



## PRÁCTICA 8. CONTRASTES DE HIPÓTESIS NO PARAMÉTRICOS.

Ejecute `library(Rcmdr)`

Continuamos trabajando con el fichero Respuestas. Los contrastes de hipótesis no paramétricos más comunes son:

1. Contraste de independencia para dos variables, que suelen ser categóricas (o numéricas, agrupadas en categorías o intervalos). *Menú estadísticos/tablas de contingencia/Tablas de doble entrada.*

El test puede escribirse:

$$\begin{cases} H_0: X \text{ e } Y \text{ son independientes} \\ H_1: X \text{ e } Y \text{ no son independientes} \end{cases}$$

donde el estadístico de contraste es:

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \rightarrow \chi^2_{(p-1)(q-1)}$$

siendo p, el número de filas y q, el número de columnas y:

$$e_{ij} = \frac{n_{i*} \cdot n_{*j}}{n}$$

Nota: Este test tiene el inconveniente de que cuando las frecuencias son inferiores a 5, no tiene validez estadística y por tanto, será necesario agrupar categorías.

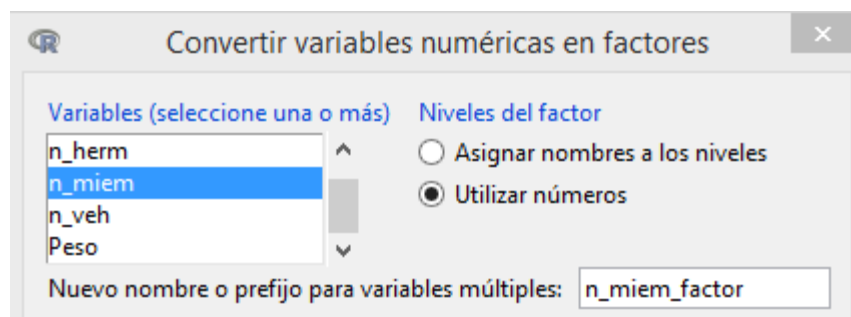
Nota: En el caso de tablas de contingencia 2x2, se utiliza el test exacto de Fisher.

2. Contraste Chi-Cuadrado de bondad de ajuste de una variable a una distribución discreta (*no disponible en R-commander. Es necesario instalar el paquete vcd y usar la orden `goodfit()`.*
3. Contraste de Shapiro-Wilk para analizar la bondad de ajuste a una distribución normal.

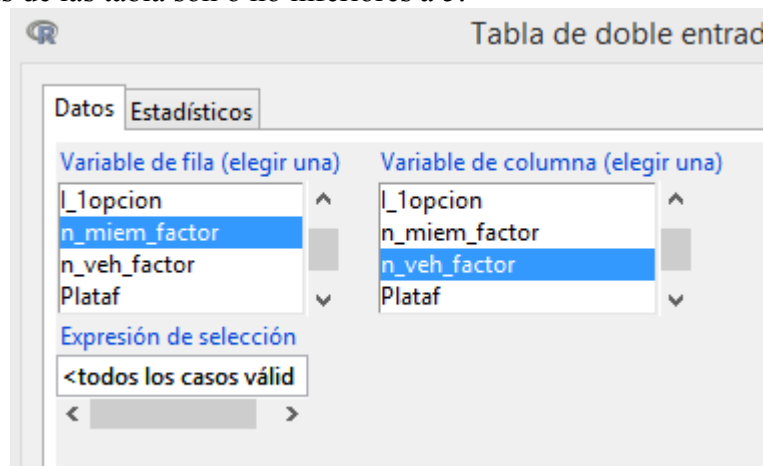


## Ejercicio 1

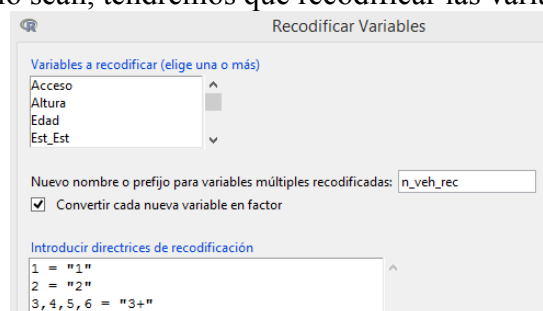
- Importamos los datos de Excel (Respuestas.xls) a R-Commander mediante el Menú Datos/Importar datos. Los visualizamos y los guardamos como fichero Respuestas.RData.
- Contraste la hipótesis de que el número de vehículos en el hogar es independiente del número de miembros de la unidad familiar, a un nivel de significación del 5%. Como las variables anteriores son numéricas, hemos de convertirlas en factores previamente (Menú Datos/Modificar):

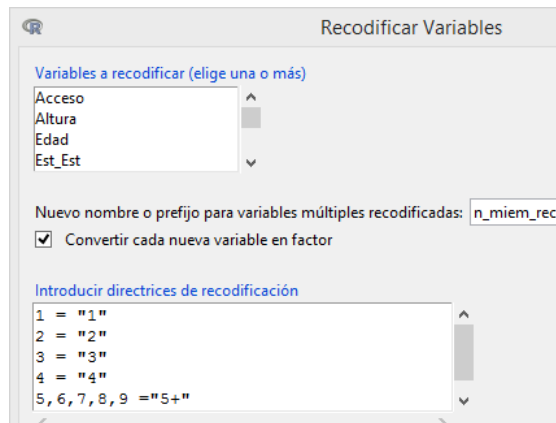


En el menú estadísticos/tablas de contingencia/tablas de doble entrada, obtenga el test chi-cuadrado y la tabla con frecuencias absolutas. Observe si las frecuencias de las tabla son o no inferiores a 5.



En el caso de que lo sean, tendremos que recodificar las variables, de la forma:





Repita el test con las nuevas variables recodificadas.

---

## Ejercicio 2

Compruebe con un nivel de significación del 5% si la variable número de vehículos en el hogar de los alumnos podría provenir de una distribución binomial.

Para obtener el contraste Chi-Cuadrado de bondad de ajuste de una variable a una distribución discreta, ejecutamos las siguientes órdenes:

```
library(vcd)
Hist(Respuestas$n_veh, col=terrain.colors(10))
goodfit(Respuestas$n_veh, type = "binomial", method = "MinChisq")
summary(goodfit(Respuestas$n_veh, type = "binomial", method = "MinChisq"))
```

---

## Ejercicio 3

Compruebe con un nivel de significación del 5% si la variable Altura de los alumnos podría provenir de una distribución normal.

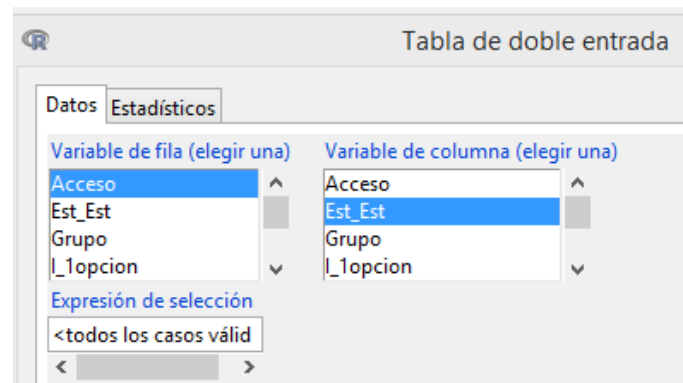
Accedemos al menú Estadísticos/Resúmenes/Test de normalidad de Shapiro-Wilk, recordando que la hipótesis nula en este contraste es la normalidad de la distribución.

---

## Ejercicio 4

Realice el test chi-cuadrado para contrastar a un nivel de significación del 5%, si la forma de acceso a la titulación, es independiente del hecho de tener conocimientos previos de estadística.





- a) Analice la tabla de frecuencias, por si existen valores muy pequeños ( $<5$ ). En caso afirmativo, soluciones el problema.
- b) Obtenga el estadístico Chi-cuadrado o test de Fisher, si finalmente se queda con una tabla de dimensiones 2x2.

