BIOSTAT 201B Homework 4

Lillian Chen (c.lillian@ucla.edu)

February 16, 2022

Warm-up Problems

1. Poisson and Negative Binomial Regression Basics:

(a) Explain what the distributions, link functions and systematic components are for Poisson and Negative Binomial GLMs.

(b) Give examples of circumstances where you might want to use each of these models.

2. Munching (Computer) Chips (AA 4.6 and 4.8):

An experimenter analyzes imperfection rates for two processes used to fabricate silicon waters for computer chips. Each treatment is applied to 10 chips and the number of imperfections are recorded. The thickness of the silicon coating is also reported (0 = thin and 1 = thick). Use the data to answer the following questions.

(a) Fit a simple Poisson regression using treatment as the predictor. Does there appear to be a significant difference between the two processes? Show two ways you could perform this test.

(b) Give a brief interpretation of the regression coefficients and their confidence intervals, both on the log scale and on the mean scale.

(c) Now add the thickness of the silicon coating as a predictor and rerun the model. Is the new model a significant improvement over the previous model? Describe as carefully as you can the effect of the coating thickness on the number of imperfections. What are you assuming about the joint effect of process and coating thickness by fitting the model this way?

3. More Sports Fanatics (AA 4.14):

British fans are mad (literally!) about their soccer, even more than New Zealanders are about rugby. This data set gives the number of fans in attendance (in thousands) and the total number of arrests in the 1987-1988 season for soccer teams in the second division of the British football league.

(a) Fit a Poisson regression model for number of arrests using attendance as an oset variable. Explain (i) why it is important to use an offset variable here and (ii) what the interpretation of the intercept from the model is. (Note that there are no predictors here other than the offset!)

(b) Plot arrests vs attendance and overlay the prediction equation. Obtain residuals from the model and use them to identify teams that had much larger or much smaller than expected number of arrests.

4. Camping Data (ATS):

This problem goes in depth into the fishing example I used in class. The outcome variable is the number of fish caught (numfish) by 250 groups of people who went to a wildlife park. The predictors are the number of people in the group (persons), the number of children in the group (children) and whether or not they were camping (camper = 1 for yes and 0 for no.) Use the data to answer the following questions.

(a) Fit a Poisson regression for the number of fish caught with number of people, number of children and whether or not the party was camping as predictors.

(b) Give careful interpretations of the coefficients of persons, children and camper status and their confidence intervals (i) on the log scale and (ii) on the mean scale. Which of these variables appear to be significant predictors?

(c) Plot the Pearson residuals versus the fitted values for this model and also obtain the deviance and Pearson goodness of fit tests. Do you think it is reasonable to use these measures here? Are there any obvious outliers? Does the model seem to be well calibrated? If not can you identify where the misfit is occurring?

(d) Now we evaluate the issue of over-dispersion. Suggest some plausible reasons why there might be over-dispersion in this data set and then check for it in the following ways:

(1) Using basic descriptive statistics obtain the mean and variance of the numfish variable. Do this for various subgroups of the data set (e.g. camping or not, with children or not, etc.) What do these statistics suggest about over-dispersion. (2) Calculate the approximate dispersion factor based on the Pearson chi-squared goodness of fit statistic. (3) Explain what your plot of residuals from part (c) suggests about over-dispersion.

(e Rerun the model using a negative binomial regression. Does this model confirm your conclusions about over-dispersion from part (d)? Explain briefly.

(f) Give brief interpretations of the coefficients from the negative binomial regression. Are your conclusions (as to magnitude, direction and significance) of the effects of the predictors any different from those in part (b)?

In the next part of the problem we consider the issue of zero inflation.

(g) Explain intuitively why we might expect zero-inflation in the context of this problem. Do you expect any of these predictors to be especially predictive of who would be a "certain 0?".

(h) Continuing with your idea of descriptive statistics from part (d) calculate the expected number of 0's for various subgroups of the data set assuming a Poisson distribution and then examine the actual number of 0's. Do we seem to have a zero-inflation problem? (Note: You may find it helpful to remember that for a Poisson random variable the probability of k events is $P(Y = k) = \mu_k e^{-\mu}/k!$).

(i) Fit a zero-inflated Poisson model to the data using persons, children and camper as the predictors for the Poisson component and experimenting with various variables the zero-inflation component of the model. Do you think this model is an improvement over the standard Poisson regression? Explain briefly.

(j) Give brief interpretations of the coefficients for each component of the zero-inflated Poisson model. Which variables seem to be significant in each part of the model? Explain as carefully as you can what this model suggests about how these variables affect numbers of fish caught.

(k) Now fit a zero-inflated negative binomial model using the same model set-up as in part (j). Is this model superior to the zero-inflated Poisson model of part (j)? Does it look superior to the plain negative binomial model of part (e)? Explain briefly in each instance. What does that tell you about the issues of overdispersion and zero-inflation? Do any of your conceptual conclusions from part (j) change with this model?

Problems to Turn In

5. I Wish I Could Play Hookey From 201B (ACM Problems 12.25 and 12.26):

This problem examines factors associated with adolescent students being absent from school without a valid excuse. For each of n = 252 students, the number of times absent in the last month was recorded (abbreviated nhookey for number of times the student "played hookey") along with possible predictors including age, sex (1 = male and 0 = female) and the degree to which the student liked school (rated from 1 = liked it very much to 5 = disliked it very much). Our goal in this problem is to build a model for school absences.

(a) Fit the Poisson model with no predictors and explain as carefully as you can the meaning of the estimated intercept. (Note–you may find it helpful to exponentiate it!)

e^{β_0} represents the mean number of times a student was absent in the last month. The Poisson model with no predictors tells us that students were absent on average 2.27 times in the last month.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8197	0.0418	0.7378	0.9017	384.34	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Figure 1: Poisson model with no predictors

(b) Plot mean number of days absent as a function of (i) likeschool and (ii) age. Do there appear to be significant relationships between these variables and numbers of days absent? Do the relationships appear linear?

There does appear to be a positive monotonic relationship between mean number of days absent and each of likeschool and age. The relationship between mean number of days absent and age looks like it could potentially be curvilinear. The relationship between mean number of days absent and likeschool looks fairly linear.

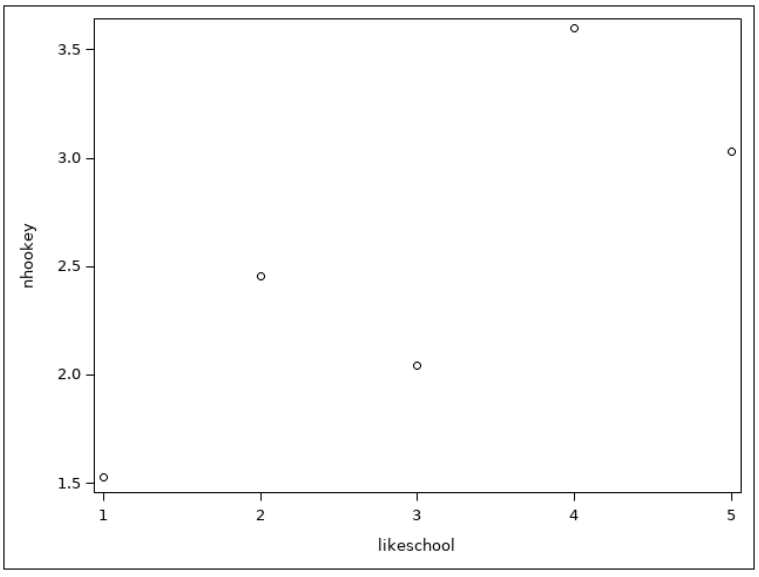

Figure 2: Mean number of days absent as a function of likeschool

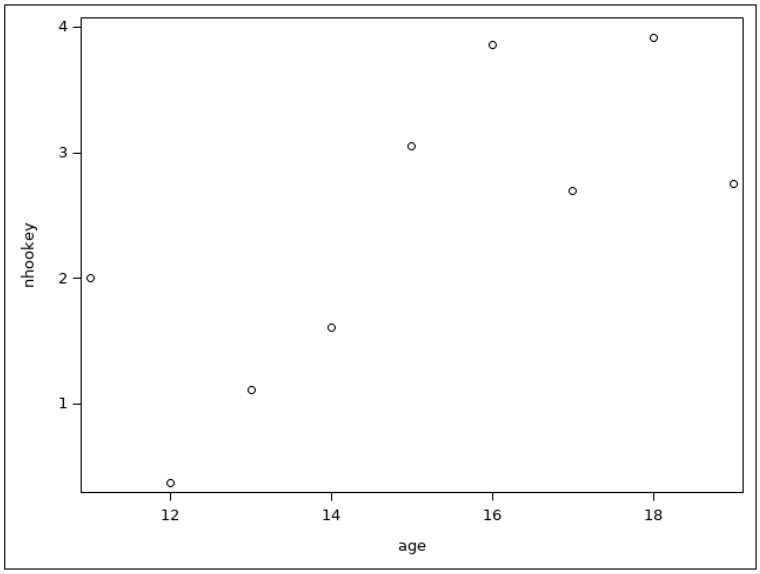

Figure 3: Mean number of days absent as a function of age

(c) Fit a Poisson regression for the number of days the adolescents in the sample were absent from school with sex, age and how much the student liked school as predictors. Obtain two versions–one in which likeschool is treated as continuous and one in which it is treated as categorical. Which version of the model do you think fits better and why?

I think the first model treating likeschool as a categorical variable fits better, but it is not so different from the second model treating likeschool as a continuous variable. While the model treating likeschool as categorical has a higher log likelihood, lower deviance, and a larger Pearson chi-square value, the differences between those goodness of fit criteria in the categorical vs continuous model are minimal. Additionally, we find all variables significant in each of the two models. We see in the categorical model that likeschool = 3 is not statistically significant, which is additional information that we cannot glean from the continuous model, but we sacrifice 3 degrees of freedom to come to a similar conclusion as the continuous model.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2817	0.3581	-3.9836	-2.5798	83.97	<.0001
likeschool	1	0.1095	0.0317	0.0474	0.1717	11.93	0.0006
age	1	0.2229	0.0221	0.1795	0.2663	101.30	<.0001
sex	1	0.6814	0.0885	0.5081	0.8548	59.34	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
likeschool	1	11.71	0.0006
age	1	105.60	<.0001
sex	1	62.73	<.0001

Figure 4: Poisson regression of number of days absent on sex, age, and likeschool, likeschool as continuous

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.2246	0.3659	-3.9418	-2.5073	77.64	<.0001
likeschool	5	1	0.5159	0.1468	0.2282	0.8037	12.35	0.0004
likeschool	4	1	0.6719	0.1731	0.3326	1.0112	15.07	0.0001
likeschool	3	1	0.2245	0.1354	-0.0409	0.4898	2.75	0.0973
likeschool	2	1	0.3705	0.1284	0.1189	0.6222	8.33	0.0039
likeschool	1	0	0.0000	0.0000	0.0000	0.0000	.	.
age		1	0.2185	0.0223	0.1747	0.2622	95.82	<.0001
sex		1	0.6769	0.0890	0.5025	0.8513	57.87	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
likeschool	4	21.17	0.0003
age	1	99.73	<.0001
sex	1	61.06	<.0001

Figure 5: Poisson regression of number of days absent on sex, age, and likeschool, likeschool as categorical

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	248	815.5013	3.2883
Scaled Deviance	248	815.5013	3.2883
Pearson Chi-Square	248	885.6149	3.5710
Scaled Pearson X2	248	885.6149	3.5710
Log Likelihood		-8.9824	
Full Log Likelihood		-611.8258	
AIC (smaller is better)		1231.6516	
AICC (smaller is better)		1231.8136	
BIC (smaller is better)		1245.7693	

(a) likeschool as continuous

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	245	806.0436	3.2900
Scaled Deviance	245	806.0436	3.2900
Pearson Chi-Square	245	882.5848	3.6024
Scaled Pearson X2	245	882.5848	3.6024
Log Likelihood		-4.2536	
Full Log Likelihood		-607.0970	
AIC (smaller is better)		1228.1940	
AICC (smaller is better)		1228.6530	
BIC (smaller is better)		1252.9000	

(b) likeschool as categorical

Figure 6: Goodness of fit statistics for Poisson regression of number of days absent on sex, age, and likeschool

From here on out, for simplicity, treat the likeschool variable as continuous rather than categorical.

(d) Give careful interpretations of the coefficients of the age, sex and likeschool variables and their confidence intervals (i) on the log scale and (ii) on the mean scale. Which of these variables appear to be significant predictors?

- age: The log mean number of days absent increases by 0.2229 units for each additional year in age, all else constant. We are 95% confident that the log mean number of days absent increases between 0.1795 and 0.2663 units for each additional year in age, all else constant. The mean number of days absent is 1.250 times as high for each additional year in age, all else constant. We are 95% confident that the mean number of days absent is anywhere between 1.197 to 1.305 times as high for each additional year in age, all else constant.
- sex: The log mean number of days absent is 0.6814 units higher for male students compared to female students, all else constant. We are 95% confident that the log mean number of days absent is between 0.5081 and 0.8548

units higher for male students as compared to female students, all else constant. The mean number of days absent is 1.977 times higher for male students as compared to female students, all else constant. We are 95% confident that the mean number of days absent for male students is between 1.662 to 2.351 times higher than for female students.

- likeschool: The log mean number of days absent increases by 0.1095 units for each additional score in how much the student liked school, all else constant. We are 95% confident that the log mean number of days absent increases between 0.0474 and 0.1717 units for each additional score in how much the student liked school, all else constant. The mean number of days absent is 1.116 times as high for each additional score in how much the student liked school, all else constant. We are 95% confident that the mean number of days absent is anywhere between 1.049 to 1.187 times as high for each additional score in how much the student liked school, all else constant.

All three variables appear to be significant predictors in this model ($p < .001$ for all three predictors).

(e) Now suppose that the number of absences for some of the students were measured over 1 month and some were measured over 3 months as indicated by the variable hmonth. Refit the model for the number of days absent using this variable as the offset. How does this change your answers to part (d)?

The impact of age on log mean number of absences per month/mean number of absences per month is noticeably decreased in this model as compared to the model without the offset term. Additionally, our interpretation of all the coefficients can now be comparable with the same denominator of "per month". Our answers in part (d) can now be about the log mean number of days absent per month/mean number of days absent per month, with the new coefficient values subbed in. We note that the direction and significance of all three predictors are the same in both models, age has a noticeably smaller magnitude (0.1881 unit increase in log mean number of absences per month for each additional year in age), and that the magnitudes of likeschool and sex are quite similar in both models.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.4546	0.3465	-4.1338	-2.7754	99.38	<.0001
likeschool	1	0.1114	0.0316	0.0496	0.1733	12.46	0.0004
age	1	0.1881	0.0221	0.1448	0.2313	72.67	<.0001
sex	1	0.6867	0.0886	0.5131	0.8604	60.09	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
likeschool	1	12.24	0.0005
age	1	74.78	<.0001
sex	1	63.51	<.0001

Figure 7: Poisson regression of number of days absent on sex, age, and likeschool, likeschool as categorical, hmonth as offset

Note: For the rest of the problem revert to thinking of the number of absences as all corresponding to one-month intervals as in parts (a)-(d).

(f) In this part of the problem we evaluate the issue of over-dispersion. First suggest some plausible reasons why there might be over-dispersion in this data set and then check for it in the following ways:

(1) Using basic descriptive statistics obtain the mean and variance of the nhookey variable. (i) What do these statistics suggest about over-dispersion and (ii) why might this assessment be too crude to evaluate

the presence of over-dispersion in the model overall?

(2) Evaluate the goodness of fit of the model using by calculating the Pearson chi-squared goodness of fit statistic and approximating the dispersion factor. Explain what each of these checks tells you.

(3) Obtain the Pearson residuals for the model and plot them as a function of the fitted values. What does this plot suggest about over-dispersion? Do there appear to be any outliers? Briefly justify your answer.

(1) The nhookey variable has a mean of 2.270 and a variance of 9.297. The variance is much bigger than the mean, which may suggest over-dispersion in the model. However, this is still a crude assessment because this removes information about the heterogeneity present in the variables included in the model, so we may be calculating the mean and variance of merged groups with very different characteristics.

(2) The Pearson chi-squared goodness of fit statistic was obtained in part (c), which was $\chi^2_{248} = 885.6149$. Divided by the degrees of freedom we get $885.6149/248 = 3.571$. If there is no over-dispersion, the dispersion factor should be equal to 1, but our value is greater than 1. The Pearson chi-squared goodness of fit statistic itself tells us that the model is not well-calibrated and is a poor fit, and the dispersion factor being greater than 1 tells us that there is proof of over-dispersion and that we would need to scale our standard errors accordingly with the square root of the dispersion factor.

(3) The Pearson residuals vs fitted values plot shows fanning of residuals increasing with larger fitted values, and the residuals are quite large, going up to 7.5 at the most fanned out data points. This suggests over-dispersion, as a model that does not have over-dispersion should have equal spread in the residuals vs fitted values plot.

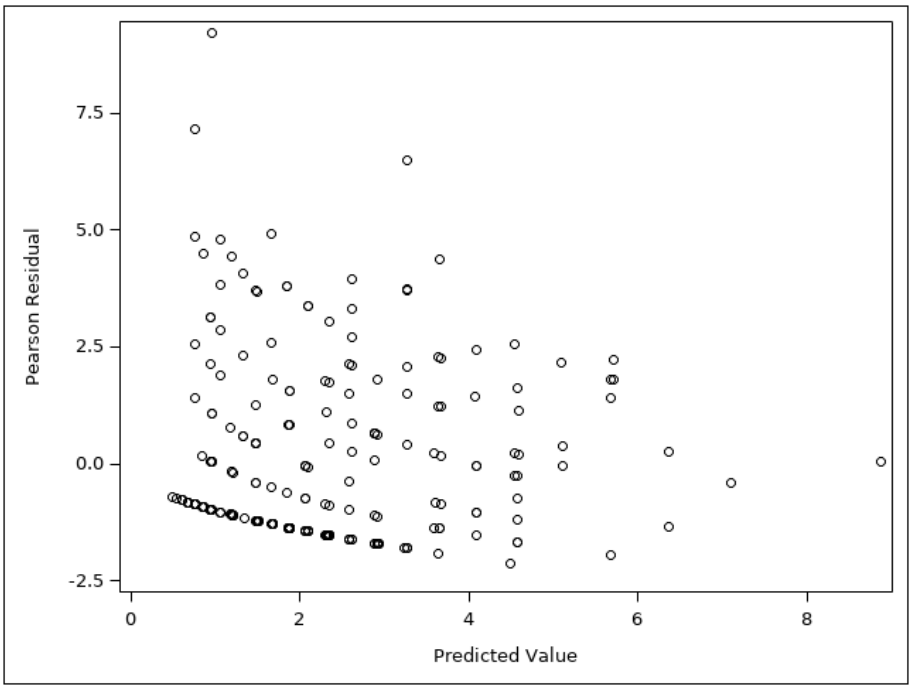

Figure 8: Pearson residuals vs fitted values for the Poisson regression model

(g) Rerun the model using a negative binomial regression. Does this model confirm your conclusions about over-dispersion from part (f)? Explain briefly.

The negative binomial model yields a dispersion parameter of 1.7512, with a 95% confidence interval of [1.3005, 2.3580]. Since this interval is entirely above 1, this model confirms my conclusions about over-dispersion from part (f) since we are 95% confident that the dispersion parameter lies between 1.3005 and 2.3580, a number greater than the desired dispersion parameter of 1.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	248	252.8602	1.0196
Scaled Deviance	248	252.8602	1.0196
Pearson Chi-Square	248	223.3531	0.9006
Scaled Pearson X2	248	223.3531	0.9006
Log Likelihood		128.7829	
Full Log Likelihood		-474.0605	
AIC (smaller is better)		958.1210	
AICC (smaller is better)		958.3649	
BIC (smaller is better)		975.7681	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.1864	0.7759	-4.7070	-1.6657	16.87	<.0001
likeschool	1	0.0939	0.0765	-0.0560	0.2439	1.51	0.2194
age	1	0.2251	0.0519	0.1233	0.3269	18.78	<.0001
sex	1	0.5329	0.1939	0.1528	0.9130	7.55	0.0060
Dispersion	1	1.7512	0.2658	1.3005	2.3580		

Note: The negative binomial dispersion parameter was estimated by maximum likelihood.

Figure 9: Negative binomial regression of number of days absent on sex, age, and likeschool, likeschool as continuous

(h) Give brief interpretations of the coefficients from the negative binomial regression. Are your conclusions (as to magnitude, direction and significance) of the effects of the predictors any different from those in part (d)?

- age: The mean number of days absent increases by a factor of $e^{0.2251} = 1.252$ for each additional year in age, all else constant.
- sex: The mean number of days absent is $e^{0.5329} = 1.704$ times higher in male students than in female students, all else equal.
- likeschool: The mean number of days absent increases by a factor of $e^{0.0939} = 1.098$ for each additional score in how much a student likes school, all else equal.

The direction of the predictors are the same as before; however, the magnitude of the predictor effect of likeschool and age is smaller than before. Additionally, likeschool is no longer significant in this model ($p = .2194$), while sex is still significant but with a larger p-value ($p = .006$), indicating that there is less evidence (but still enough evidence to determine a positive association) for the effect of sex on mean number of days absent. This is possibly in part due to the larger standard error estimates in the negative binomial model as compared to the Poisson model.

In the next part of the problem we consider the issue of zero inflation.

(i) Explain intuitively why we might expect zero-inflation in the context of this problem. Of our three predictors, which would you expect to be most associated with a "certain zero"? Explain briefly.

Zero-inflation in the context of this problem may be that some students will never miss a day of school, either because they have always have perfect attendance, parents do not allow them to play hookey, they like attending class, or

something else. Of the three predictors, I would most expect likeschool to be associated with a "certain zero" since those that like school would be more inclined to go to class, regardless of gender or age.

(j) Using the estimated rate of unexcused absences from your model in part (a) calculate the expected number of 0's we would see in a sample of 252 students if absences have a Poisson distribution. You may find it helpful to remember that for a Poisson random variable the probability of k events is $P(Y = k) = \mu^k e^{-\mu}/k!$). How many 0's did we actually observe? What does this suggest about whether we have a zero-inflation issue? (Note: This is a fairly crude approximation because we haven't accounted for the covariates but it is rather suggestive. You can of course take this approach further and break it down by sex, age bin and likeschool categories.)

The expected number of 0's we would see in a sample of 252 if absences have a Poisson distribution would be $P(Y = 0) = (2.27)^0 e^{-2.27}/0! = 0.103$, suggesting that 10.3% of the sample would be 0's. However, in the sample we observed $121/252 = 48\%$, suggesting that there is zero-inflation present here, even without the finer breakdown into variable categories.

(k) Fit a zero-inflated Poisson model to the data using age, sex and likeschool as the predictors for both the Poisson component and the zero-inflation component of the model. Does this model seem like an improvement over a standard Poisson regression? Explain briefly.

The model does seem like a better fit over the standard Poisson regression –we note that the log likelihood is significantly higher in the zero-inflated Poisson model as compared to the standard Poisson model, and the Pearson chi-square test statistic is lower in this model vs the standard Poisson model.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8904	0.4072	0.0922	1.6886	4.78	0.0288
likeschool	1	0.0319	0.0339	-0.0346	0.0985	0.88	0.3472
age	1	0.0037	0.0268	-0.0488	0.0562	0.02	0.8902
sex	1	0.6743	0.1022	0.4740	0.8745	43.55	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	7.7887	1.4080	5.0291	10.5483	30.60	<.0001
likeschool	1	-0.2503	0.1264	-0.4980	-0.0025	3.92	0.0477
age	1	-0.5007	0.0963	-0.6894	-0.3120	27.05	<.0001
sex	1	0.0900	0.3202	-0.5375	0.7175	0.08	0.7787

Figure 10: Zero-inflated Poisson model of number of days absent on sex, age, and likeschool

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance		908.1374	
Scaled Deviance		908.1374	
Pearson Chi-Square	244	323.5140	1.3259
Scaled Pearson X2	244	323.5140	1.3259
Log Likelihood		148.7747	
Full Log Likelihood		-454.0687	
AIC (smaller is better)		924.1374	
AICC (smaller is better)		924.7300	
BIC (smaller is better)		952.3728	

Figure 11: Goodness of fit criteria for zero-inflated Poisson model of number of days absent on sex, age, and likeschool

(l) Give brief interpretations of the coefficients for each component of the zero-inflated Poisson model. Which variables seem to be significant in each part of the model? Explain as carefully as you can what this model suggests about how these variables affect school absences.

- Count
- age: Among students with absences, the mean number of days absent is $e^{0.0037} = 1.004$ times higher for each additional year in age, all else equal.
- sex: Among students with absences, the mean number of days absent is $e^{0.6743} = 1.963$ times higher in male students than in female students, all else equal.
- likeschool: Among students with absences, the mean number of days absent is $e^{0.0319} = 1.032$ times higher for each additional score in how much a student likes school, all else equal.
- Inflate
- age: All else equal, each additional year in age is associated with a 39.4% reduction in the odds of having no absences (odds ratio $e^{-0.5007} = 0.606$).
- sex: All else equal, male students have 9.4% higher odds of having no absences than female students (odds ratio $e^{0.09} = 1.094$).
- likeschool: All else equal, each additional score in how much a student likes school is associated with a 22.1% reduction in the odds of having no absences (odds ratio $e^{-0.2503} = 0.779$).

In the count component of the ZIP model, sex is the only variable that is significant ($p < .001$). In the zero-inflated component of the model, both likeschool and age are significant ($p = .048$ and $p < .001$, respectively). This model suggests that students that are older or like school more are both more likely to never have absences. Of the students that do have absences, male students are more likely to have higher counts of absences than female students.

(m) Now fit a zero-inflated negative binomial model using the same model set-up as in part (k). Is this model superior to the zero-inflated Poisson model of part (k)? Does it appear to be an improvement over the plain negative binomial model of part (g)? Explain briefly in each instance. What does that tell you about the issues of over-dispersion and zero-inflation? Do any of your conceptual conclusions from part (l) change with this model?

This model is superior to the zero-inflated Poisson model since it has a dispersion parameter that is greater than 0 (dispersion = 0.3711, 95% CI = [0.2091, 0.6585]), so this model better accounts for the over-dispersion. This model appears to be an improvement over the plain negative binomial model, as we see that there is significance observed in the coefficients for age and sex in the inflated component. This tells us that there was over-dispersion and zero-inflation in our data and desired model that the ZINB model is best suited to address. In this model, likeschool is no longer significant in either the count or zero-inflated components, so that would change our conclusions about likeschool being associated with increases in mean number of days absent. Additionally, sex is now significant in both

the count and zero-inflated components. However, the standard error of sex in the zero-inflated component is so large that we cannot make a new conclusion about its role in determining a "certain 0" or not. So, we lose the conclusion about likeschool and keep the same direction and significance for sex in the count portion of the model and age in the zero-inflated portion of the model, as compared to in part (l).

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.1554	0.6293	-0.0779	2.3888	3.37	0.0663
likeschool	1	0.0635	0.0531	-0.0405	0.1675	1.43	0.2314
age	1	-0.0349	0.0396	-0.1126	0.0427	0.78	0.3776
sex	1	0.9023	0.1413	0.6255	1.1792	40.80	<.0001
Dispersion	1	0.3711	0.1086	0.2091	0.6585		

Note: The negative binomial dispersion parameter was estimated by maximum likelihood.

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	13.0952	2.5239	8.1485	18.0420	26.92	<.0001
likeschool	1	-0.2157	0.1601	-0.5295	0.0980	1.82	0.1777
age	1	-0.9588	0.1951	-1.3411	-0.5765	24.16	<.0001
sex	1	1.3119	0.5371	0.2593	2.3645	5.97	0.0146

Figure 12: Zero-inflated negative binomial model of number of days absent on sex, age, and likeschool

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance		873.8883	
Scaled Deviance		873.8883	
Pearson Chi-Square	244	292.6531	1.1994
Scaled Pearson X2	244	292.6531	1.1994
Log Likelihood		-436.9442	
Full Log Likelihood		-436.9442	
AIC (smaller is better)		891.8883	
AICC (smaller is better)		892.6322	
BIC (smaller is better)		923.6532	

Figure 13: Goodness of fit criteria for zero-inflated negative binomial model of number of days absent on sex, age, and likeschool