Plasticity in Littorina offspring, the size issue

Samuel

4/2/2020

For now, the data and results belong to the island CZA.

## Part one: Correlation and multicollinearity

A plastic response to a change in the environment can be observed in one or more phenotypes. Ideally, we would like to understand to what extent each phenotype is plastic but in some (perhaps many) cases the plastic responses are not independent because the phenotypic values are correlated.  
As expected, size, thickness and weight result to be highly correlated within generations (Fig. 1) and between generations (Fig. 2). One way to detect whether they are also collinear is to check the Choleski decomposition of the correlation matrix - if there is (multi)collinearity there will be some diagonal elements that are close to zero (Table 2). There is also another way to assess multicollinearity and that is by computing a score called the variance inflation factor (VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model. I have not exactly figured out how to implement this second test for multicollinearity to our case but, for example, if we consider the outcome variable to be the thickness of the offspring, then we can build a regression model with four predictors, thickness, weight, shape (PC2) and size of the parents, and calculate the VIF score after model fitting. Does it make sense?

## [[1]]

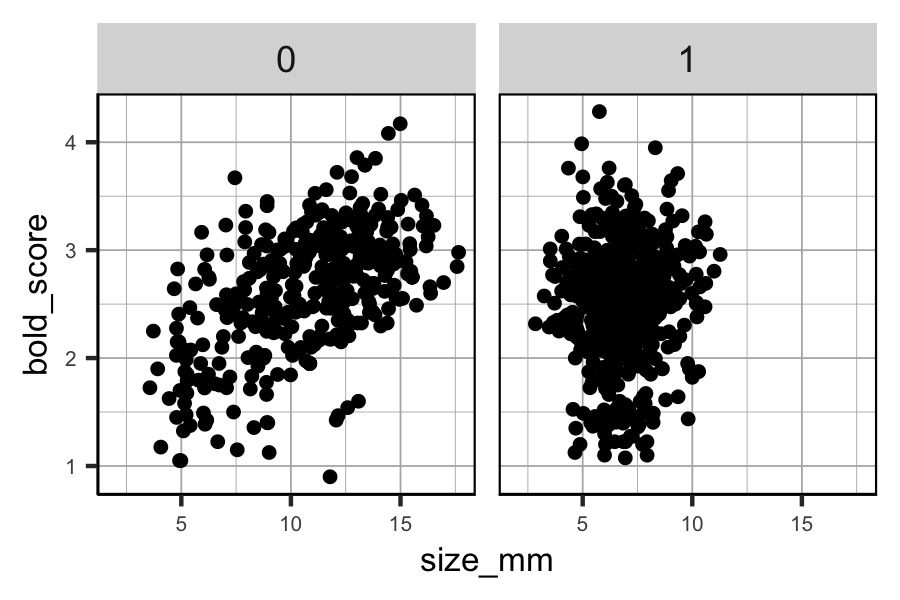


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[2]]

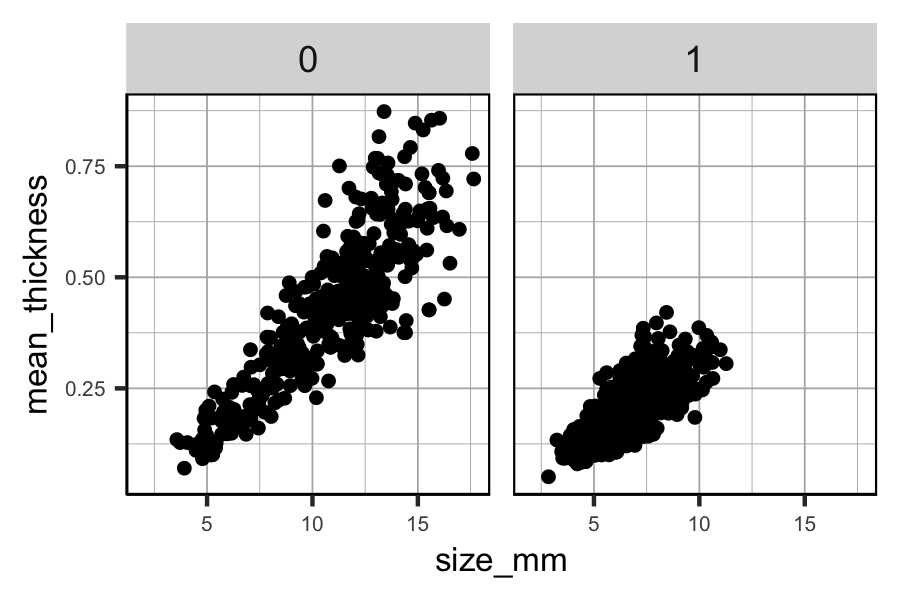


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[3]]

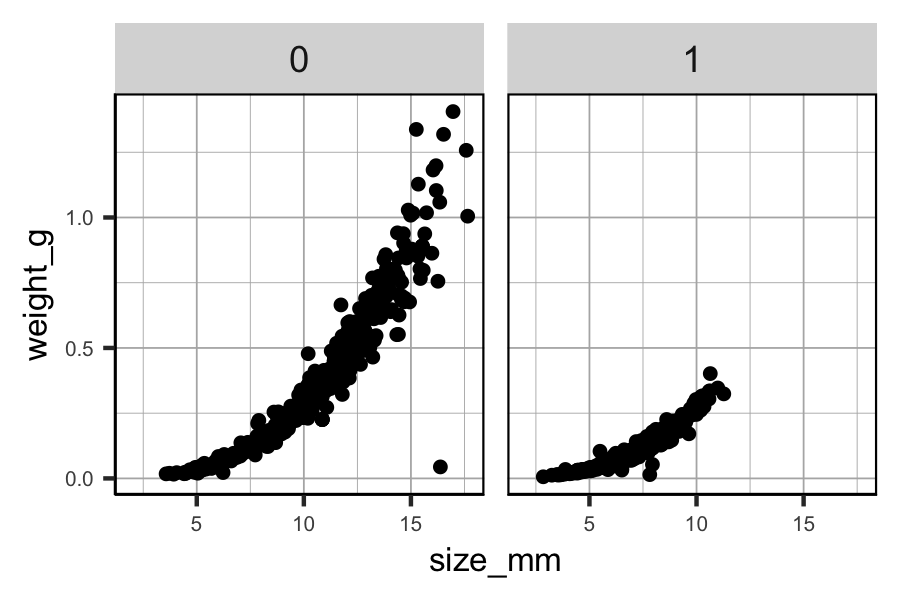


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[4]]

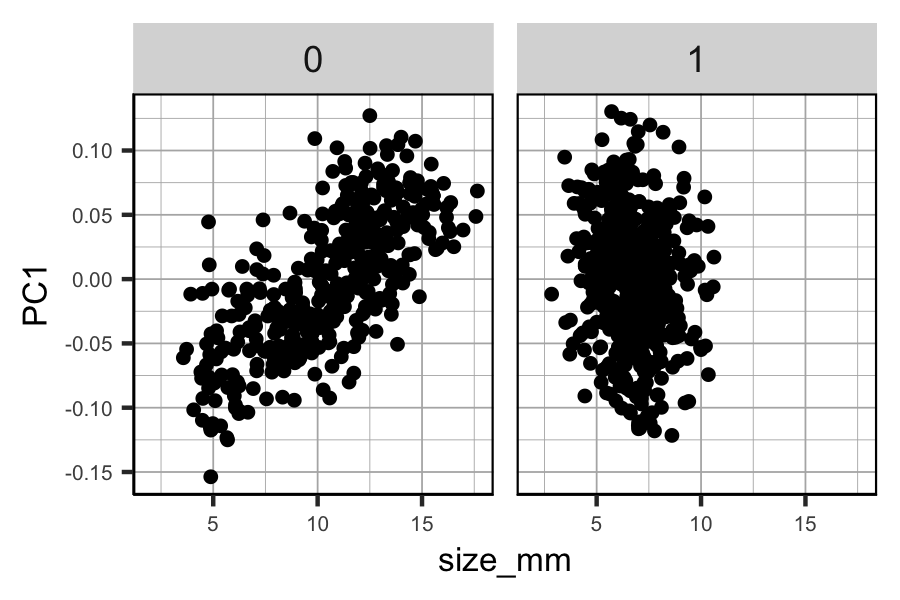


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[5]]

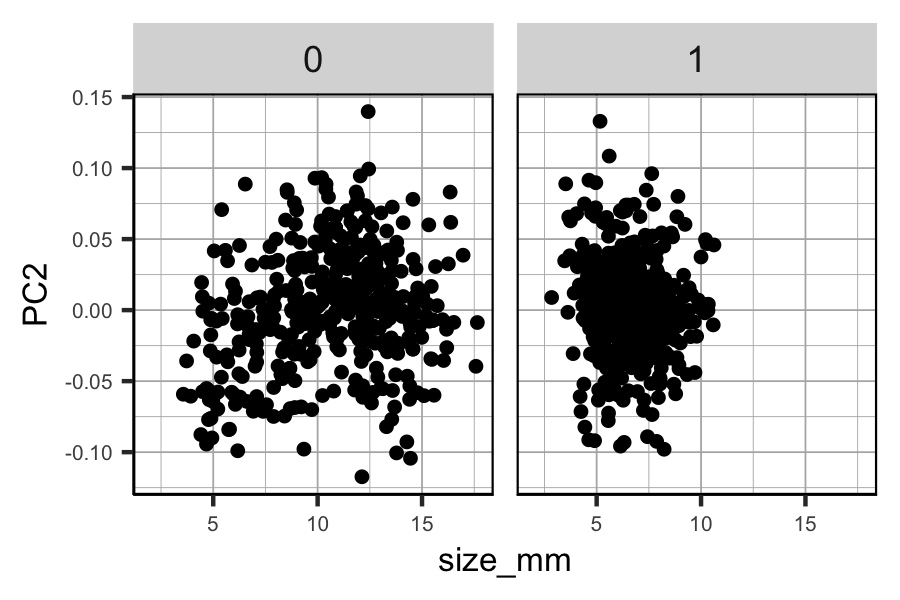


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

None of the diagonal elements are close to zero but the closest values are found in size, thickness and weight. Given these results, we can exclude mutlicollinearity and focus on the consequences of using highly correlated variables in the model. However, I am quite skeptical about the abscence of multicollinearity because in general, it is expected that an absolute correlation coefficient of among two or more predictors indicates the presence of multicollinearity (Dohoo et al. 1997). Perhaps I am performing the Choleski decomposition of the correlation matrix on the wrong set of variables. What would you suggest?

Dohoo, I., Ducrot, C., Fourichon, C., Donald, A. and Hurnik, D. (1997), “An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies”, Preventive Veterinary Medicine, Vol. 29 No. 3, pp. 221-239.

## Part two: Size at maturation

Just briefly because I think we should clarify part one first but for this section, the plan is to perform an ordinal logistic regression between the multinomial outcome variable (mature, immature and juvenile) and the predictor variable size. The result should give us an estimate of the probability of maturation depending on size. Not sure about how to split the data though. By generation and by tanks or just by tanks? I guess that in the first case we might be able to see whether the probablity of maturation differ between generations and not only between tanks and thus ecotypes but there are compromises to make.

1. In the generation 0 parents there is not enough information about maturity (Table 2) but we could use CZA data from the contact zone experiment.
2. CZA data were collected continuously along the shore and not discretely by patches but we could imitate the sampling that was done for the offspring experiment and select CZA individuals within a circle with center equal to the mean path of each population and radius of 2m (or maybe a different radius?).

Does it make sense to try by “generation” and by tanks?

Table 2. Frequency of maturity classes in generation 0 parents and generation 1 offspring.

| Maturity | Count (0) | Count (1) |
| --- | --- | --- |
| female | 261 | 204 |
| immature | 0 | 62 |
| juvenile | 2 | 719 |
| male | 177 | 234 |
| missing | 4 | 20 |