



## Guía 8 - Análisis de datos científicos en R

1. El comando `system.time()` sirve para calcular el tiempo que tarda en ejecutar una instrucción de R (ver la ayuda del comando). En las próximas dos preguntas, usarlo para comparar diferencias de velocidad de ejecución de distintos comandos.
  - a. ¿Cuánto demoran los resultados de `mean()` y `mean.default()` con 100 observaciones y con 10000? Calcular el tiempo que demora cada bucle. Investigar con la ayuda el porqué de la diferencia.
  - b. Comparar la velocidad de `apply(x, 1, sum)` con `rowSums(x)` para un vector atómico `x` de tipo *numeric*, de distintos tamaños (por ejemplo, de  $10^1$ ,  $10^2$ ,  $10^3$  y  $10^4$  (si su máquina aguanta, si no usar base 2:  $2^1$ ,  $2^2$ ,  $2^3$  y  $2^4$ . Recordar que  $2^4$  es  $2^4$ ).
2. El comando `rpois()` nos devuelve una ‘tirada’ de variable aleatoria (como cuando tiramos una moneda) de una distribución de Poisson (ver la ayuda del comando). El código a continuación simula el desempeño de un *t-test* para datos que vienen de una distribución de este tipo.

```
trials <- replicate(  
  100,  
  t.test(rpois(10, 10), rpois(7, 10)),  
  simplify = FALSE  
)
```

Usar `sapply()` y una función (que puede tener nombre o que puede ser anónima, o sea, sin nombre) para extraer el *p-value* de cada “tirada”.

3. **POSGRADO** – Los datos *Dnase*, del paquete *datasets*, contienen información de ensayos bioquímicos de una proteína. Específicamente, contiene un índice que indica el número de experimento, concentración de la proteína y densidad.
  - a. Calcular, usando `tapply()`, la concentración media para los distintos experimentos.
  - b. Usar el comando `which()` para identificar para qué experimento la densidad media es máxima.