

Results section of IA1_Bi610

Dane Dewees

Fri Oct 27 21:36:04 2017

Background

This study was used to determine whether or not feeding zebrafish with enriched brine shrimp nauplii (treated with a particular supplement) affects mean zebrafish size. One of the main objectives is to determine whether size might be affected differentially in terms of overall length of the fish, as determined by its standard length in millimeters (mm), as compared to body mass, as measured by wet weight in milligrams (mg).

We sampled 200 newly hatched zebrafish from the same clutch and randomly assign each to either an enriched shrimp diet or an unenriched shrimp ("control") diet. Made sure that the experiment was carried out carefully and not confounded diet with any other variables that may affect size. After two months we measured the standard length and weight of all 200 fish.

Analysis

We use R for all of the statistical analysis and computations in this paper. This includes various functions and commands in order to convey the desired trends found in this dataset. I used a distribution model, paired with a non-parametric approach (bootstrap) as well as a linear regression model. The corresponding R code can be found in the RMD file paired with this paper.

In order to statistically address our question of interest (i.e., How does diet affect weight and physiological changes - SL?) a 1-sample t-test is performed with subsequent null and alternative hypotheses:

Weight

The Null hypothesis assumes that there is no mean difference in weight between control and enriched. The alternative hypothesis assumes that there is a non-zero mean difference in zebrafish weights between the two groups (categorical).

1. Null = $\text{weight}(\text{control}) - \text{weight}(\text{enriched}) = 0$
2. Alternative = $\text{weight}(\text{control}) - \text{weight}(\text{enriched}) \neq 0$

Standard Length

The Null hypothesis assumes that there is no mean difference in SL between control and enriched. The alternative hypothesis assumes that there is a non-zero mean difference in zebrafish standard length between the two groups (categorical).

1. Null = $\text{SL}(\text{control}) - \text{SL}(\text{enriched}) = 0$
2. Alternative = $\text{SL}(\text{control}) - \text{SL}(\text{enriched}) \neq 0$

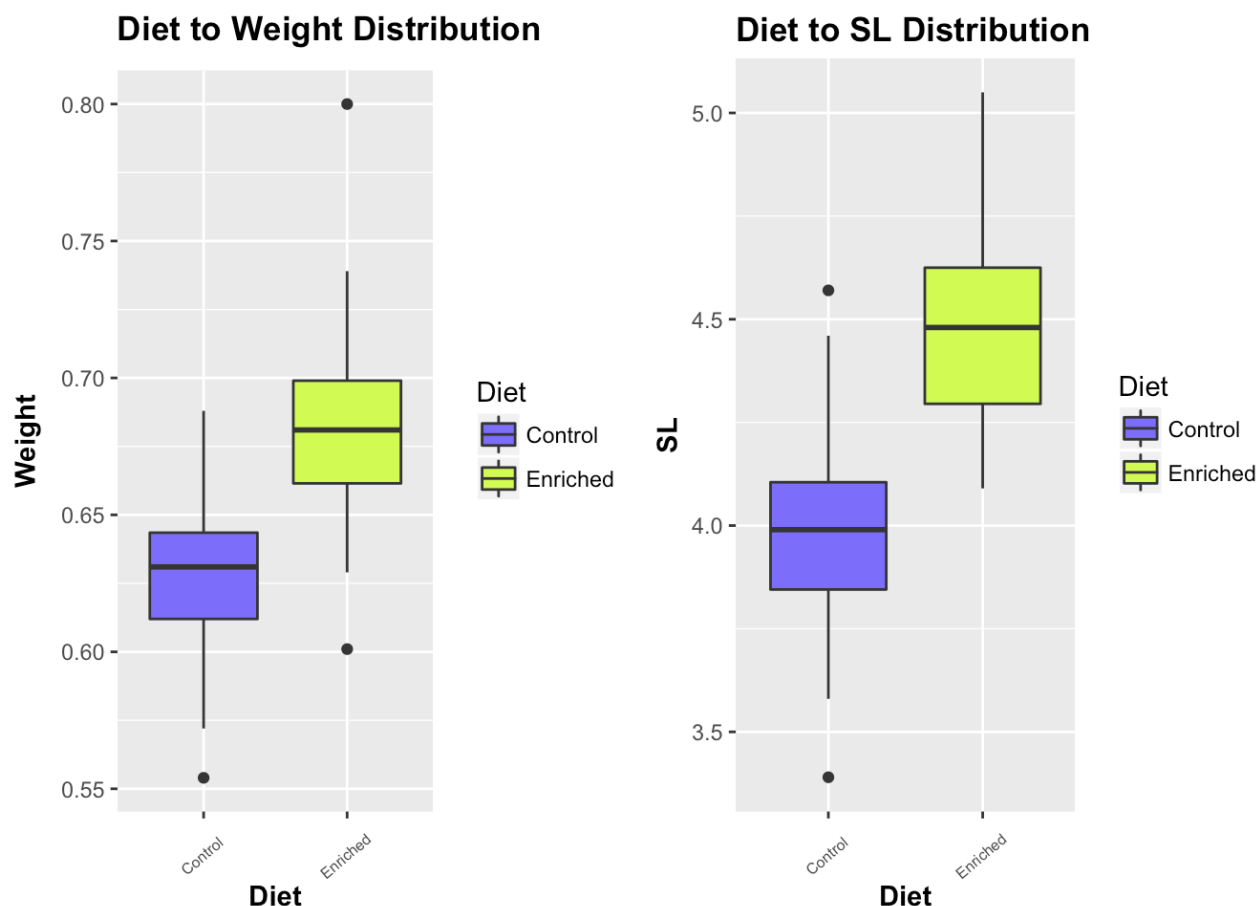
The following dot_plot was constructed using the original dataset (while removing 2 outliers), for graphical interpretation of any interesting or unusual points. Here we see the distribution of weight versus standard length given the categorical variable (diet) and it's corresponding subgroups (control & enriched) and can tell that the affects of diet before running any parameters is indicative of the diets impact on weight and physiological changes in Zebrafish (Figure 1):

Figure 1



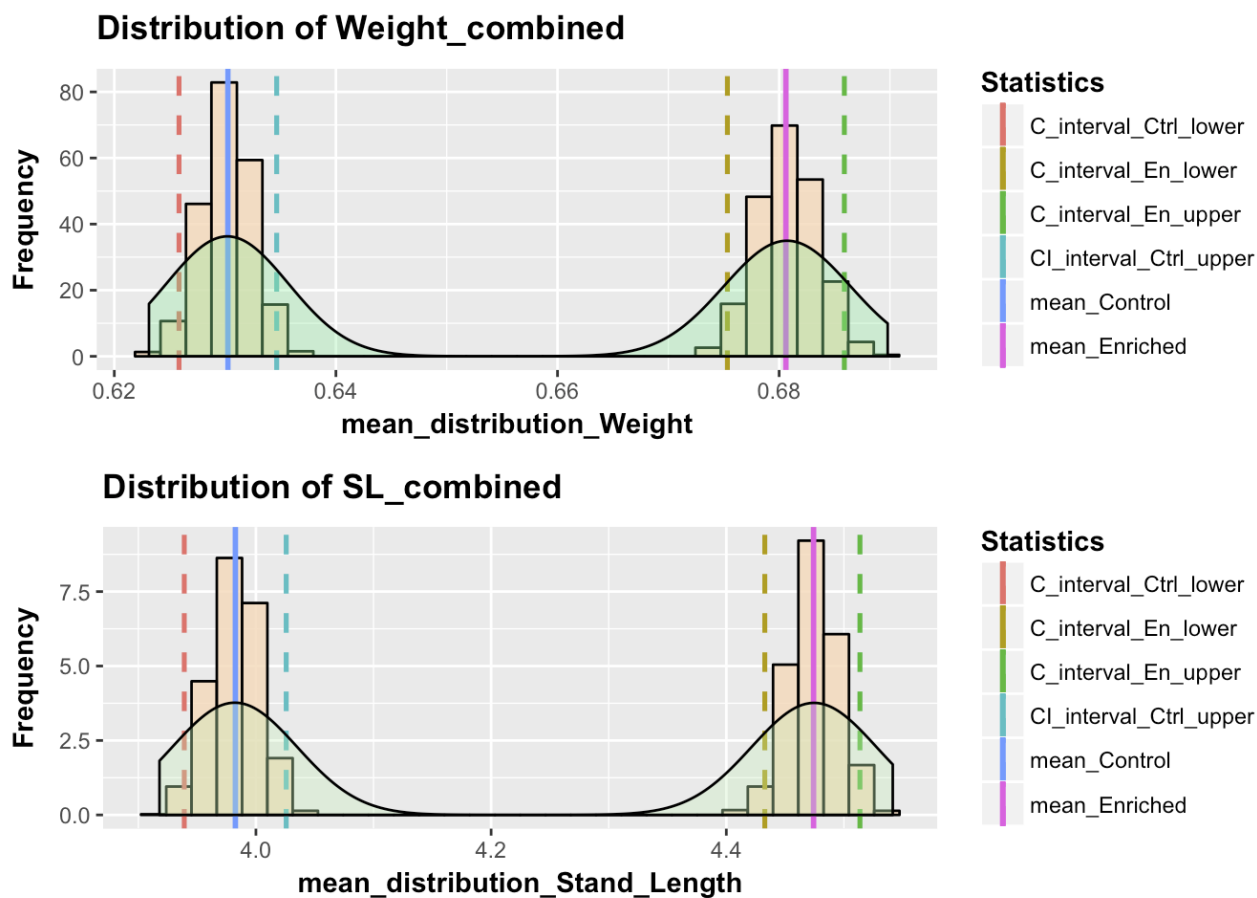
Isolating the categorical variable and it's paired distributions among the continuous variables, we found that there was a correlation with regards to diet influencing both weight and SL (Figure 2), both the SL and weight continuous variables showed changes when diet changed from control to enriched. The enriched group for both SL and Weight had a higher overall mean difference than with the control subgroup (Fig. S1). Following this subset plot, we ran a 2-sample ttest for both subset datasets for SL & Weight and found that there was a statistically significant difference in mean SL of Zebrafish between the control & enriched groups ($t_{195.98} = -16.27$, $p < 2.2e-16$) and same for Weight ($t_{190.9} = -14.08$, $p < 2.2e-16$).

Figure 2



Factoring in a non-parametric (like bootstrapping) in which it relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy, and in doing so we were able to see a distribution across both SL & Weight with regards to Diet. Resampling the data showed a positive trend of an even distribution when combining the datasets for all SL and Weight values for both control and enriched diets (Figure S2). Instead of viewing each variable independently, we appended the corresponding data in order to create a proper bootstrap function. Using statistical functions, we performed the same hypothesis test, and resampled the dataset 1000 times. Plotting the distribution of the resampled parameter with its mean and standard deviation along with the confidence interval (95%) gave us a visual representation of the overarching spread among the continuous variables (Figure 3). The corresponding statistical functions (i.e., mean, variance, CI) are shown as overlapping dashed and solid lines on the histogram, respectively.

Figure 3



In order to evaluate the strength of the relationship between the two continuous variables, we used a simple linear regression assuming normality. I decided to run two packages in R (lmodel1 & 2) in order to see if there was any difference. In reference of the model 2 linear regression, I overlayed the SMA line which shows the reduced model axis and was applied via this equation: (weight) = a+b(SL) or more commonly known as $y=mx +b$ (Figure 4). From this, we were able to conclude that the model 1 fit was more sufficient in regards to illustrating the strength of the relationship between the two continuous variables (SL & Weight). Referencing model1 figures can be found in supplemental material (Figure S3).

Supplemental Material

Figure S1

##	Variable	Diet	Variance	Mean	SD	median
## 1	SL	Control	0.0005324090	0.6301717	0.02307399	0.631
## 2	SL	Enriched	0.0007404115	0.6806566	0.02721050	0.681
## 3	Weight	Control	0.0446139353	3.9823232	0.21122011	3.990
## 4	Weight	Enriched	0.0455431870	4.4734343	0.21340850	4.480

Figure S2

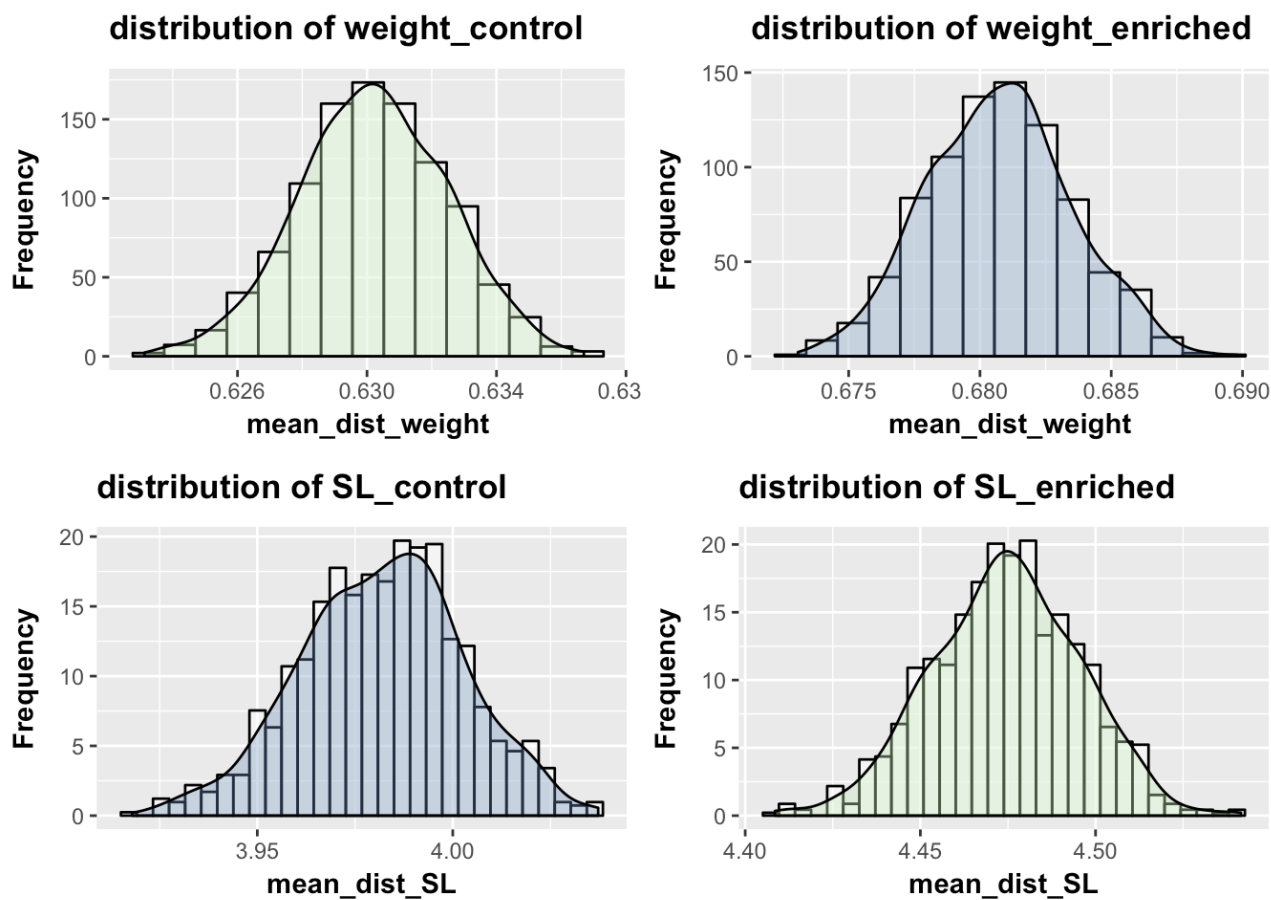


Figure S3

