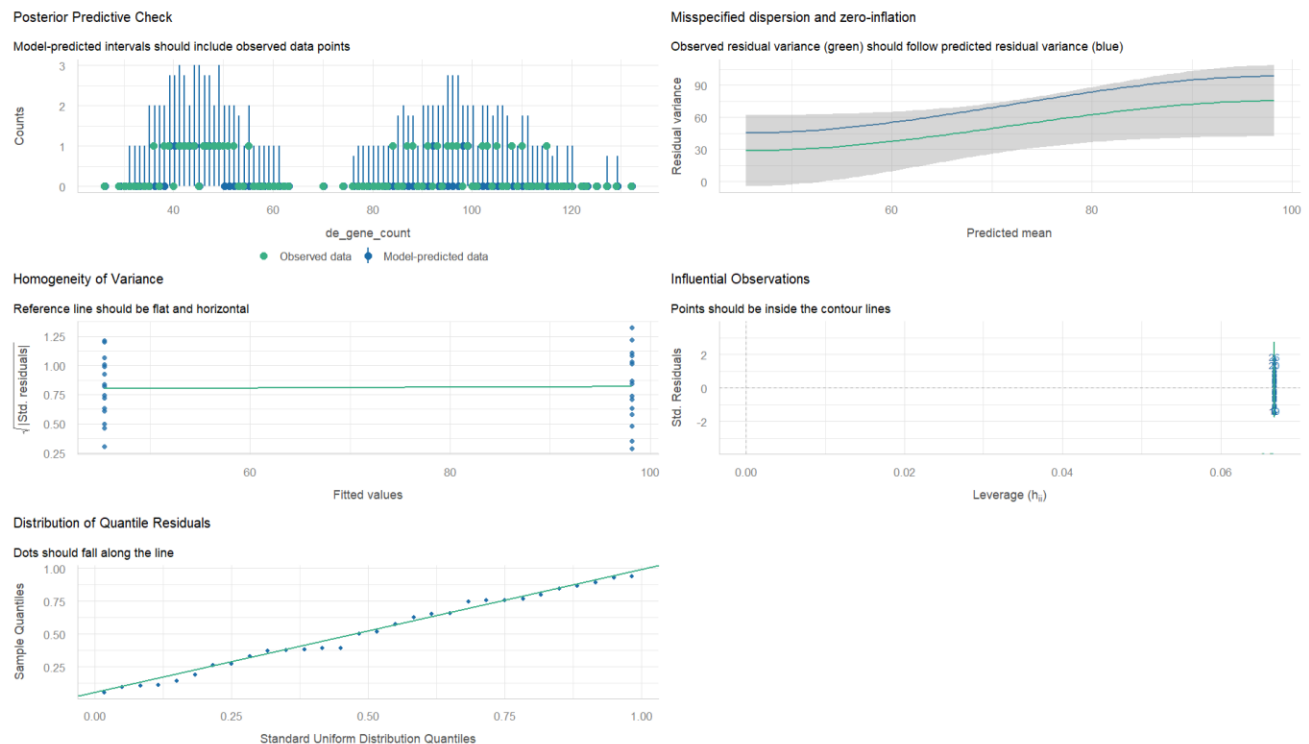
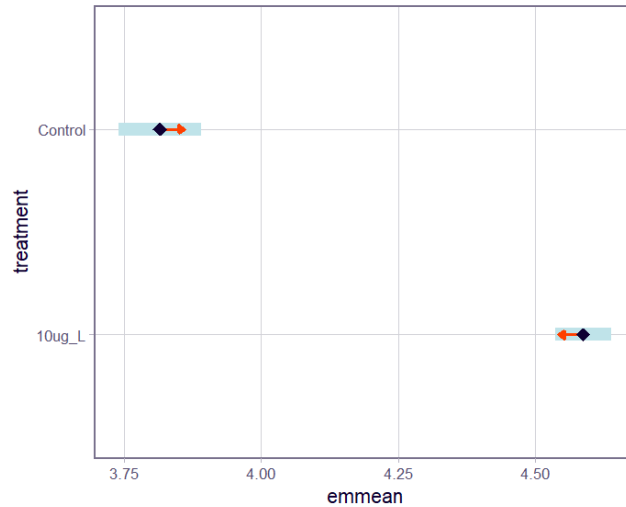


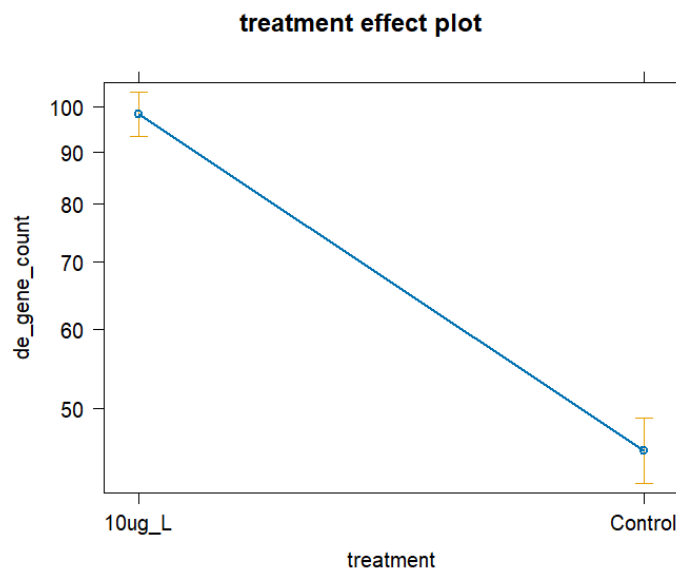
For this assignment I used different data as the data I've been using previously doesn't include discrete counts or binomials. To generate the dataset, I gave the following prompt to ChatGPT. Generate a dataset for me to use for practicing generalized linear models in csv format. Have there be two categorical variables (control and 10 micrograms/L) and a response variable (number of differentially expressed genes between male and female fish) measured in the form of counts. The following charts are models ran from this generated dataset. All data related comments are obviously not true.



The zero-inflation plot shows that the data has a slightly lower residual variance than the model predicts, which might indicate some under dispersion but it still is within the confidence intervals so it's not something to worry about at this point. Had it been outside the intervals it would have been better to run a quasipoisson distribution or a negative binomial. The homogeneity of variance chart shows that both treatment groups are homoscedastic. The influential observations plot shows that there are no overly influential residuals in the data as all points fall within the Cook's distance lines. The distribution of quantile residuals shows that the residuals are distributed equally throughout the data set.



This estimated marginal means comparison chart shows the statistical significance between the control and 10 microgram/L treatment group. From this it would be interpreted that control and treatment groups are significantly different with the treatment group causing an increase in the differentially expressed genes.



This effects chart shows the biological effects of the treatment group on the number of differentially expressed genes. From this we can tell that the treatment group results in almost a doubling of the number of differentially expressed genes from 50 to 100.