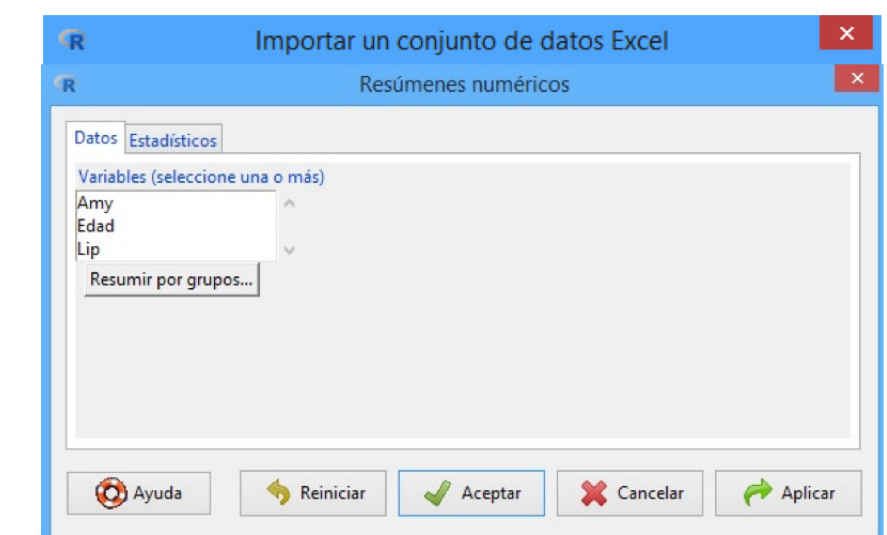


CASO PRÁCTICO: TEST DE HIPÓTESIS

En este ejercicio vamos a utilizar una base de datos que contiene los resultados de pacientes que acuden a urgencias con sospecha de pancreatitis. A todos se les realiza una analítica al entrar en urgencias y se determinan los niveles de amilasa y de lipasa, y también se recoge si se le realizó o no una prueba de imagen.

Para el estudio comenzamos cargando el archivo "Pancre". Para ello lo que vamos a hacer es abrir el programa de R, cargamos el Rcmdr, bien con el comando `>library(Rcmdr)` o bien desde el menú de la consola, con las pestañas de R [paquetes](#) → [Cargar paquete](#), seleccionamos Rcmdr y nos aparecerá la ventana de R Commander. En esta nos vamos a la pestaña de [Datos](#) → [Importar datos](#) → [Desde un archivo de Excel...](#) damos nombre a nuestro archivo (pancre) y nos aparecerá el explorador, buscamos nuestro archivo "Pan-cre" y aceptamos la carga del archivo.



Lo primero que vamos a hacer es resumir los datos que nos dan, para ello Estadísticos → Resúmenes → Resúmenes numéricos, una vez que hacemos esto, se nos abre un cuadro con dos pestañas: Datos y Estadísticos. En la pestaña Datos podemos seleccionar las variables cuantitativa a ordenar, y en el botón "Resumir por grupos..." nos permite hacer un resumen de la variable seleccionada en función de una variable cualitativa. La pestaña Estadísticos nos permite configurar el resumen que R nos realice del modo que queramos, definiendo el cuartil de interés, etc. Veamos un ejemplo, si selecciono "Amy" y la agrupo por sexo, los resultados que R me muestra son los siguientes:

```
Rcmdr> numsummary(pancre[, "Amy"], groups=pancre$Sexo, statistics=c("mean", "sd",
Rcmdr+   "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
      mean      sd IQR 0% 25% 50% 75% 100% data:n
F    484.9189 499.4011 697 40 131 251 828 1867    37
M    535.3415 501.0053 495 64 202 386 697 2293    41
```

Donde nos muestra que los valores medios de Amilasa en el grupo de hombres fue mayor que en el de mujeres (535 en hombres frente a 484 en mujeres), con una desviación estándar similar, IQR es el rango intercuartilico y los otros valores son los datos de los diferentes cuartiles. Hagamos nuevamente una prueba, pero esta vez para conocer qué valores de amilasa y lipasa tenían los individuos en función de si se realizó o no la prueba de imagen. Seleccionamos "Amy" y "Lip" y agrupamos por pruebas de imagen y obtenemos:

```
Rcmdr> numSummary(pancre[,c("Amy", "Lip")], groups=pancre$Pruebas.de.imagen,
Rcmdr+   statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

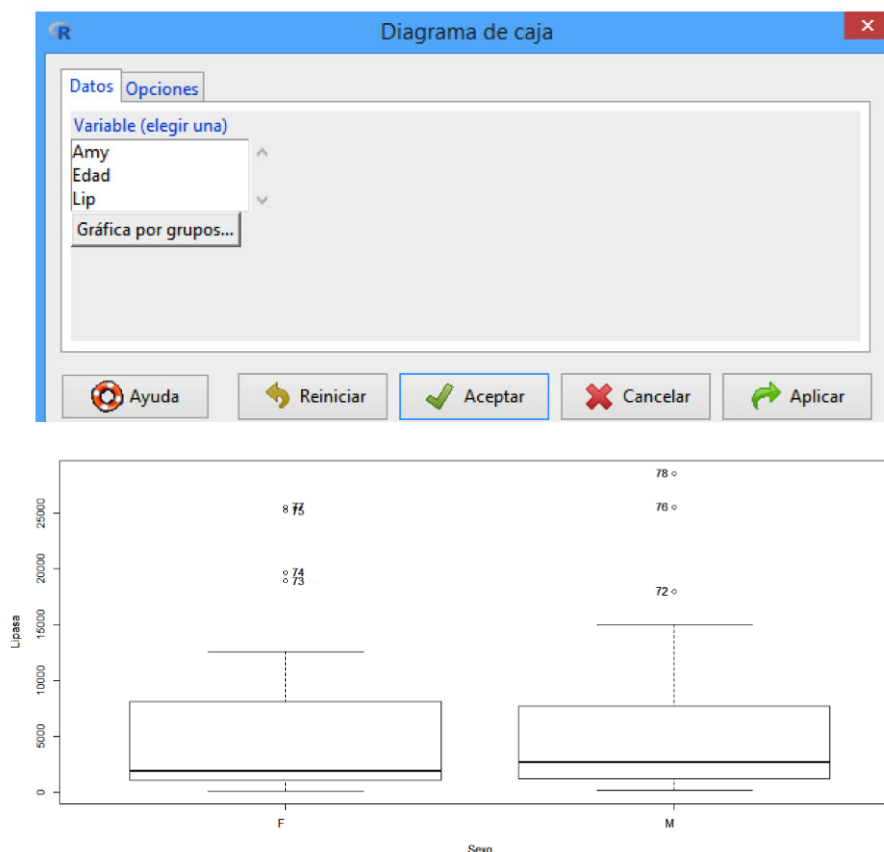
Variable: Amy

	mean	sd	IQR	0%	25%	50%	75%	100%	n
NO	562.9318	530.0241	681.75	40	173.25	379.5	855	2293	44
SI	444.7647	451.2790	401.00	40	151.00	245.5	552	1837	34

Variable: Lip

	mean	sd	IQR	0%	25%	50%	75%	100%	n
NO	6046.136	7022.554	7267.50	120	1102.75	2695	8370.25	28544	44
SI	4678.412	6463.519	4603.75	112	1101.75	1450	5705.50	25527	34

Pero si queremos tener los resultados de forma gráfica, que es más visual, nos vamos a [Gráficas](#)→[Diagrama de cajas](#) y aquí agrupamos por sexo, en esta ventana tenemos nuevamente dos pestañas Datos y Opciones, en Opciones podemos configurar si queremos que la gráfica aparezca con algún nombre o qué hacer con los puntos que se salgan del diagrama, etc. Para poder ver gráficamente la agrupación de Lipasa en función del sexo seleccionamos "Lip" y en "Gráfica por grupos" seleccionamos "Sexo". Ahora pulsamos la pestaña opciones y al eje de las X lo llamaremos Sexo y al eje de las Y lo llamaremos Lipasa y deberíamos obtener esto:

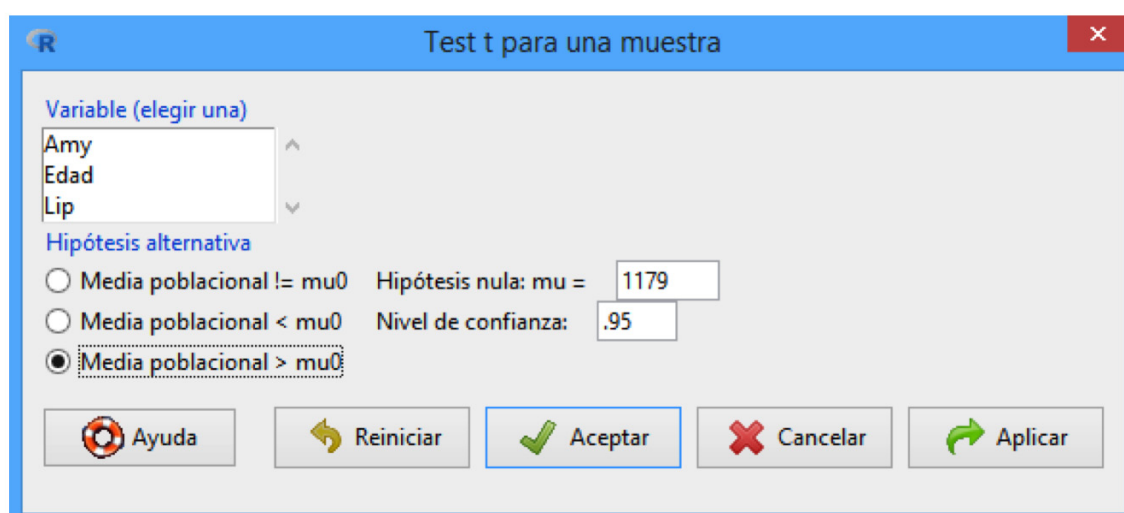


Una vez que hemos recordado un poco de estadística descriptiva vayamos al apartado de inferencia estadística, en el párrafo anterior hablábamos de cómo calcular la media muestral de nuestros datos. Pero la inferencia estadística la usamos para extrapolar el valor obtenido en nuestra muestra a una población de estudio. Uno de los criterios en los que se apoya el diagnóstico de pancreatitis es la elevación de las enzimas pancreáticas tres veces por encima del límite superior de referencia. Si el valor de referencia que consideramos para la Lipasa es 393 U/L, nuestro punto de corte para pancreatitis es de 1179 U/L. Queremos saber si el valor medio en nuestra muestra es mayor que 1179 U/L, para ello plantearemos el contraste de hipótesis de la siguiente manera.

$$H_0 \leq 1179 \text{ U/L}$$

$$H_1 > 1179 \text{ U/L}$$

Para ello nos vamos a [Estadísticos Medias Test t para una muestra..](#) y se nos abre un cuadro similar a este:



En este caso vamos a realizar un contraste de hipótesis muy sencillo, para ello seleccionaremos mi variable Lipasa primero, luego, para comprobar si nuestro valor observado es superior a tres veces el límite superior de referencia, asumiendo una probabilidad de error del 5 % ($\alpha=0,05$), donde pone "Hipótesis nula.mu" colocamos el valor que queremos contrastar (en este caso 1179). Seleccionamos la opción "Media poblacional > mu0", al tratarse de un contraste de hipótesis de una sola cola, debido a que queremos saber si nuestro valor medido supera o no un valor esperado, pero ya lo explicaremos con detalle en el tema próximo. Y en "Nivel de confianza" lo dejamos en ".95" ya que esto le indica al programa que el IC es al 95 %

```
data: Lip
t = 5.5669, df = 77, p-value = 1.816e-07
alternative hypothesis: true mean is greater than 1179
95 percent confidence interval:
 4172.647      Inf
sample estimates:
mean of x
 5449.949
```

Como se puede observar el valor medio obtenido en nuestra muestra (5449,9 U/L), es mayor a tres veces el límite superior de referencia con una $p=1,81 \times 10^{-7}$. Y nuestro intervalo de confianza nos indica que la media de la población es superior a 4172 (IC $\rightarrow 4172, +\infty$).

¿Qué resultados obtendríamos si hiciésemos el mismo planteamiento, pero con la Amilasa (cuyo límite superior de referencia es 115 UI/L) y con un intervalo de confianza al 99 %?