

Proiect

Problema 1¹

Populațiile diverselor specii de *Drosophila melanogaster* (musculița de vin/oțet sau a fructelor) prezintă variații ale formei aripilor. Forma aripilor este cuantificată prin măsurarea locațiilor unde nervurile se intersectează între ele sau se intersectează cu marginile acestora (vezi Figura 1). Se obține un total de 15 astfel de puncte de intersecție, fiecare având două coordonate, prin urmare forma aripilor este caracterizată de un vector din \mathbb{R}^{30} .

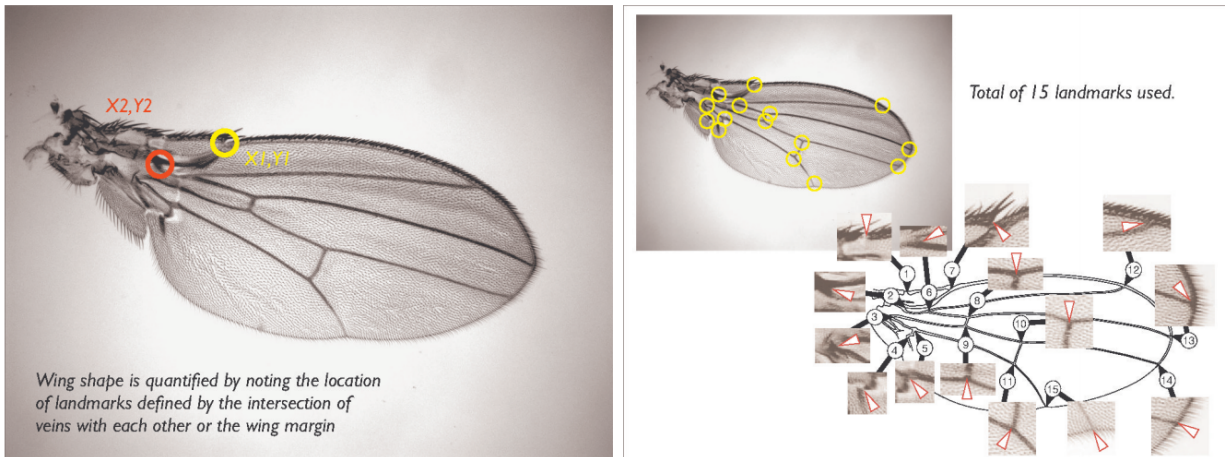


Figura 1: Cuanatificarea formei unei aripi

Ne propunem să investigăm dacă forma aripilor poate diferenția următoarele trei specii de musculițe: *D. Mauritiana*, *D. Sechellia* și *D. Simulans*. Pentru aceasta considerăm setul de date `wing.dat` care conține cuantificarea formei aripilor a 138 de musculițe de oțet diferite, fiecare făcând parte dintr-una din cele 3 specii enumerate mai sus astfel: de la 1 la 42 din specia *D. Mauritiana*, de la 43 la 90 din specia *D. Sechellia* iar restul din specia *D. Simulans*. Fie \mathbf{Y} matricea observațiilor și \mathbf{X} matricea *centrată* obținută prin

$$\mathbf{X}_{i,j} = \mathbf{Y}_{i,j} - \bar{\mathbf{Y}}_{\cdot,j}$$

1. Calculați matricea $\mathbf{C} = \frac{1}{137} \mathbf{X}^T \mathbf{X}$. Ce reprezintă această matrice ?
2. Calculați λ_{\max} și \mathbf{v}_{\max} cea mai mare valoare proprie a matricii \mathbf{C} și vectorul propriu corespunzător (folosiți funcția `eigen`).
3. Pentru fiecare musculiță i , calculați proiecția pe direcția \mathbf{v}_{\max} a vectorului centrat X_i (linia i din matricea \mathbf{X}) care cuantifică forma aripilor acesteia, $l_i = \langle X_i, \mathbf{v}_{\max} \rangle$. Acest scalar va fi folosit ca un cuantificator al formei aripilor pentru musculița i .
4. Verificați grafic și printr-un test adecvat dacă descriptorii l_i sunt repartizați normal pentru fiecare din cele trei specii în parte.
5. Folosiți un test adecvat pentru a verifica dacă mediile $\mu_{Mauritiana}$ și $\mu_{Sechellia}$ sunt egale, unde $\mu_{Mauritiana}$ reprezintă media descriptorilor l_i pentru specia Mauritiana și respectiv $\mu_{Sechellia}$ este media lui l_i pentru specia Sechellia. Repetați procedura și pentru celelalte două perechi.
6. Verificați dacă cele trei medii sunt egale între ele.

¹Raportul poate fi scris în Word sau L^AT_EX (pentru ușurință recomand folosirea pachetului `rmarkdown` din R - mai multe informații găsiți pe site la secțiune *Link-uri utile*). Toate simulările, figurile și codurile folosite trebuie incluse în raport. Se va folosi doar limbajul R.

Să presupunem că vrem să studiem efectul raportului dintre mărimea aripilor pe mărimea corpului pentru musculița de oțet și abilitatea acesteia de a scăpa de un păianjen. Testăm ipoteza că musculițele cu un corp mai greu relativ la mărimea aripilor sunt mai vulnerabile. Mai multe musculițe cu un raport scăzut între mărimea aripilor și a corpului (clasificate ca și *ușoare*) au fost închise într-o cutie împreună cu 2 păianjeni pentru o perioadă de 8 zile. Un experiment similar a avut loc pentru musculițe cu un raport ridicat (clasificate ca *grele*). Numărul de musculițe care au fost consumate de păianjeni este prezentat în Tabelul 1:

Tabelul 1: numărul de musculite consumate

	Usoare	Grele	Total
$\leq 48h$	16	27	43
48-96h	6	10	16
96-144h	7	6	13
144-192h	7	12	
Supraviețuit după 192h	71	55	126
Total	105	105	210

7. Verificați dacă rata de incidență a consumului rămâne constantă de-a lungul timpului. Calculați p-valoarea prin trei teste diferite.