



---

**Re: Question about Negative Binomial Model with Interactions**

---

**From** Thompson, Paige A <thomp732@purdue.edu>

**Date** Mon 5/6/2024 8:10 PM

**To** Craig, Bruce A. <bacraig@purdue.edu>

**Cc** Guda, Sumeeth Krishna <sguda@purdue.edu>

Dr. Craig,

Thank you so much for that information! Sumeeth was a great help!

Best,  
Paige

--

Paige A. Thompson, M.S., ACSM EP-C  
Graduate Student/ Research Assistant  
Department of Health & Kinesiology  
Purdue Motor Development Lab

---

**From:** Craig, Bruce A. <bacraig@purdue.edu>

**Sent:** Friday, May 3, 2024 5:42 PM

**To:** Thompson, Paige A <thomp732@purdue.edu>

**Cc:** Guda, Sumeeth Krishna <sguda@purdue.edu>; Craig, Bruce A. <bacraig@purdue.edu>

**Subject:** RE: Question about Negative Binomial Model with Interactions

Paige:

These AIC values confirm that the reduced model adequately describes the data relative to the more complicated model. It is used solely for model comparison purposes and does not assess whether the model fits the data well. In other words, we could be finding the better of two poorly fitting models.

The scaled deviance and scaled Pearson statistics can be used for assessing goodness of fit to the binomial or Poisson distributions in most settings. It is suspect for the negative binomial situation because the distribution has a second parameter that describes the extra dispersion. Instead, a test that the negative binomial is preferred to the Poisson is often done. Perhaps you saw a test that the dispersion parameter is equal to 1 or 0 (on the log scale)? The other thing you can do is consider zero-inflated negative binomial model and see if its AIC is better than just the negative binomial model result. This is comparable to a chisq test where, for a particular combination of parameters, you compare the expected number of 0's, 1's, 2's etc with the actual counts you observed.

Hope that helps. I've included Sumeeth on this email so updates to your directory can be made

Sincerely,

Prof Craig

---

**From:** Thompson, Paige A <[thomp732@purdue.edu](mailto:thomp732@purdue.edu)>  
**Sent:** Friday, May 3, 2024 1:29 PM  
**To:** Craig, Bruce A. <[bacraig@purdue.edu](mailto:bacraig@purdue.edu)>  
**Subject:** Re: Question about Negative Binomial Model with Interactions

Hi Dr. Craig,

Thank you so much for your response!

We calculated dispersion by using the sum of the squared residuals and dividing by the degrees of freedom. The model with all interactions had a value of 0.85 and the reduced model had a value of 0.97. I also compared the AIC values of both models (410 for the full model and 405 for the reduced) and looked at the Q-Q plots.

I believe both issues you pointed out are true for my data. The count of the behavior I am looking at is very low for each subject, it only happens about 4 times a session. But I can't do anything about that since it's an observational study. Therefore, I do have many instances of combinations with Age and position etc. where there are many zeros.

Best,

Paige

--

Paige A. Thompson, M.S., ACSM EP-C  
Graduate Student/ Research Assistant  
Department of Health & Kinesiology  
Purdue Motor Development Lab

---

**From:** Craig, Bruce A. <[bacraig@purdue.edu](mailto:bacraig@purdue.edu)>  
**Sent:** Thursday, May 2, 2024 4:23 PM  
**To:** Thompson, Paige A <[thomp732@purdue.edu](mailto:thomp732@purdue.edu)>  
**Cc:** Craig, Bruce A. <[bacraig@purdue.edu](mailto:bacraig@purdue.edu)>  
**Subject:** RE: Question about Negative Binomial Model with Interactions

Paige:

The approach of a full model with interactions and then peeling away the insignificant ones sounds reasonable.

How are you assessing fit? Are you using standard approaches that rely on large-sample asymptotics (like the chi square distribution)? It turns out that when the mean of the negative binomial/Poisson distribution is small, sample sizes need to be pretty large and as a result standard test results can be misleading. That may be part of the problem. The other issue is that when fitting the full model, you likely have many cells (combination of Age, position, size, etc) where all you have are 0's (no variation) but as you reduce things down you start to see variation in all the remaining cells.

That's my quick answer. I'm happy to consider this further but need more information about fits etc.

Sincerely,

Prof Craig

---

**From:** Thompson, Paige A <[thomp732@purdue.edu](mailto:thomp732@purdue.edu)>

**Sent:** Thursday, May 2, 2024 1:00 PM

**To:** Craig, Bruce A. <[bacraig@purdue.edu](mailto:bacraig@purdue.edu)>

**Subject:** Question about Negative Binomial Model with Interactions

Hello Dr. Craig

I was working with a student consultant the past few weeks and I just wanted to have someone double-check the work that we did and make sure that what I am doing is appropriate. If you would be able to help, I would greatly appreciate it.

I have used a negative binomial model for count data that is over dispersed. I have one outcome variable (Value) and multiple predictor variables that each have levels (Age(12-24); Position (sit, stand, squat, walk); Toy1Size (small, medium, large); Toy1Weight (light, moderate, heavy); Toy2Size (small, medium, large); Toy2Weight (light, moderate, heavy)).

I am interested in the interaction effects of Age on all other predictors for theoretical reasons. To have a model with interactions I first included all age interactions and then did backward elimination on interactions with high p-values. My main question is if this process is appropriate? Including all interactions in my model seems to give me a model with not good fit if I look at the dispersion. The model from backward elimination seems to have good dispersion. I have never used this model before so any insight would be helpful. I can provide the code if needed!

Best,  
Paige

--

Paige A. Thompson, M.S., ACSM EP-C  
Graduate Student/ Research Assistant  
Department of Health & Kinesiology  
Purdue Motor Development Lab