BIO 708 QMEE
Final Project

Submitted 23 April, 2019
by David Filice & Susan Marsh-Rollo

Background:

In many species, the context of a social environment can have substantial effects on individual reproductive behaviours. In fruit flies, males that are housed in a vial with other competitors (perceive high-risk of sperm competition) transfer a greater volume of ejaculate compared to flies housed in a vial alone (perceive low-risk of sperm competition). The contents of this ejaculate can have negative effects on the fitness of females by reducing their fecundity and lifespan. In this project, we study how the effect of previous competition influences male reproductive behaviours, and how these behavioural differences influence the fitness (lifespan and offspring production) of the females that they mate with.

Prediction 1: Males previously housed with rival males will reduce the lifespan of their mates compared to males housed alone

Statistical approach: Here, we analyzed two models to estimate the influence of various factors on female survivorship. The first approach involved constructing a linear mixed model to analyze the square-root transformed lifespan, assuming a gaussian distribution. The second approach involved using a cox survival model to analyze the hazard of mortality over time.

Important assumptions and decisions: The first major decision was to decide which modelling approach we would use. In practice, we recognize the proper statistical approach is to decide your type of analysis before you analyze the data and to stick with it. However, for the sake of learning in this class (and having a bit more to write about), we decided to compare and contrast two different methods: a linear mixed model, and a cox proportional hazard mixed model. When constructing our models, we had to consider which random factors to include. Our maximal model included every interaction combination, but we decided to systematically reduce these interactions until we were left with no singular fits. Next, we had to decide if any transformations of our response variable were necessary. First we log transformed our data, then performed a box-cox transformation and calculated a lambda value very close to 0.5, so we decided to simply square-root transform our response. After a series of diagnostics, we found that the square-root transformation did indeed best fit the assumptions of a linear model, so we decided to proceed with our analysis.

Results:

Table 1: The fixed effects from our linear mixed model analyzing female lifespan. The coefficients were determined using the summary() function, while the 95% CI were calculated using the confint(,family = boot) function, and p-values using the lmerTest package.

Source of Variance	Coefficient Estimate	Upper & Lower 95% CI	Degrees of Freedom	p-value
sqrt(Lifespan)				
Intercept	4.530	4.968 4.092	37.638	<0.0001
Treatment (single)	0.312	0.894 -0.269	231	0.295
Population (wild)	1.923	2.505 1.341	231	<0.0001
Treatment x Population	0.155	0.978 -0.667	231	0.712

Table 2: The random effects from our linear mixed model analyzing female lifespan. Variance components were estimated using the summary() function, 95%CI using the confint(,family = boot) function, and p-values estimated using a permutation approach with 1000 repetitions.

Source of Variance	Variance	Upper & Lower 95% CI	p-value
Lifespan			
Line	0.041	0.542 0	0.215
Residual	2.659		

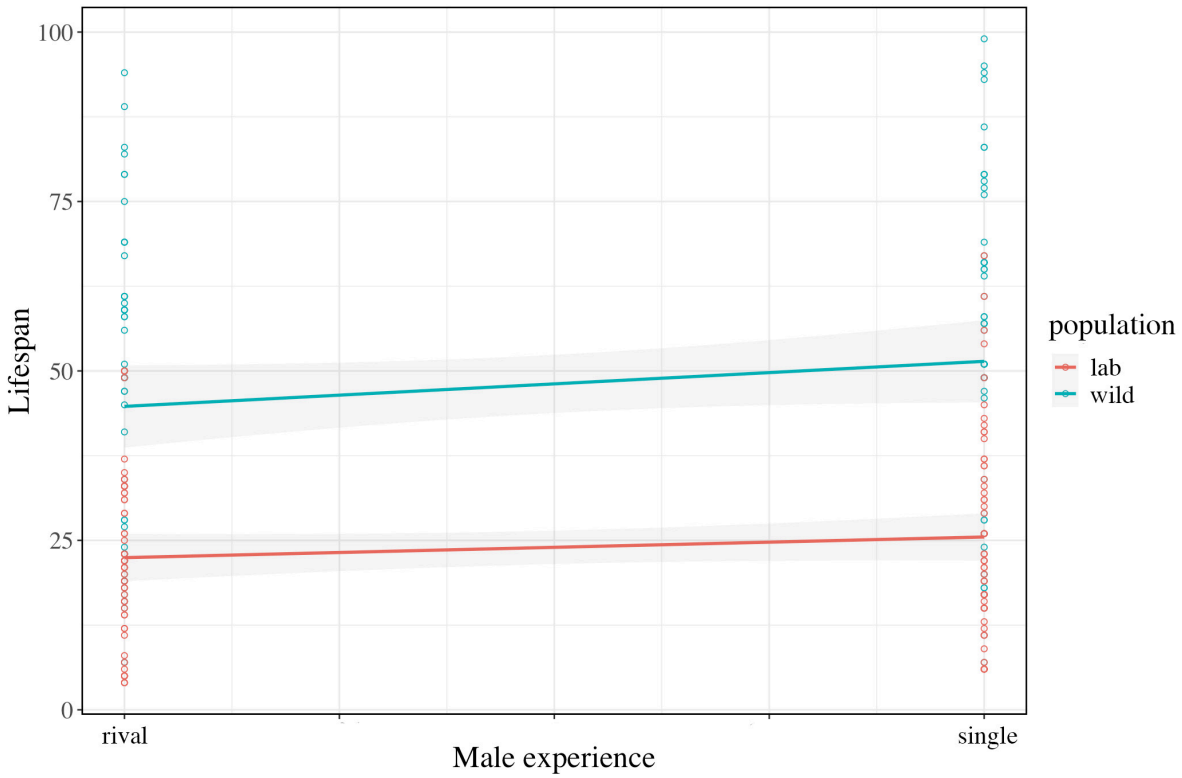


Figure 1: Regression lines predicted by our linear model for female lifespan. The shaded region around each line represents the 95% CI, and the points on each side of the figure represent the actual lifespan of each individual.

Table 3: The fixed effects of our cox survival model analyzing the hazard of female mortality. Test statistics were determined from using the summary() function.

Source of Variance	Coefficient Estimate	Standard Error	p-value
Surv(Lifespan)			
Treatment (single)	-0.230	0.184	0.21
Population (wild)	-1.335	0.203	<0.0001
Treatment x Population	-0.096	0.261	0.71

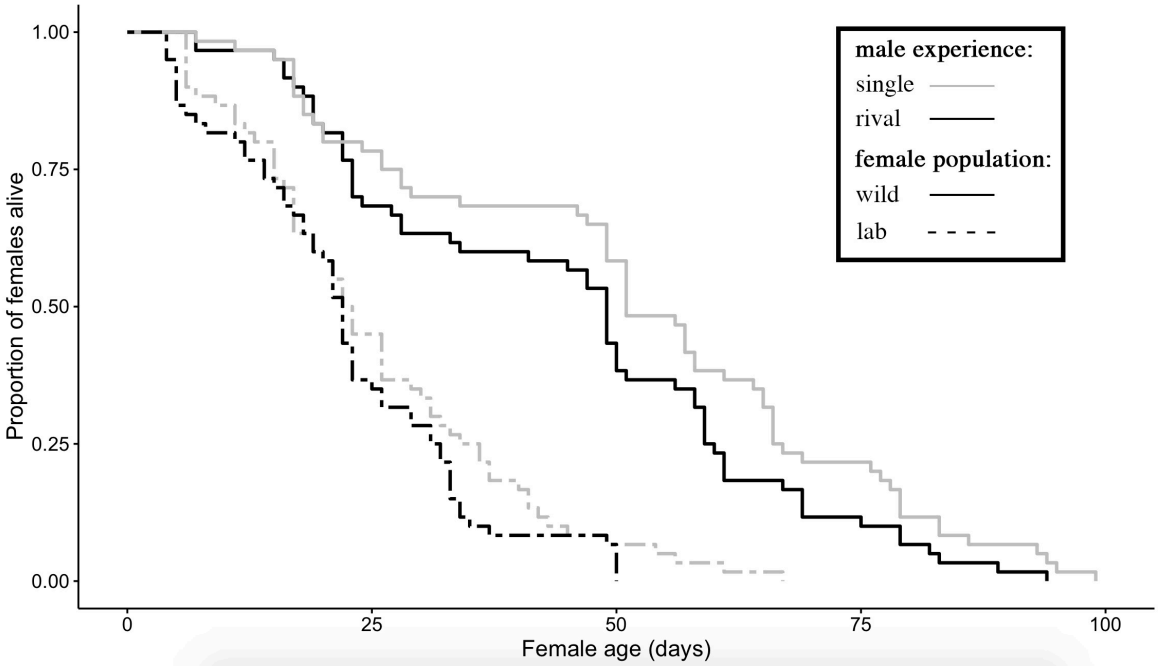


Figure 2: Survival curves demonstrating the hazard of mortality over time for each treatment combination

Conclusions: Although they look at slightly different response variables, our two models yield similar results. The biggest factor influencing female survivorship appears to be the base population she descended from, where females coming from the wild population live longer on average compared to females from the lab population. The effect of male experience has a much smaller effect, where females that mate with males housed alone live slightly longer on average compared to females mated with males housed with rivals, as we predicted. Finally, it appears that male genetic background has a very small effect on the survivorship of his mates.

Prediction 2: Males previously housed with rivals will reduce the lifetime offspring production of their mates compared to males housed alone

Statistical approach: We constructed a generalized linear mixed model analyzing offspring production as negative binomial response.

Important assumptions and decisions: Firstly, we had to consider the syntax for our fixed and random effects in our model. Initially we considered to input day as a fixed quadratic factor, but decided linear was indeed a better fit. Similar to our strategy above we reduced the number of possible interactions in all combinations one by one until we were left with no singular fits. In this case, even though including random effect of individual female produced a singular fit, we decided to keep it in our model as this is the most effective way to account for repeated measures of a single individual. Since this is a discrete count response, our first intuition was to use a poisson distribution. Given that the data set contains many zeroes, there was a high risk of overdispertion, and upon testing we did indeed find our model was overdispersed. Therefore, we looked into other options and decided the best way forward was to use a negative binomial distribution to account for this. As expected, it had no issues with overdispersion.

Results:

Table 3: Summary of test statistics for our generalized linear mixed model analyzing offspring production. The coefficients and p-values were determined using the summary() function, while the 95% CI were calculated using the confint(,family = uniroot) function.

Source of Variance	Coefficient Estimate	Upper & Lower 95% CI	p-value
Offspring			
Intercept	5.317	5.477 5.160	<0.0001
Treatment (single)	-0.141	0.078 -0.361	0.207
Day	-0.099	-0.087 -0.111	<0.0001
Population (wild)	-0.856	-0.640 -1.073	<0.0001
Treatment x Day	0.020	0.036 0.003	0.016
Treatment x Population	-0.028	0.274 -0.330	0.856
Day x Population	0.102	0.118 0.086	<0.0001
Treatment x Day x Population	-0.008	0.013 -0.030	0.445

Table 4: The random effects from our generalized linear mixed model analyzing offspring production. Variance components were estimated using the summary() function, 95%CI using the confint(,family = uniroot) function.

Source of Variance	Variance	Upper & Lower 95% CI
Offspring		
Line	1.413×10^{-9}	inf 0

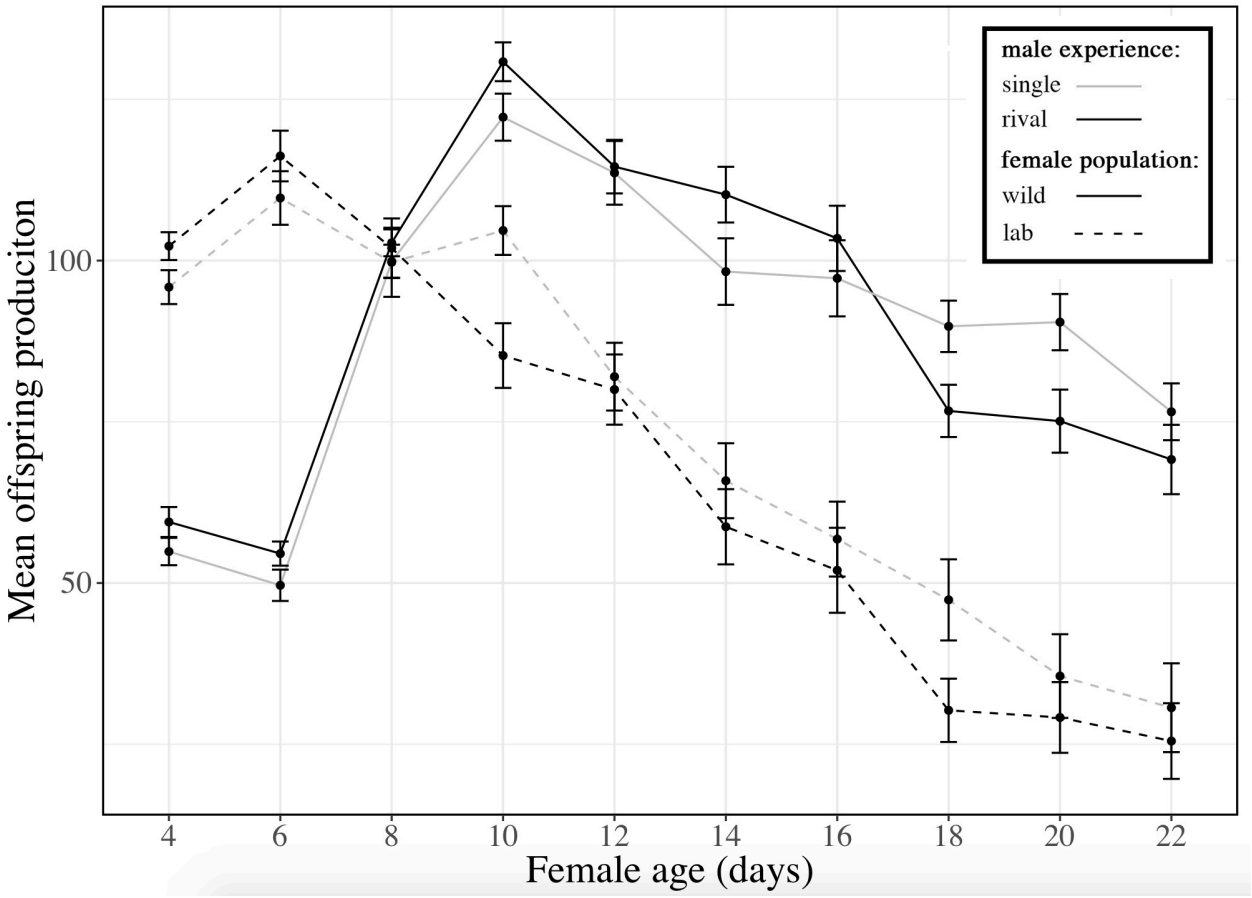


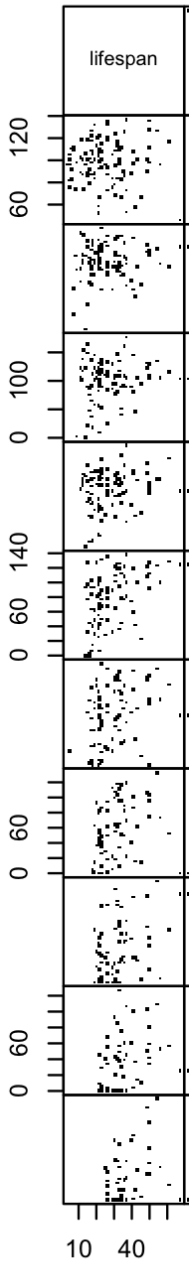
Figure 3: Offspring curves demonstrating the mean amount of offspring produced over time for each treatment combination. Error bars represent the standard error of each mean.

Conclusions: Not surprisingly, females varied significantly in their offspring production depending on the day of egg-laying. It was however interesting to see how big the difference was between females from the lab and wild populations. Overall, females from the wild population appear to produce more offspring over their lifetime. It is also interesting to see a strong interaction between population and day. This can be seen by looking at the population trends over time, where the wild population has a quadric shape, and the lab population has a negative linear curve.

Contrary to our prediction, it appears the effect of male experience does not have a strong effect on the overall lifespan production of his mates. However, there is an interesting significant interaction between male experience and day, where in both populations, it appears that females produce more offspring early on in life and less later on in life when mated to males housed with rivals. Similar to lifespan, it appears that male genotype does not have a very strong influence on the offspring production of his mates.

Further analysis:
If given more time, we would be interested in constructing some sort of multivariate mixed model in order to analyze the relationship between lifespan and offspring production over time. A quick analysis of the correlation matrix between these factors highlights an interesting trend, whereby early-life reproduction doesn't appear to be a good predictor of lifespan, but high production later in life becomes an increasingly good predictor of living longer.

```
> cor(data[, 6:16], use = "pairwise.complete.obs")
```




```
lifespan      1.000000000
offspring.1  -0.009139966
offspring.2   0.119138298
offspring.3   0.178361836
offspring.4   0.199899710
offspring.5   0.432094472
offspring.6   0.353024543
offspring.7   0.378799181
offspring.8   0.424503357
offspring.9   0.459553445
offspring.10  0.372593196
```