

Práctica 5

Estadística descriptiva bidimensional: tablas de contingencia

En esta práctica aprenderemos a usar R Commander para obtener tablas de contingencia. En particular, veremos cómo calcular las tablas de frecuencias absolutas y relativas (en porcentaje) que nos muestran las distribuciones conjunta, marginales y condicionadas.

Contenido de la práctica

5.1. Tabla de contingencia a partir de las frecuencias absolutas conjuntas	35
5.1.1. Distribución conjunta	36
5.1.2. Distribuciones marginales	36
5.1.3. Distribuciones condicionadas	37
5.2. Tablas de contingencia a partir de un fichero de datos	37
5.2.1. Distribuciones conjunta, marginales y condicionadas	37
5.3. Variables cuantitativas en tablas de contingencia	39

5.1. Tabla de contingencia a partir de las frecuencias absolutas conjuntas

Dadas dos variables cualitativas X e Y y sus modalidades, es decir, los distintos valores que puede tomar cada una de las variables, es posible que conozcamos las frecuencias conjuntas n_{ij} de todas ellas. Para introducir en R Commander directamente los valores de una tabla de contingencia utilizamos la opción:

Estadísticos > Tablas de contingencia > Introducir y analizar una tabla de doble entrada...

Al pulsar en dicha opción, se abre una ventana con dos pestañas: *Tabla* y *Estadísticos*. En la pestaña *Tabla* se introducen los datos de las frecuencias absolutas conjuntas de los que dispongamos. Además, permite especificar el nombre de las variables y sus posibles valores (es recomendable que pongáis toda la información posible). La pestaña *Estadísticos* la utilizaremos para seleccionar el tipo de tabla de contingencia que queremos obtener (también permite marcar opciones relativas a *Tests de hipótesis* pero, de momento, no las vamos a utilizar).

5.1.1. Distribución conjunta

Para obtener la distribución conjunta de frecuencias absolutas basta con marcar, en la pestaña *Estadísticos*, la opción *Sin porcentajes*. Para obtener la distribución conjunta de frecuencias relativas (en porcentaje) hay que marcar la opción *Porcentajes totales* (observa que en este último caso se proporcionan también las distribuciones marginales).

Ejercicio 34: Introduce en R Commander la siguiente tabla de contingencia, que contiene información relativa al número de estudiantes matriculados en la Facultad de Ciencias Sociales y del Trabajo de Zaragoza durante el curso 2020/2021. Las variables que se han tenido en consideración son *Estudios* y *Sexo*, que indican en qué estudios se matriculó el estudiante y su sexo, respectivamente.

Estudios	Sexo	
	Hombre	Mujer
RLRH	280	461
TS	128	608
Máster	4	33

Después, haz que muestre por pantalla tanto la tabla de contingencia con las frecuencias absolutas como con las frecuencias relativas y responde:

- ¿Qué porcentaje del estudiantado matriculado en la FCST es mujer y estudia Relaciones Laborales y Recursos Humanos?
- ¿Qué porcentaje del estudiantado matriculado en la FCST es estudiante de máster y hombre?

5.1.2. Distribuciones marginales

Para obtener la distribución marginal de frecuencias absolutas de la variable X (aquella cuyos valores hemos colocado en la primera columna) seleccionamos la opción *Porcentajes por filas* que aparece en la pestaña *Estadísticos*. Si observas el resultado, en la columna *count* aparecen los valores de la distribución marginal.

Análogamente, para obtener la distribución marginal de frecuencias absolutas de la variable Y (aquella cuyos valores hemos colocado en la primera fila) seleccionamos la opción *Porcentajes por columnas* que aparece en la pestaña *Estadísticos*. Si observas el resultado, en la fila *count* aparecen los valores de la distribución marginal.

La forma de obtener las distribuciones marginales de frecuencias relativas (en porcentaje) es más directa, basta con seleccionar la opción *Porcentajes totales* de la pestaña *Estadísticos*.

Ejercicio 35: Calcula, a partir de la información proporcionada en el ejercicio 34, las distribuciones marginales para las variables *Estudios* y *Sexo*, tanto con frecuencias absolutas como con frecuencias relativas. Después, responde:

- ¿Qué porcentaje se matriculó en un máster de la FCST?
- ¿Cuántos estudiantes hay en Trabajo Social?

5.1.3. Distribuciones condicionadas

Para obtener la distribución de X condicionada a cada uno de los posibles valores de la variable Y marcaremos la opción *Porcentaje por columnas* que aparece en la pestaña *Estadísticos*.

Análogamente, para obtener la distribución de Y condicionada a cada uno de los posibles valores de la variable X marcaremos la opción *Porcentaje por filas* que aparece en la pestaña *Estadísticos*.

Ejercicio 36: Calcula, a partir de la información proporcionada en el ejercicio 34, las distribuciones condicionadas para las variables *Estudios* y *Sexo*. Después, responde:

- ¿Qué porcentaje de los hombres matriculados en la FCST ha elegido Relaciones Laborales y Recursos Humanos?
- ¿Qué porcentaje el estudiantado de Trabajo Social es mujer?

5.2. Tablas de contingencia a partir de un fichero de datos

Generalmente, en lugar de partir de una tabla de contingencia con las frecuencias absolutas conjuntas, partiremos de un conjunto de datos. Para obtener las distintas tablas de contingencia, hay que pulsar la opción:

Estadísticos > Tablas de contingencia > Tabla de doble entrada...

Al pulsar en dicha opción se abre una ventana con dos pestañas: *Datos* y *Estadísticos*. En la pestaña *Datos* se seleccionan las dos variables a tener en cuenta y en la pestaña *Estadísticos* se indica el tipo de tabla que se quiere calcular. Observa que en la pestaña *Datos* es posible indicar que se quiere utilizar solamente una parte de los datos, ya que en la casilla *Expresión de selección* podemos escribir una expresión para filtrar los datos.

5.2.1. Distribuciones conjunta, marginales y condicionadas

Los pasos a seguir para calcular las distintas distribuciones son los mismos que los descritos en la sección 5.1. Es decir, se debe hacer uso de la pestaña *Estadísticos*. A modo de resumen, la siguiente tabla relaciona las opciones marcadas en *Estadísticos* con la información que R Commander proporciona como resultado.

Opción	Resultado
Sin porcentajes	Frecuencias absolutas conjuntas
Porcentajes totales	Frecuencias relativas conjuntas Distribuciones marginales relativas
Porcentajes por filas	Distribución marginal de la variable X Distribución de Y condicionada a X
Porcentajes por columnas	Distribución marginal de la variable Y Distribución de X condicionada a Y

Ejercicio 37: 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. Realiza un breve análisis descriptivo de las variables.
2. Calcula la tabla de contingencia con frecuencias absolutas para las variables *Sexo* y *Dia* (recuerda que si una variable tiene orden, entonces debe aparecer ordenada).
3. Completa todos los valores de las siguientes tablas.

		Y (Sexo)	
X (Dia)			

		Y (Sexo)	
X (Dia)			

		Y (Sexo)	
X (Dia)			

		Y (Sexo)	
X (Dia)			

4. A la vista de las tablas del apartado anterior, responde:
 - a) ¿Cuántos alumnos comenzaron el test?
 - b) ¿Qué día se empezaron a hacer mas test?
 - c) ¿Cuántos hombres hicieron el test en lunes?
 - d) ¿Qué porcentaje de los test se comenzó el lunes?
 - e) De los hombres, el % comenzó el test el jueves.
 - f) ¿Cuántas mujeres han hecho el test?
 - g) Entre los que hicieron el test el miércoles, ¿que porcentaje eran hombres?
 - h) ¿Qué porcentaje de las mujeres comenzó el test el jueves?
 - i) Dentro del conjunto de alumnos que comenzó el test el viernes, las mujeres representan un %.
 - j) De los que comenzaron el test en martes, eran mujeres.
5. Crea una variable, llamada *NotaC*, que contenga la calificación cualitativa de los distintos alumnos (suspense, aprobado, notable y sobresaliente).

6. Responde a las siguientes preguntas calculando previamente la tabla de contingencia oportuna:

- a) ¿Qué porcentaje de los alumnos que comenzó el test en miércoles ha suspendido?
- b) ¿Cuántos sobresalientes ha habido entre las mujeres?
- c) ¿Cuántas personas han sacado notable?
- d) ¿Cuál es el día que más sobresalientes ha habido?
- e) ¿Qué porcentaje de notables ha habido el día con menos notables?
- f) ¿Qué día ha sido menor el porcentaje de notables?

5.3. Variables cuantitativas en tablas de contingencia

También es posible realizar tablas de contingencia cuando alguna de las variables es cuantitativa, aunque teniendo en cuenta que:

- Si la variable toma pocos valores distintos, como suele ser el caso de algunas variables cuantitativas discretas, entonces habrá que considerar dicha variable numérica como factor:

Datos > Modificar variables del conjunto de datos activo > Convertir variable numérica en factor...

- Si la variable toma muchos valores distintos, como suele ser el caso de las variables cuantitativas continuas, entonces será necesario agrupar en intervalos los distintos valores que toma la variable:

Datos > Modificar variables del conjunto de datos activo > Recodificar variables...

Ejercicio 38: Utilizando la información proporcionada en el archivo *test1.xlsx*:

1. Calcula una tabla de contingencia con las frecuencias relativas de la distribución conjunta y de las distribuciones marginales para las variables *Nota* y *Sexo*.
2. ¿Cuál es la nota más frecuente para cada sexo?
3. Agrupa la variable *Tiempo* en los intervalos $[0, 30]$, $(30, 60]$, $(60, 90]$, $(90, 120]$ y $(120, +\infty)$, y llámala *TiemposInter*. Después, crea una variable llamada *Aprueba* que tome los valores *Sí* y *No*. Por último, responde las siguientes preguntas calculando previamente la tabla de contingencia oportuna:
 - a) ¿Cuántos estudiantes han suspendido tardando más de dos horas?
 - b) ¿Qué porcentaje de los que tardaron menos de media hora han aprobado?
 - c) ¿Qué porcentaje de los alumnos ha tardado entre media hora y una hora?
 - d) ¿En qué rango de tiempo se sitúa la mayor parte de los que han aprobado el test? ¿Y los que han suspendido?