

## Segundo Informe de Bioestadística.

### Sobre el informe:

- El informe se debe entregar el **lunes 28 de Noviembre de 8:00 a 22:00 horas**, como una tarea en EVA. Solo se aceptarán archivos en formato .pdf. Lo entregará **un solo** representante de cada grupo. Se debe poner el número del grupo en el nombre del archivo. Ejemplo: **Grupo27.pdf**
- Se deben incluir los script utilizados (en el mismo documento).
- Se podrán utilizar a lo sumo **6 carillas** incluyendo el gráfico solicitado.
- Deberá contener un título, fecha, número de grupo y autores (nombre y cédula de identidad).
- Se evaluará: prolijidad del informe, utilización de las funciones presentadas en los tutoriales, calidad de los gráficos, conclusiones.

En este informe analizaremos resultados de experimentos del largo de vida en *Drosophila*, enmarcados en la línea de Investigación: <sup>1</sup>

*“Neuroprotección frente a enfermedades neurodegenerativas como el Parkinson y el Alzheimer: modelos de neurodegeneración en Drosophila”*,

del Departamento de Biología del Neurodesarrollo del Instituto de Investigaciones Biológicas Clemente Estable ([Enlace al proyecto](#)).

La mosca de la fruta *Drosophila melanogaster* es ampliamente utilizada como organismo modelo en biología, dado que posee un reducido número de cromosomas, breve ciclo de vida y más de la mitad de los genes asociados a enfermedades humanas que se conocen tienen un homólogo identificable en el genoma de la mosca.

Estos datos fueron obtenidos durante ensayos de “largo de vida”, en los que se evaluó el largo de vida promedio de distintas cepas moscas, cada una con un genotipo diferente. En estos experimentos se busca identificar genes con potencial neuroprotector. La hipótesis que se busca poner a prueba con los experimentos es que las moscas con cierto genotipo tienen alterado su largo de vida promedio.

### Parte 1: Cargar los datos

- (a) Descargar el archivo *Datos-largo-de-vida-Drosophila.xls* que contiene los datos que analizaremos en este informe y observar los datos. Notar que cada columna contiene la **cantidad** de moscas con el genotipo correspondiente a esa columna que van quedando luego de transcurridos los días indicados en la columna 1. Notar también que solo se contaron las moscas en días impares.
- (b) Cargar los datos a R. *Sugerencia*: ver el vídeo correspondiente al Laboratorio 4 en donde se explica cómo cargar datos a RStudio.

### Parte 2: Tiempos de vida de cada una de las moscas del experimento

Nos interesa obtener a partir de esos datos el tiempo de vida (en días) de cada una de las moscas del experimento. Por ejemplo, si la tabla correspondiente al Genotipo A fuera:

N moscas vivas	
día	Genotipo A
1	20
3	17
5	15
7	11
9	3
11	0

<sup>1</sup>Agradecemos a Constanza Silvera y María José Ferreiro, quienes nos proporcionaron los datos para este informe.

entonces sabríamos que en el frasco en el que se analizaron las moscas con el Genotipo A habían inicialmente 20 moscas, de las cuales:

- $3 = 20 - 17$  moscas que vivieron 1 día,
- $2 = 17 - 15$  moscas que vivieron 3 días,
- $4 = 15 - 11$  moscas que vivieron 5 días,
- $8 = 11 - 3$  moscas que vivieron 7 días,
- 3 moscas que vivieron 9 días.

Si le llamamos  $a_1, a_2, \dots, a_{20}$  a la muestra ordenada en la que  $a_i$  representa el tiempo de vida (en días) de la mosca  $i$ , entonces la muestra correspondiente a la tabla anterior es:

$i$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$a_i$	1	1	1	3	3	5	5	5	5	7	7	7	7	7	7	7	7	9	9	9

- (a) Crear un algoritmo que obtenga a partir de la columna **Genotipo A** la muestra ordenada:

$$MuestraA = (a_1, a_2, \dots, a_{57})$$

correspondiente al tiempo de vida  $a_i$  de cada mosca con el Genotipo A.

- (b) Idem para el **Genotipo B**. Llamarle a esta muestra:

$$MuestraB = (b_1, b_2, \dots, b_{58})$$

correspondiente al tiempo de vida  $b_i$  de cada mosca con el Genotipo B.

- (c) Idem para el **Genotipo C**. Llamarle a esta muestra:

$$MuestraC = (c_1, c_2, \dots, c_{66})$$

correspondiente al tiempo de vida  $c_i$  de cada mosca con el Genotipo C.

- (d) Idem para el **Genotipo D**. Llamarle a esta muestra:

$$MuestraD = (d_1, d_2, \dots, d_{69})$$

correspondiente al tiempo de vida  $d_i$  de cada mosca con el Genotipo D.

- (e) Idem para el **Genotipo E**. Llamarle a esta muestra:

$$MuestraE = (e_1, e_2, \dots, e_{66})$$

correspondiente al tiempo de vida  $e_i$  de cada mosca con el Genotipo E.

### Parte 3: Estimación del tiempo medio de vida de moscas de cada genotipo

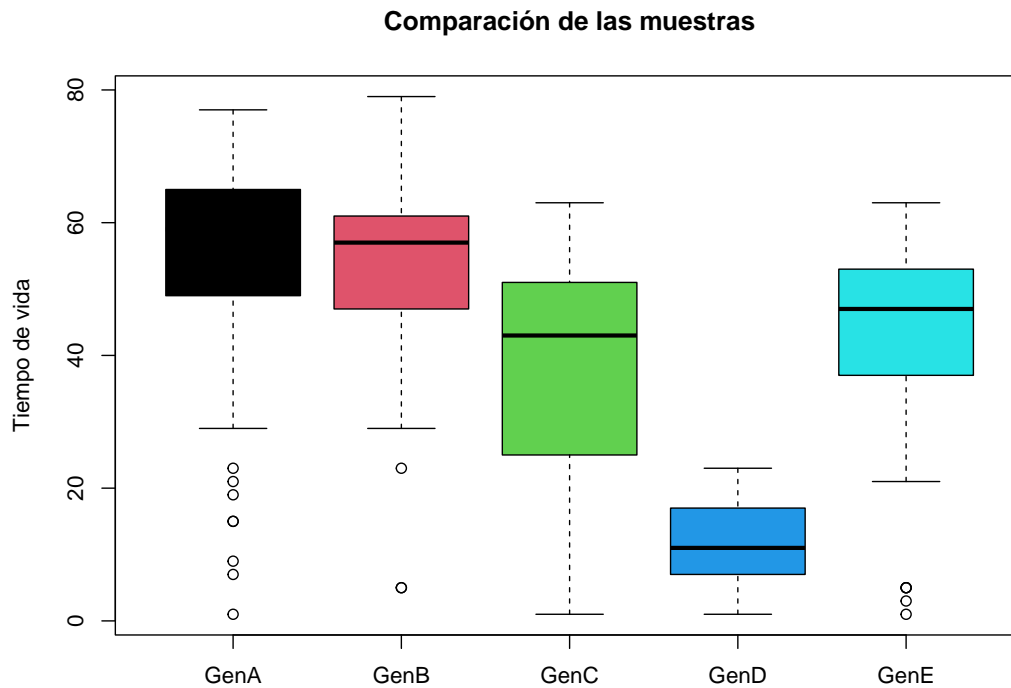
En esta parte estimaremos el tiempo medio de vida de moscas de cada genotipo, a partir de las muestras obtenidas en la parte anterior.

- (a) A partir de la *MuestraA*, obtener un estimador de  $\mu_A$ , el tiempo medio de vida de una mosca con el Genotipo A. Construir también un intervalo de confianza aproximado a nivel 95% para  $\mu_A$ .
- (b) A partir de la *MuestraB*, obtener un estimador de  $\mu_B$ , el tiempo medio de vida de una mosca con el Genotipo B. Construir también un intervalo de confianza aproximado a nivel 95% para  $\mu_B$ .
- (c) A partir de la *MuestraC*, obtener un estimador de  $\mu_C$ , el tiempo medio de vida de una mosca con el Genotipo C. Construir también un intervalo de confianza aproximado a nivel 95% para  $\mu_C$ .
- (d) A partir de la *MuestraD*, obtener un estimador de  $\mu_D$ , el tiempo medio de vida de una mosca con el Genotipo D. Construir también un intervalo de confianza aproximado a nivel 95% para  $\mu_D$ .

- (e) A partir de la *MuestraE*, obtener un estimador de  $\mu_E$ , el tiempo medio de vida de una mosca con el Genotipo E. Construir también un intervalo de confianza aproximado a nivel 95% para  $\mu_E$ .
- (f) A partir de las partes anteriores, ¿para qué genotipo el tiempo de vida esperado es mayor y para cuál es menor?, ¿pueden sacar alguna otra conclusión?.

**Parte 4: Comparación de las muestras**

Construir un boxplot para cada una de las muestras obtenidas en la Parte 2. Presentar todos los boxplots en un mismo gráfico, debería quedar similar al que se muestra a continuación:



¿Qué conclusiones se pueden sacar a partir de esta comparación?

**Parte 5: Comparación de los diferentes genotipos respecto del Genotipo E, utilizado como control**

En esta parte se compararán todas las muestras con la muestra correspondiente al Genotipo E para ver si ocurre un cambio significativo con respecto al mismo.

- (a) Realizar un test de comparación de dos muestras de Kolmogorov-Smirnov para testar si se puede asumir que las muestras *MuestraA* y *MuestraE* obtenidas en la Parte 2 provienen de variables aleatorias con la misma distribución. Escribir cuáles son las hipótesis, indicar el valor del estadístico para el test, indicar el valor del p-valor y tomar una decisión a nivel  $\alpha = 0.05$ . *Sugerencia:* mirar el vídeo correspondiente al Laboratorio 5.
- (b) Idem para las muestras *MuestraB* y *MuestraE*.
- (c) Idem para las muestras *MuestraC* y *MuestraE*.
- (d) Idem para las muestras *MuestraD* y *MuestraE*.

**Ejercicio 6**

*Conclusiones:* Realizar todas las conclusiones que les parezcan pertinentes respecto a los datos observados. Si desean realizar otro tipo de test para estas muestras, incluirlos en esta parte.