**BIOS6643 Homework 5, Due Nov. 2 at 5pm Fall 2022**

Trent Hawkins

***In this question you will analyze only the week 1 (baseline) data, C1. There was some question as to whether the experimental and control groups were comparable at this time because some aspects of the intervention may have been done before week 1. Carry out the following analyses to compare the groups at baseline, using data only from Kid1 (FamMem=3) – families were sampled based on this child. For each model, write the model equation, and write a sentence describing the results, understandable by dietitians.***

***Use a Poisson GzLM (i.e. Poisson regression) to estimate the association between condition and number of breakfast servings, adjusting for sex and weight.***

This model will follow the form of:

Where Y is the bowls of cereal consumed by Kid 1 at week 1, and X are our covariate values.

| **Term** | **Estimate** | **Std. Error** | **Statistic** | **p.value** | **Conf.low** | **Conf.high** |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 2.224 | 0.213 | 3.760 | <0.001 | 1.458 | 3.356 |
| Experimental | 2.374 | 0.135 | 6.405 | <0.001 | 1.838 | 3.123 |
| Weight | 1.218 | 0.132 | 1.491 | 0.136 | 0.937 | 1.575 |
| Sex (Male) | 1.220 | 0.082 | 2.418 | 0.016 | 1.039 | 1.434 |

*Interpretation:*

The results in the table above have been exponentiated to obtain a more desirable interpretation. The rate of cereal bowls eaten in the treatment group is 2.374 times (95% CI: 1.838, 3.123; p < 0.001) the rate in the control group after adjusting for sex and weight.

***Is the Poisson distribution sufficient in modeling the data? How do you know? If it is insufficient, what impact do you expect this to have on the standard errors? (Subsequent parts may help you to answer this.)***

I expect that there is overdispersion in this model. This model assumes that E[Bowls Eaten] = Var (Bowls eaten). The Mean bowls of cereal eaten by the first child at baseline is 6.55 bowls, but the variance is 23.54 bowls. Since the variance is *much* higher than the mean, we have evidence of overdispersion. In the case of this model, I expect that the standard error is small and is producing overly narrow confidence intervals.

***Repeat (a) allowing for overdispersion by using quasilikelihood with the Poisson GzLM. Use the Deviance method for estimating the scale parameter. Show algebraically the relation between the QL SEs for the betas and those from the Poisson model in (a).***

This model takes the same form as the model above:

but it also estimates a scale parameter (. The new Standard Errors for this model can be represented as:

The results below have been exponentiated to obtain a more desirable interpretation. After exponentiation the null value becomes 1.

| **Term** | **Estimate** | **Std. Error** | **Statistic** | **p.value** | **Conf.low** | **Conf.high** |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 2.224 | 0.365 | 2.191 | 0.031 | 1.070 | 4.481 |
| Experimental | 2.374 | 0.232 | 3.732 | < 0.001 | 1.546 | 3.852 |
| Weight | 1.218 | 0.227 | 0.869 | 0.387 | 0.774 | 1.886 |
| Sex (Male) | 1.220 | 0.141 | 1.409 | 0.162 | 0.927 | 1.613 |

*Interpretation:*

The rate of cereal bowls eaten in the experimental group is 2.374 times (95% CI: 1.546, 3.852; p < 0.001) the rate in the control group after adjusting for sex and weight.

***Repeat (a) allowing for overdispersion by using a Negative Binomial GzLM. This is a likelihood-based method.***

The results below have been exponentiated to obtain a more desirable interpretation. After exponentiation the null value becomes 1 to reflect the multiplicative nature of exponentiation.

| **Term** | **Estimate** | **Std. Error** | **Statistic** | **p.value** | **Conf.low** | **Conf.high** |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | 2.546 | 0.343 | 2.722 | 0.006 | 1.312 | 4.917 |
| Experimental | 2.357 | 0.195 | 4.386 | 0.000 | 1.603 | 3.469 |
| Weight | 1.105 | 0.234 | 0.425 | 0.671 | 0.689 | 1.781 |
| Sex (Male) | 1.187 | 0.144 | 1.191 | 0.234 | 0.898 | 1.567 |

*Interpretation:*

The rate of cereal bowls eaten in the experimental group is 2.357 times (95% CI: 1.603, 3.469; p < 0.001) the rate in the control group after adjusting for sex and weight.

***Use the chart below to summarize the models you fit previously, comparing Conditions and adjusting for Sex and Weight. In each entry except ‘Intercept’ and ‘Dispersion param’ give the estimate of the rate ratio and a 95% CI (these are easy to calculate from the beta and SE). For ‘Intercept’ give the beta and its SE. For ‘Other param’, give the Scale parameter for QL, the SD for the normal error model, or the dispersion parameter for NB, but not the SE. QL=Quasilikelihood, NB=Negative Binomial.***

|  |  |  |  |
| --- | --- | --- | --- |
|  | Poisson regression | Poisson QL | NB NLMIXED |
| Intercept | 2.224 (0.213) | 2.224(0.365) | 2.224 (0.343) |
| Condition (Experiment) | 2.374 (1.838, 3.123) | 2.374 (1.546, 3.852) | 2.374 (1.603, 3.469) |
| Sex (Male) | 1.220 (1.039, 1.434) | 1.220 (0.927, 1.613) | 1.220 (0.898, 1.567) |
| Weight | 1.218 (0.937, 1.575) | 1.218 (0.774, 1.886) | 1.218 (0.689, 1.781) |
| Dispersion param | NA | 2.945 | 3.164 |

***Write a short paragraph summarizing the results of comparing conditions, e.g. which condition gives higher consumption and by how much.***

Fixed effect estimates and overall inference remain the same across all our models. The cereal bowl consumption rate in the experiment group is roughly 2.4 times that of the control group when adjusting for Sex and Weight. The conditions of the experiment are not well defined, though. The experimental group could be eating something like Frosted Flakes, while the control is eating Fiber-one. The jury is out on the cause of this doubling of cereal consumption.

***Write a short paragraph comparing the model fits, e.g. differences between models.***

While the fixed effects remain the same across the models, the confidence intervals change. The effect that the experiment group has on the rate of cereal consumption is large, so the inference is relatively stable. The term that is most effected by accounting for overdispersion is Sex. Looking only at the Poisson Regression model, the rate of male cereal consumption is 1.2 times (95% CI: 1.039, 1.434) that of females. That confidence interval does not contain 1, so it can be considered statistically significant. After accounting for overdispersion, the confidence interval widens and agrees inferentially between the Poisson QL and Negative-Binomial models. It is clear from this model comparison that Standard Errors are underestimated in the case of overdispersion and tend to increase (and therefore, widen confidence intervals) when accounted for by estimating a dispersion/scale parameter.