hubo una proporción mayor de suspensos?
2. ¿Qué día se hicieron menos tests?
3. ¿Qué día hubo una proporción mayor de notas por debajo del 7 (suspensos + aprobados)?

Compara las respuestas que has dado ahora con las respuestas dadas al observar el diagrama de barras apiladas.

### 6.1.4. Diagramas de barras agrupadas condicionadas

Un diagrama de barras agrupadas condicionadas permite representar las frecuencias relativas de una distribución condicionada. A diferencia del diagrama de barras apiladas condicionadas, que permite acumular visualmente las frecuencias de barras contiguas, en este diagrama se facilita la comparación entre distintas frecuencias condicionadas. Para realizarlo se utiliza la opción:

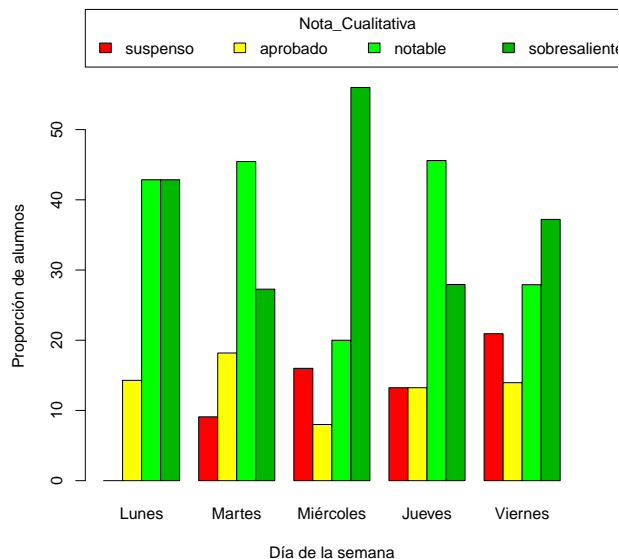
*Gráficas > Gráfica de barras...*

En la pestaña *Datos* se elige una la variable a la que se quiere condicionar. Después, se pulsa el botón *Gráfica por grupos...* y se selecciona la otra variable a representar. Por otro lado, en la pestaña *Opciones*:

- Marca la opción *Porcentajes* que aparece en *Escala de los ejes*.
- Marca la opción *Lado a lado (paralelo)* que aparece en *Estilo del grupo de barras*.
- Marca la opción *Condicional* que aparece en *Porcentajes para barras agrupadas*.
- Marca la opción *Encima del gráfico* que aparece en *Posición del texto*.

Observa que hay otras opciones que puedes modificar: colores, títulos, etc.

**Ejercicio 42:** Realiza un diagrama de barras agrupadas condicionadas para las variables *Día* y *Nota\_Cualitativa*, de manera que represente las frecuencias relativas en porcentaje de la variable *Nota\_Cualitativa* condicionada a la variable *Día*, es decir:



A la vista del gráfico, responde:

1. ¿Cuál fue la nota más frecuente el viernes? ¿y la menos frecuente?
2. ¿Qué día hubo mas sobresalientes?
3. ¿Qué día hubo mayor proporción de sobresalientes?
4. ¿Qué día hubo mayor proporción de notables?

## 6.2. Ejercicios propuestos

**Ejercicio 43:** A partir del conjunto de datos utilizado a lo largo de la práctica (correspondiente al fichero *test1.xlsx*), crea un nuevo conjunto de datos llamado *aprobados*, que contenga únicamente los datos de los alumnos que han obtenido una nota igual o superior a 5. Después, realiza los siguientes apartados utilizando el conjunto filtrado:

1. Representa, utilizando un diagrama de barras apiladas, las frecuencias relativas en porcentaje de la distribución conjunta de las variables *Sexo* y *Nota\_Cualitativa*, de manera que se aprecie la distribución marginal de la variable *Sexo*.

Nota: utiliza la opción *Datos > Modificar variables de conjunto de datos activo... > Descartar niveles sin uso...* para eliminar los posibles valores de una variable de tipo factor que ya no tiene sentido representar.

2. A la vista del gráfico anterior, ¿qué porcentaje de los datos corresponde a mujeres que han sacado sobresaliente? ¿qué porcentaje corresponde a hombres que han realizado el test?
3. Representa gráficamente la variable *Nota\_Cualitativa* condicionada a los valores de la variable *Sexo* utilizando un diagrama de barras apiladas condicionadas.
4. A la vista del gráfico anterior, ¿qué porcentaje de las mujeres ha sacado sobresaliente? ¿qué porcentaje de los hombres ha sacado sobresaliente?
5. ¿Qué relación existe entre las variables *Nota\_Cualitativa* y *Sexo*?
6. Representa gráficamente la variable *Sexo* condicionada a los valores de la variable *Nota\_Cualitativa* y observa de nuevo el aspecto del gráfico obtenido.

**Ejercicio 44:** El archivo *salarios2018.RData* ha sido generado a partir de los datos publicados por el INE en la Encuesta de Estructura Salarial correspondiente al año 2018. Cada una de las filas del conjunto de datos contiene la información de un trabajador, siendo las variables:

- *Edad*: edad del trabajador.
- *Estudios*: nivel máximo de estudios alcanzado.
- *Zona*: zona geográfica según la Nomenclatura de las Unidades Territoriales Estadísticas (NUTS).
- *Responsable*: responsabilidad en organización y/o supervisión.
- *Sexo*: sexo del trabajador.
- *Duracion.Contrato*: duración del contrato (indefinida o determinada).
- *Tipo.Jornada*: completa o parcial.
- *Salario*: salario bruto anual, en euros.

Responde a las siguientes preguntas:

1. Realiza un breve análisis descriptivo de las variables.

2. Dibuja un diagrama de barras apiladas para las variables *Edad* y *Responsable*, que tenga una barra para cada rango de edad:
  - a) ¿Qué rango de edad tiene mas trabajadores?
  - b) ¿En qué rango de edad hay mas personas con responsabilidad?
3. Dibuja un diagrama de barras apiladas que represente la variable *Responsable* condicionada a la variable *Edad*:
  - a) ¿Son independientes ambas variables?
  - b) ¿Qué relación observas?
  - c) ¿En qué rango de edad hay una proporción mayor de personas con responsabilidad?
4. Repite los dibujos de los dos apartados anteriores (2 y 3) utilizando las variables *Estudios* y *Responsable*:
  - a) ¿Qué nivel de estudios es más habitual entre los trabajadores?
  - b) ¿Cuál es nivel de estudios con mas personas con responsabilidad? ¿y el siguiente?
  - c) ¿Cuál es el nivel de estudios con mayor proporción de personas con responsabilidad? ¿y el siguiente?
5. Realiza un diagrama de barras que represente la variable *Duracion* condicionada a la variable *Sexo* y otro que represente la variable *Tipo.Jornada* condicionada a la variable *Sexo*. ¿Para que variable se observan mayores diferencias entre sexos?
6. Crea una nueva variable denominada *Salario.Niveles* que tome los valores *Bajo*, *Medio* y *Alto* según si el salario del trabajador está por debajo de 14000 €, entre 14000 y 30000 € o por encima de 30000 €, respectivamente (observa que 14000 € es, aproximadamente, el primer cuartil de la variable *Salario* y 30000 € el tercero). Después, indícale a R Commander que se trata de una variable cualitativa ordinal.
7. Estudia la relación que existe entre la variable *Salario.Niveles* y el resto de variables cualitativas. Para ello, responde a las siguientes preguntas realizando previamente el diagrama de barras que creas conveniente en cada caso:
  - a) ¿Qué nivel de estudios tiene la mayor proporción de salarios altos?
  - b) ¿Qué rango de edad tiene la mayor proporción de salarios bajos?
  - c) ¿Cuándo son mayores los salarios, en contratos de duración indefinida o determinada?
  - d) ¿Cuál es la zona geográfica con mayor número de salarios altos? ¿Y con menos salarios bajos?
  - e) ¿Cuál es la zona geográfica con mayor proporción de salarios altos? ¿Y de salarios bajo?
  - f) ¿Qué sexo tiene mayor proporción de salarios altos?
  - g) ¿Observas independencia entre la variable *Salario.Niveles* y alguna de las otras las variables cualitativas?
  - h) Indica si son verdaderas o falsas las siguientes frases:
    - Más de la mitad de los trabajadores con responsabilidad tiene un salario *Alto*.
    - El número de trabajadores con responsabilidad es superior al número de trabajadores sin responsabilidad que recibe un salario alto.