Stats question responses:

**Y/N flight stats:**

Coefficients between the interactions host \* average\_mass\_trans and sym\_dist \* average\_mass\_trans for the Males top model are conflicting. Problematic?

Coefficients between the interactions host \* wing2body and sym\_dist \* wing2body for the Males top model are conflicting. Problematic?

*The pieces here: 1) increasing wing2body ratio increases flight probability.*

*2) The positive effect of wing2body ratio on flight probability is more pronounced on GRT*

*3) The positive effect of wing2body ratio on flight probability decreases moving away from the sympatric zone*

*I do think all of these things can be simultaneously true, if difficult to visualize.*

They definitely make sense conceptually, but was I curious about the coefficient values themselves (e.g. host\*wing2body = -9.46 vs. sym\_dist\*wing2body = 7.45). I’m curious to know if this is something to be concerned about. If the coefficients maybe need to be scanned for a masking effect or other some other statistical weakness because of the relationship between host plant and distance from the sympatric zone? Maybe sym\_dist is not too effective either in itself? When I dropped host from the model, the sym\*wing2body effect dropped an order of magnitude and both the sym\_dist effect and its interaction term was not significant. If you drop sym\_dist, host is still significant but its interaction term isn’t.

noE vs. E source script? Why use the + E source script?

*Because we know that days from start is a meaningful covariate that can’t be explained by the other factors we’re including in the model, and know that it is partially conflating host plant due to non-random mortality, in my opinion it really has to be left in the final model set, even if it doesn’t come out as significant in the final model.*

**Flight speed:**

Difference in pool sizes between sexes causes this hard-to-parse-out comparison?  
*We certainly have more power to detect what’s happening with males. The unequal sample sizes is something that could be more easily dealt with in non-parametric tests.*

What to do about potential “outliers”? I don’t see them as outliers, just some powerful male flyers.

*I believe the outlier removal was originally to avoid including bursts of speed that were the result of us blowing on the bugs (which I think is now being handled more rigorously in the python part of the data processing), so we could drop these lines. However, in the speed script, all of the bugs we removed as outliers were bursters, which weren’t included in the speed analyses anyway, so I don’t think we need to worry about it.*

Use max speed instead of average speed?

*In principal, I think it would be cool to look at max speed in addition to average speed. I believe we originally didn’t look at max speed because of uncertainty about whether they were the result of us blowing on them. The highly bimodal distribution of max speed, within both males and females, does look suspicious. I am also cautious because, if I recall correctly, there is a cut-off for these in the python scripts that excludes some very high speeds? I wouldn’t want that choice to influence our interpretation. I would be interested in talking about your observations from working with the actual data about when these kinds of high speeds occur.*

*Something worth noting is the relationship between max speed and average speed (see figure). In the lower part of the max speed distribution, there is a very nice positive relationship between max speed and average speed (roughly, the black points here). In the upper range of max speeds, however, this relationship gets a lot muddier (red points). This suggests to me that these max speeds may not be indicators of actual strong flight. But, this could also be that different fliers have different patterns of acceleration; ie, some individuals might start very fast but then slow way down, while others may be more consistent.*

**

**Distance**

MLC pointed out that distance is not normal even after transforming it, but I’m really not sure if it’s ok to let it pass. Seems like it’s better not. We might have to use nonparametric tests like the Kruskal-Wallis.

*I’m fine with changing the approach to this one. I haven’t really used Kruskal-Wallis tests before – if I understand these, you essentially rank all the responses and look at whether different groups are coming out earlier in the order? I see how this could work for basically any type of distribution, but that you lose information (ie, the difference between 0.1 m and 0.2 m is treated the same as the difference between 1000m and 1500m, if there is nobody else in between those values). We could also consider going Bayesian for this one.*