

# Análise de Sobrevivência - Seminário 1

Matheus T. Baumgartner

2024

## Banco de Dados

Os dados utilizados neste trabalho foram obtidos de:

- Rijn e Wäckers (2016) - Nectar accessibility determines fitness, flower choice and abundance of hoverflies that provide natural pest control. *Journal of Applied Ecology*, 53, 925-933.

E estão disponíveis em <https://datadryad.org/stash/dataset/doi:10.5061/dryad.d37n0>.

Os dados representam observações da sobrevivência de indivíduos de mosca (*Episyrphus balteatus* (De Geer, 1776); Figura 1) diante de diferentes espécies de planta como recurso alimentar. A maioria das plantas selecionadas fazem parte da dieta natural de *E. balteatus*, embora algumas outras plantas também foram consideradas no estudo.

O procedimento experimental se deu, de forma resumida, através da inoculação de indivíduos de *E. balteatus* com um dia de vida em gaiolas feitas de gaze com a presença de uma espécie de planta como recurso. Então, foi medido o número de dias até os indivíduos morrerem de fome ou não serem mais avistados pelos observadores.

Os dados originais estão armazenados no arquivo **flower.xlsx** na aba **data**. As variáveis originais são:

- **year**: (num) ano de realização do ensaio experimental
- **season**: (fator) estação do ano (A: antes de julho, B: de julho a agosto, C: depois de agosto)
- **cage**: (fator) identificador da gaiola
- **food species**: (fator) espécie de planta disponibilizada como recurso alimentar
- **sex**: (fator) sexo do indivíduo de *E. balteatus*
- **introduce**: data de quando o indivíduo foi introduzido na gaiola
- **lastdate**: data de quando o indivíduo morreu ou não foi mais avistado
- **time**: (num) tempo estimado (em dias) de sobrevivência dos indivíduos de *E. balteatus*
- **dead**: (binário) indicadora de morte (1: morte, 0: desaparecimento ou censura)

## Análise Descritiva

O banco de dados original é composto por observações de 842 indivíduos de *E. balteatus* dos quais 93,82% são observações do evento (morte do indivíduo). Para as análises, os tratamentos do tipo controle onde apenas água ou sacarose estavam disponíveis para a alimentação dos indivíduos foram removidos. O tempo máximo de sobrevivência atingido por um indivíduo foi 34 dias.

Como agrupamento, as 30 espécies de plantas foram classificadas e agrupadas de acordo com a família a qual pertenciam (variável **family**), resultando em um agrupamento de 8 famílias. Os resultados descritivos estão resumidos na Tabela 1.



Figure 1: Fotografia de *E. balteatus*. Fonte: Picture Insect.

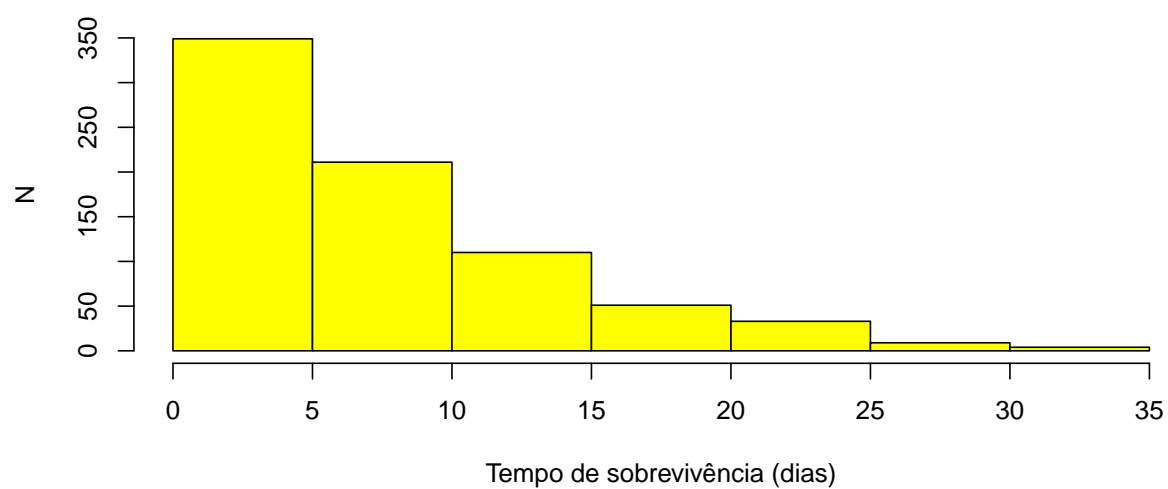


Figure 2: Histograma do tempo de sobrevivência (em dias) dos indivíduos de *E. balteatus*.

Table 1: Resultados das estatísticas descritivas para os indivíduos de *E. balteatus*.

| Familia         | Nspp | Nind | Mortes | Censura | PropCens |
|-----------------|------|------|--------|---------|----------|
| Apiaceae        | 6    | 150  | 139    | 11      | 0.0733   |
| Asteraceae      | 15   | 350  | 323    | 27      | 0.0771   |
| Boraginaceae    | 1    | 56   | 54     | 2       | 0.0357   |
| Caryophyllaceae | 1    | 13   | 13     | 0       | 0.0000   |
| Fabaceae        | 4    | 88   | 86     | 2       | 0.0227   |
| Hydrophyllaceae | 1    | 25   | 25     | 0       | 0.0000   |
| Polygonaceae    | 1    | 69   | 68     | 1       | 0.0145   |
| Ranunculaceae   | 1    | 16   | 15     | 1       | 0.0625   |

## Função de Sobrevivência

A função de sobrevivência  $S(t)$  foi estimada com o método Kaplan-Meier utilizando a função `survfit()` do pacote `survival`, para cada uma das 8 famílias de plantas, e os resultados estão na Figura 2. O tempo médio de sobrevida dos indivíduos variou entre 2,44 dias (Hydrophyllaceae) e 13,35 dias (Apiaceae), enquanto o tempo mediano de vida variou entre 3 dias (Hydrophyllaceae) e 12 dias (Boraginaceae e Caryophyllaceae).

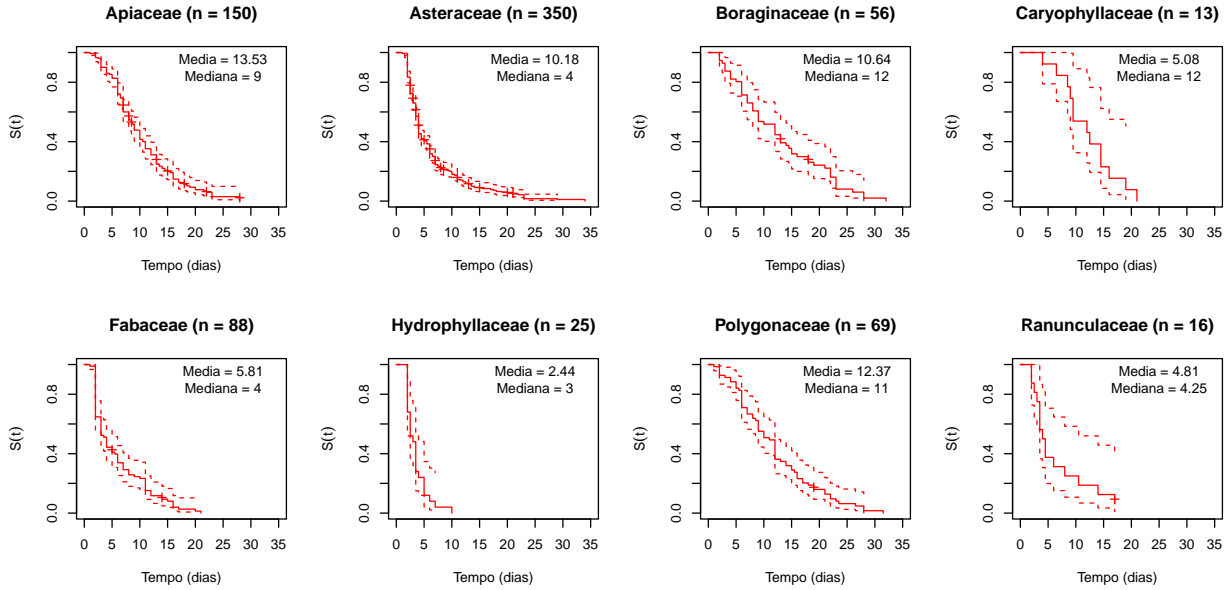


Figure 3: Funções de sobrevivência  $S(t)$  estimadas através do método Kaplan-Meier para os indivíduos de *E. balteatus* para as condições experimentais de cada uma das 8 famílias de plantas como recurso alimentar. As linhas pontilhadas representam os intervalos de confiança das estimativas.

## Função de Risco

A função de risco  $h(t)$  foi estimada utilizando a função `muhaz()` do pacote de mesmo nome. Os resultados estão na Figura 3.

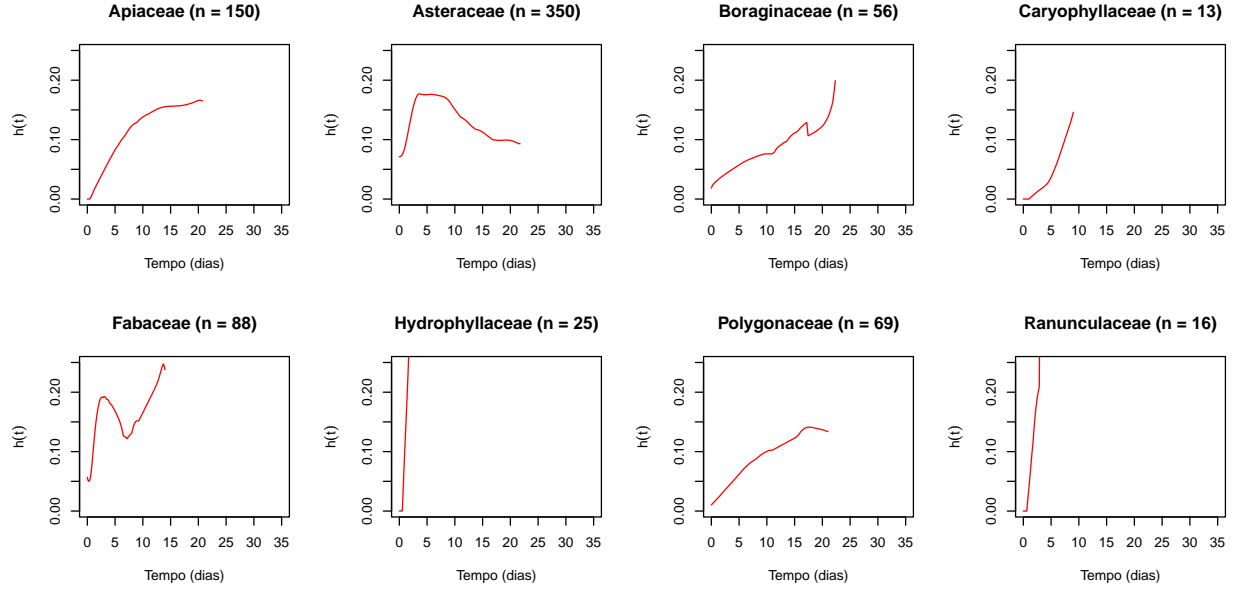


Figure 4: Funções de risco  $h(t)$  estimadas para os indivíduos de *E. balteatus* para as condições experimentais de cada uma das 8 famílias de plantas como recurso alimentar.

## Modelos Paramétricos

O ajuste da função de sobrevivência e da função de risco foi realizado utilizando 7 modelos paramétricos (pacotes):

- Exponencial (`parmsurvfit::fit_data()`)
- Weibull (`parmsurvfit::fit_data()`)
- Gama (`flexsurv::flexsurvreg()`)
- Log-Normal (`parmsurvfit::fit_data()`)
- Log-logístico (`parmsurvfit::fit_data()`)
- Gompertz (`flexsurv::flexsurvreg()`)
- Gama generalizada (`flexsurv::flexsurvreg()`)

Em termos da função de sobrevivência, o modelo que mais se ajustou às curvas de sobrevivência estimadas pelo Kaplan-Meier foi o modelo Gompertz. Os outros modelos estimaram curvas muito distantes do modelo não-paramétrico. No entanto, em relação à função de risco, os resultados foram consideravelmente divergentes. Por conta do formato da função de risco empírica, o ajuste foi dificultado. Para algumas famílias como Apiaceae, Asteraceae, Caryophyllaceae, Polygonaceae e até para Fabaceae, os modelos ajustados mais visualmente plausíveis foram Weibull, Gama, Gama generalizada e log-Normal, porém sem performance superior entre eles.

## Interpretação dos resultados

As curvas de sobrevivência mostraram diferentes padrões para a sobrevivência dos indivíduos de *E. balteatus*, dependendo do recurso alimentar disponível. As maiores sobrevivências (em média) foram observadas quando as plantas pertenciam às famílias Apiaceae, Polygonaceae, Boraginaceae e Asteraceae, com tempos médios de sobrevida acima dos 10 dias de experimento. Em contrapartida, indivíduos com recurso alimentar pertencente

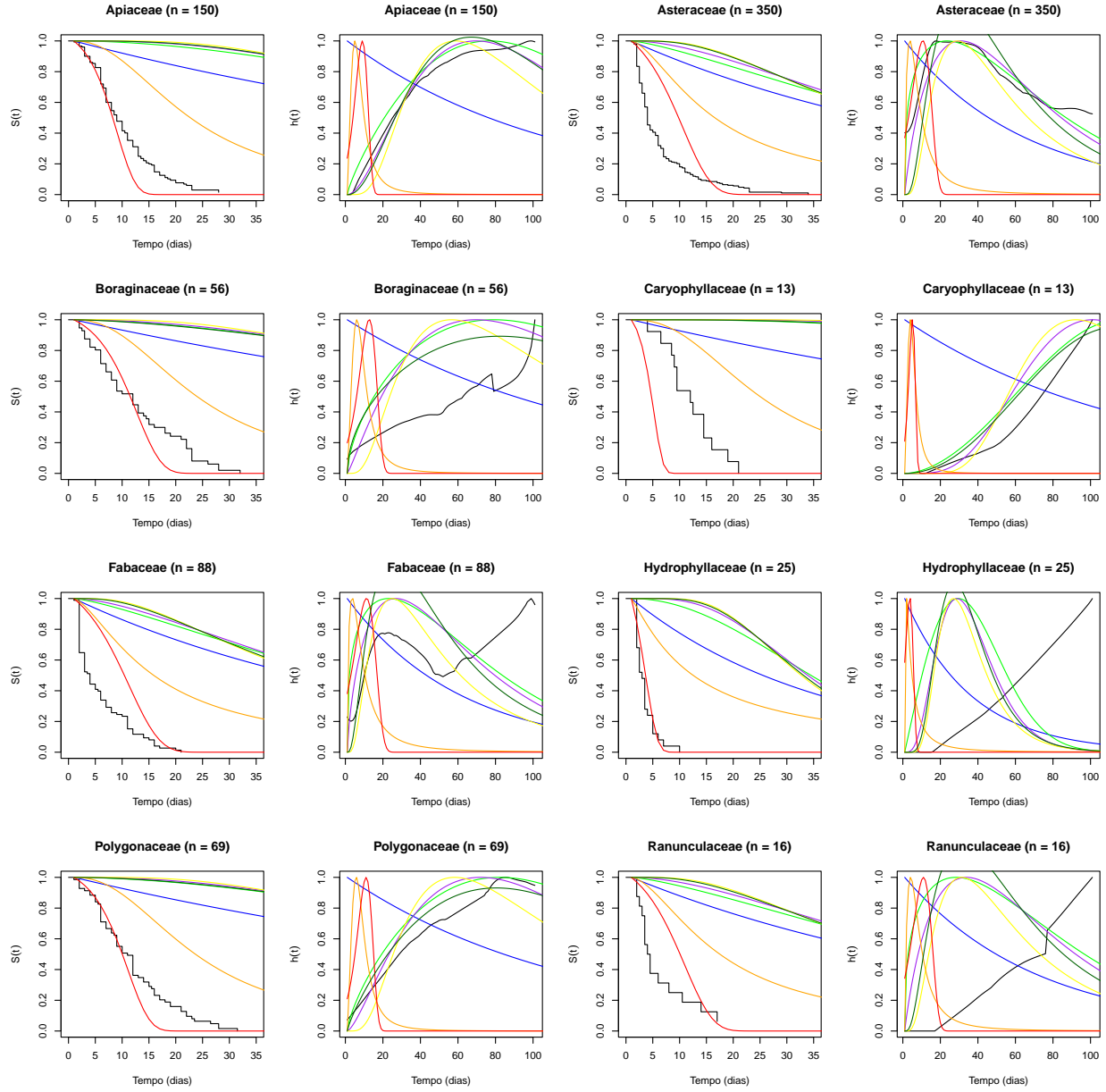


Figure 5: Funções de sobrevivência  $S(t)$  e de risco  $h(t)$  estimadas para os indivíduos de *E. balteatus* para as condições experimentais de cada uma das 8 famílias de plantas como recurso alimentar, utilizando 7 modelos paramétricos: exponencial (azul), Weibull (verde), Gama (roxo), log-Normal (amarelo), log-logístico (laranja), Gompertz (vermelho) e Gama generalizada (verde escuro).

às famílias Hydrophyllaceae, Fabaceae e Ranunculaceae apresentaram sobrevivência menor. É importante ressaltar o contraste entre as estimativas de sobrevida média e mediana para a família Asteraceae.

A maior dificuldade foi com relação aos modelos paramétricos. Embora o ajuste do modelo Gompertz tenha fornecido as melhores estimativas da função de sobrevivência, a função de risco foi modelada de forma muito divergente dos valores observados. Muito provavelmente isso se deve ao fato de que as funções empíricas de risco crescem indefinidamente para quatro das oito famílias. À exceção de Asteraceae e ligeiramente Apiaceae, todas as outras famílias apresentaram funções de risco com comportamento pouco caracterizado por nenhum dos modelo paramétrico utilizados.

## Conclusão

A sobrevivência dos indivíduos de *E. balteatus* diante de diferentes recursos alimentares depende da família a qual pertence a planta que provê o recurso. Os ajustes de modelos probabilísticos para as funções de sobrevivência e de risco revelaram que o modelo Gompertz foi o mais adequado para modelar a sobrevivência, mas apresentou performance muito divergente para modelar o risco de morte dos indivíduos. Já os modelos Exponencial, Weibull, Gama e log-Normal apresentaram padrões opostos, com ajuste ruim para a função de sobrevivência, mas melhor ajuste para a função de risco.