Extra-Binomial Variation and Extensions

Tate Jacobson

Department of Statistics
Oregon State University, Corvallis, Oregon

Spring 2026

Reading: Ch 21.5 – 21.7 of *The Sleuth*

Outline

1. Extra-Binomial Variation
    - Extra-Binomial Variation
    - Quasi-Likelihood
    - Estimating Dispersion
    - Case Study 21.1.2: Modeling Predation Rates
    - Interpreting the Model
2. Multinomial Regression

Extra-Binomial Variation

Extra-Binomial Variation

Recall that if $X \sim Bin(m, \pi)$, then

$$mean\{X\} = m\pi \text{ and } Var\{X\} = m\pi(1-\pi).$$

Often, however, the single parameter, π, is inadequate for describing *both* the mean and the variance of a count random variable, X.

- In particular, there is often *more variability in count data than can be described by the binomial distribution*.
- This is called **extra-binomial** variation.

What’s the Problem?

When extra-binomial variation is present, the standard errors for the logistic regression coefficient estimates are TOO SMALL.

- This means that p-values based on these SE’s are
  and confidence intervals built with these SE’s are
- In this case, Wald tests and confidence intervals are not just approximate, they are misleading.

The Quasi-Likelihood Model

We can deal with extra-binomial variation by instead using a **Quasi-Likelihood Model**:

In this model, we don't assume that the responses come from a Binomial distribution, we just specify their mean and variance (which happen to be fairly similar to—i.e. quasi—the binomial mean and variance):

$$\begin{aligned}
\mu\{Y_i|X_{i1},\ldots,X_{ip}\} &= m_i\pi_i \\
Var\{Y_i|X_{i1},\ldots,X_{ip}\} &= \psi m_i\pi_i(1-\pi_i) \\
\text{logit}(\pi_i) &= \beta_0+\beta_1X_{i1}+\cdots+\beta_pX_{ip}
\end{aligned}$$

- The parameter ψ is called the **dispersion** parameter.
- $\psi > 1$ represents extra-binomial variation (or **over-dispersion**).

Checking for Over Dispersion

Extra-binomial variation is a form of **over-dispersion.** To check for over-dispersion:

1. Think about whether it might be likely for this particular response (dependence among binary random variables comprising the counts; unaccounted for explanatory information).
2. Examine the deviance goodness-of-fit after fitting a rich model.
3. Examine the residuals—over dispersion could just be due to one or two large outliers.

An Estimate for the Dispersion Parameter

Remember that under the quasi-likelihood parameterization, the variance of a count, Y_i, is assumed to be $\psi m_i \pi_i(1-\pi_i)$, where m_i is the size (denominator) associated with Y_i.

- One estimate for ψ is:

  $$\hat{\psi} = \frac{\text{Deviance}}{\text{Degrees of Freedom}}$$

  (recall that the degrees of freedom are $n-p-1$).
- $\hat{\psi}$ should be approximately equal to one if the data are binomial, and larger that one if there is over dispersion.
- Note: Some estimate ψ using the sum of squared Pearson residuals rather than the deviance. We'll use the deviance version throughout these slides.

Maximum Quasi-Likelihood Estimates

Standard errors of the maximum quasi-likelihood estimates are $\sqrt{\hat{\psi}}$ times the standard errors for the maximum likelihood estimates, i.e.

$$SE(\hat{\beta}_j^{QL}) = \sqrt{\hat{\psi}} \times SE(\hat{\beta}_j^{MLE})$$

- This makes some intuitive sense: if we have more variation in the responses, we should have more variation in our estimates.

Adjusted Tests and CIs for Coefficients

We can perform tests and build confidence intervals for the regression coefficients as before, with two changes:

1. We **use the adjusted standard errors**, so our test statistic for testing H$_0 : \beta_j = 0$ is

$$t = \frac{\hat{\beta}_j^{MLE} - 0}{SE(\hat{\beta}_j^{QL})} = \frac{\hat{\beta}_j^{MLE} - 0}{\sqrt{\hat{\psi}} \times SE(\hat{\beta}_j^{MLE})}$$

2. We use a **t-distribution** with $n - p - 1$ degrees of freedom to perform the tests and build the confidence intervals (t-distribution is more conservative than a standard normal distribution), so our two-sided p-value is

$$p\text{-value} = 2 \times P(t_{n-p-1} > |t|)$$

and we use t distribution critical values in the CI.

Drop-in-Deviance F-tests

Performing a drop in deviance test in the presence of over dispersion also involves an adjustment.

- Instead of using the unscaled drop in deviance (i.e., just the difference in deviance between two models), we use the following F-statistic:

  $$F = \frac{(deviance_{reduced} - deviance_{full})/d}{\hat{\psi}},$$

  where d is the number of parameters different between the two models and $\hat{\psi}$ is the estimate of the dispersion parameter from the full model.
- The p-value for this test is obtained from an **F-distribution** on d and $n - p - 1$ degrees of freedom:

  $$p\text{-value} = Pr(F_{d,n-p-1} > F).$$

Case Study 21.1.2: Moth Predation Data

A **cline** is a region where two color or physiological morphs of one species arrange themselves at opposite ends of an environmental gradient. A cline exists near Liverpool, England:

- In Liverpool, trees trunks are blackened by smoke, and there is a high percentage of dark morph moths.
- In the countryside around Liverpool, there is a higher percentage of the typical salt-and-pepper moths.

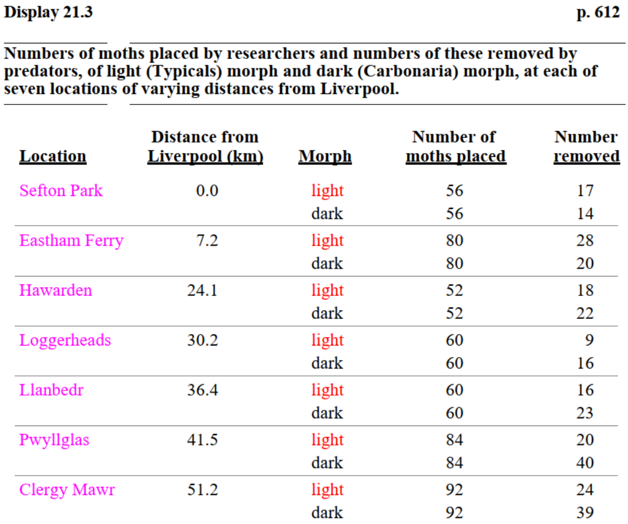
An Experiment to Study Natural Selection

Consider the following study:

- Seven locations, progressively farther from Liverpool.
- Eight trees chosen randomly at each location.
- Equal number of frozen light and dark colored moths were glued to each tree in “life-like” positions.
- After 24 hours, a count was taken of the number from each morph on each tree that was removed (presumably by predation).

Question: Do the relative odds of predation for the two species change with distance from Liverpool?

Moth Predation Data



Display 21.3 **p. 612**

Numbers of moths placed by researchers and numbers of these removed by predators, of light (Typicals) morph and dark (Carbonaria) morph, at each of seven locations of varying distances from Liverpool.

Location	Distance from Liverpool (km)	Morph	Number of moths placed	Number removed
Sefton Park	0.0	light	56	17
		dark	56	14
Eastham Ferry	7.2	light	80	28
		dark	80	20
Hawarden	24.1	light	52	18
		dark	52	22
Loggerheads	30.2	light	60	9
		dark	60	16
Llanbedr	36.4	light	60	16
		dark	60	23
Pwyllglas	41.5	light	84	20
		dark	84	40
Clergy Mawr	51.2	light	92	24
		dark	92	39

Thinking about Over-Dispersion

There are a few possible sources of over-dispersion in the moth data:

1. The predation acts may not be independent, and therefore the binary predation events that make up the counts are probably not independent.
2. The binary predation events may not have the same probabilities at the same site.
3. The logistic regression model may not be correct (two sites equidistant from Liverpool may not have the same tree coloration properties).

Any and all of these may lead to over-dispersion in the counts.

Thinking About our Explanatory Variables

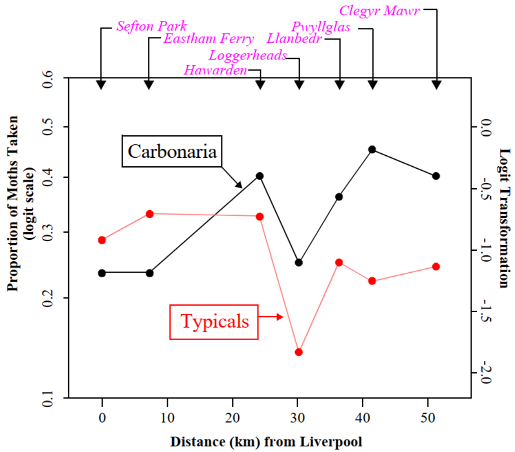
Remember that a big question we have is: do the predation rates of the differently colored moths change with tree color? Specifically: are the light moths more heavily predated than dark moths close to Liverpool AND are they less heavily predated than dark moths farther from Liverpool?

- Distance from Liverpool is being used as a proxy for tree color—blacker trees are closer to Liverpool, naturally colored trees are farther out.
- Can we visualize whether relative predation rates change with distance?

Evidence of a Change in Relative Predation?

Display 21.4 **p. 613**

Proportions of moths of two color morphs taken by predators at seven locations near Liverpool, England



Explanatory Variable Interactions

In terms of a logistic regression model, what we're after is whether there is evidence of an **interaction** between moth color and distance.

- Recall that if X_1 and X_2 are two explanatory variables that we're using to model a response, Y, then X_1 and X_2 are said to *interact* if the effect of X_1, say, on Y changes for different values (levels) of X_2.
- In our example, we can ask: is the effect of tree color (distance) on the odds of predation change different for the different colored moths? Based on the previous plot, this seems plausible.

Inference for the Moth Predation Model

Let's consider the following model with a dark×distance interaction:

$$\text{logit}(\pi) = \beta_0 + \beta_1 dark + \beta_2 distance + \beta_3 dark \times distance.$$

Let's fit the model and plot the deviance residuals in R.

Based on possible correlation between the Bernoulli trials and some large residuals, we decide to account for over-dispersion.

Exercise (10 minutes): Using the model output on the next slide:

1. Estimate the dispersion parameter $\hat{\psi}$ using the deviance.
2. Compute the SE of $\hat{\beta}_3$ under the quasi-likelihood model.
3. Test the hypothesis $H_0 : \beta_3 = 0$ vs $H_a : \beta_3 \neq 0$.

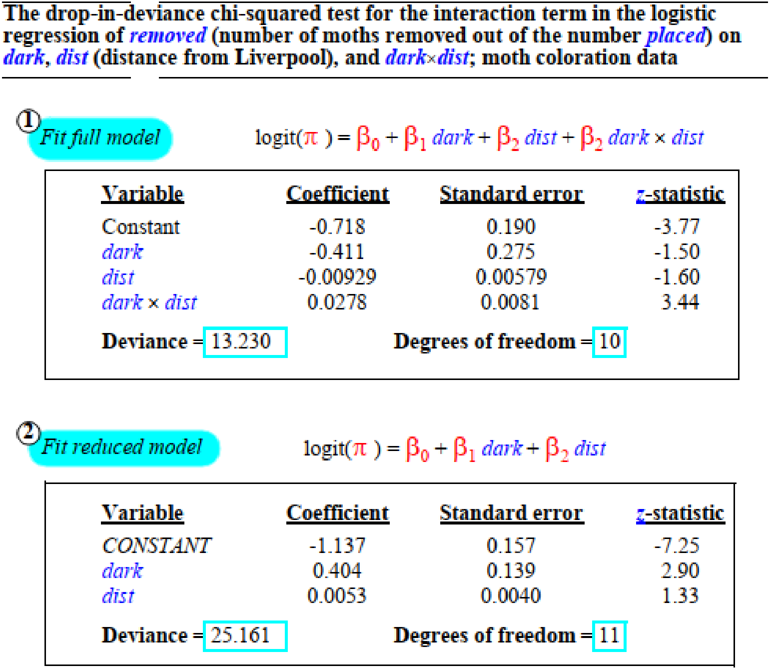
```
Call:
glm(formula = resp ~ dark * Distance, family = binomial)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.717729   0.190205  -3.773 0.000161 ***
darkdark          -0.411257   0.274490  -1.498 0.134066
Distance          -0.009287   0.005788  -1.604 0.108629
darkdark:Distance  0.027789   0.008085   3.437 0.000588 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 35.385  on 13  degrees of freedom
Residual deviance: 13.230  on 10  degrees of freedom
AIC: 83.904

Number of Fisher Scoring iterations: 4
```

Solution:

- Estimate the dispersion parameter $\hat{\psi}$ using the deviance:
- Compute the SE of $\hat{\beta}_3$ under the quasi-likelihood model:
- Test the hypothesis $H_0 : \beta_3 = 0$:

Drop in Deviance Test with Extra-Binomial Variation

Exercise (4 minutes): Using the model output on the next slide, perform a drop-in-deviance F-test to account for extra-binomial variation.



The drop-in-deviance chi-squared test for the interaction term in the logistic regression of *removed* (number of moths removed out of the number *placed*) on *dark*, *dist* (distance from Liverpool), and *dark*×*dist*; moth coloration data

① *Fit full model*

$$\text{logit}(\pi) = \beta_0 + \beta_1\, dark + \beta_2\, dist + \beta_2\, dark \times dist$$

Variable	Coefficient	Standard error	z-statistic
Constant	-0.718	0.190	-3.77
dark	-0.411	0.275	-1.50
dist	-0.00929	0.00579	-1.60
dark × *dist*	0.0278	0.0081	3.44

Deviance = 13.230 **Degrees of freedom =** 10

② *Fit reduced model*

$$\text{logit}(\pi) = \beta_0 + \beta_1\, dark + \beta_2\, dist$$

Variable	Coefficient	Standard error	z-statistic
CONSTANT	-1.137	0.157	-7.25
dark	0.404	0.139	2.90
dist	0.0053	0.0040	1.33

Deviance = 25.161 **Degrees of freedom =** 11

Drop in Deviance Test with Extra-Binomial Variation

Solution:

What's our conclusion?

Let’s go to R to see how we can perform the coefficient test and F test that account for over-dispersion.

Quasi-Likelihood Model in R

We've seen how to account for extra-binomial variation by-hand. The `glm` function also includes a `family = quasibinomial` option

```
glm(resp ~ dark*Distance, family = quasibinomial)
```

which accounts for over-dispersion (though with a different estimate) in the coefficient SEs and tests.

Let's go to R to see how to fit a `quasibinomial` model.

The `quasibinomial` model output looks similar, but the SEs account for extra-binomial variation and the dispersion parameter estimate is not taken to be 1.

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.717729   0.214423  -3.347   0.0074 **
dark          -0.411257   0.309439  -1.329   0.2134
Distance      -0.009287   0.006525  -1.423   0.1851
dark:Distance  0.027789   0.009115   3.049   0.0123 *
---

(Dispersion parameter for quasibinomial family taken to be 1.270859)

Null deviance: 35.385  on 13  degrees of freedom
Residual deviance: 13.230  on 10  degrees of freedom
```

Interpreting the Moth Predation Model

Our model is:

$$\text{logit}(\pi) = \beta_0 + \beta_1 dark + \beta_2 distance + \beta_3 dark \times distance.$$

- Here, $dark$ is an indicator variable that is equal to one for the dark morphs and zero for the light morphs (i.e., the light morph is the reference level).

Interpreting the Moth Predation Model

Our model is:

$$\text{logit}(\pi) = \beta_0 + \beta_1 dark + \beta_2 distance + \beta_3 dark \times distance.$$

- Notice that β_0 is the log odds of predation of light morphs at *Sefton Park* (for which $distance$ is zero) and $(\beta_0 + \beta_1)$ is the log odds of predation of dark morphs at *Sefton Park*.
- Therefore $\beta_1 = (\beta_0 + \beta_1) - \beta_0$ is the difference in the log odds of predation for dark morph and light morph moths at *Sefton Park*.

Interpreting an Interaction

Our model is:

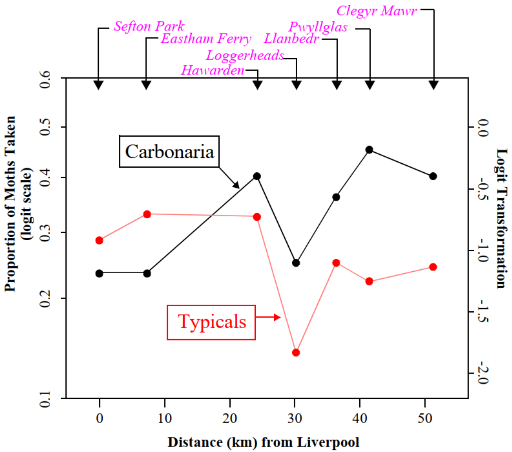
$$\text{logit}(\pi) = \beta_0 + \beta_1 dark + \beta_2 distance + \beta_3 dark \times distance.$$

- The log odds of predation of the light morph at distance D is $\beta_0 + \beta_2 D$, and the log odds of predation of the dark morph at distance D is $(\beta_0 + \beta_1) + (\beta_2 + \beta_3)D$.
- Due to the interaction, the log odds change at different rates ($\beta_2 + \beta_3$ vs β_2) for the dark and light morphs. That is, they have different *slopes*.
- In particular, β_3 is the difference in the slopes (of the log odds as a function of distance) for the dark and light morphs.

Including an interaction in our model of moth predation rates allows for some greater flexibility. However, looking at Display 21.4 we might wonder whether we need a more complex model?

Display 21.4 **p. 613**

Proportions of moths of two color morphs taken by predators at seven locations near Liverpool, England



Including Location and Distance

Suppose we create indicator variables for 6 of the 7 locations. Using those, we can allow each location to have it's own odds of predation for the light morphs.

- Then, by retaining the terms $dark$ and $dark \times distance$, we can let those location-specific odds change according to morph and distance from Liverpool.
- Notice that it's important to keep $dark \times distance$ in the model, because it's relevant to the research question we're after.
- So now, we'll look at intermediate model with 9 parameters.
- Note: the goodness-of-fit test has a p-value of 0.72 suggesting a good model fit. As such, we won't worry about over-dispersion.

Model Results

```
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.766921   0.246214  -3.115 0.001840 **
EF             0.020374   0.274405   0.074 0.940814
HW             0.163394   0.305404   0.535 0.592643
LH            -0.797830   0.329669  -2.420 0.015517 *
LB            -0.282134   0.321641  -0.877 0.380395
PG            -0.213749   0.313693  -0.681 0.495620
CM            -0.434293   0.339809  -1.278 0.201232
dark          -0.405455   0.275187  -1.473 0.140648
dark:Distance  0.027739   0.008097   3.426 0.000613 ***
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 35.3853  on 13  degrees of freedom
Residual deviance:  2.8757  on  5  degrees of freedom
```

Interpreting Results

- Once we decide which model to use, we can take the estimate of the regression coefficient for the interaction term, $dark \times distance$, and use it to report the change in the odds ratio of dark morph predation to light morph predation with distance.
- For example, from the last model, the estimated odds ratio of dark to light predation increase by almost 32% with a 10km increase in distance from Liverpool (this is just $\exp(10 \times 0.0277)$). An approximate 95% confidence interval for this increase is 13% to 55%. (Q: How can you get this CI?)

  Let's see how to get this CI in R.

In our last lecture on binary logistic regression, we introduced the **binary logistic regression modeling workflow**.

With binomial logistic regression, checking for over-dispersion adds a few extra steps and can affect how we perform statistical inference.

Binomial Logistic Regression Modeling Workflow

1. Determine whether binomial distribution is reasonable for the response
2. Examine exploratory plots
3. Consider transformations of explanatory variables
4. Fit a rich model (with interactions, polynomial terms)
5. Check for over-dispersion using rich model (goodness of fit, test residuals)
6. Determine whether a quasi-likelihood model is needed. If so, use it for the rest of your analysis.
7. Perform variable selection (drop-in-deviance, AIC (if applicable))
8. Pick a final model and report estimates
9. Interpret results to answer research questions

Multinomial Regression

Multinomial Responses

- So far, we've been talking about binary and binomial responses—essentially, responses that can take one of two values: 0/1, yes/no, survived/died, etc.
- A **multinomial** response can take on 3 or more values. Some examples:
  1. Bush/Gore/Nader (2000 Presidential candidates)
  2. Plant/Animal/Both (stomach contents)
  3. Pool with wood/Riffle/Other pool (stream habitat)

The Multinomial Distribution

The Binomial probability function is a special case of the multinomial probability function.

- Remember that the Binomial probability function is:

$$Pr(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}.$$

  where π is the probability of success, or the binomial probability.
- In the Binomial case, X just counts up the number of successes.
- With multinomial responses, we have to count up the numbers of observations in each of several (3 or more) categories.

The Multinomial Distribution

Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a multinomial random variable, with size M and probabilities, $(\pi_1, \pi_2, \ldots, \pi_k)$. It's important to note that:

$$\sum_{j=1}^{k} X_j = M \text{ and } \sum_{j=1}^{k} \pi_k = 1.$$

- The multinomial probability distribution is just:

$$Pr(\mathbf{X} = \mathbf{x}) = \frac{n!}{X_1! X_2! \cdots X_k!} \pi_1^{x_1} \pi_2^{x_2} \cdots \pi_k^{x_k}.$$

- Instead of one probability, π to estimate, we have $k-1$ probabilities, $\pi_1, \pi_2, \ldots, \pi_{k-1}$.

Multinomial Logistic Regression

In a multinomial logistic regression model, we:

1. pick one category as the **reference category**, say category k.
2. form each of the *multinomial logits*:

$$\log\left(\frac{\pi_1}{\pi_k}\right), \log\left(\frac{\pi_2}{\pi_k}\right), \ldots, \log\left(\frac{\pi_{k-1}}{\pi_k}\right).$$

3. then, for each of these logits, form the models:

$$\log\left(\frac{\pi_j}{\pi_k}\right) = \beta_{0j} + \beta_{1j}X_1 + \cdots + \beta_{pj}X_p.$$

Notice that **each multinomial logit has its own logistic regression coefficients.**

Recap

Material covered: Ch 21.5 - 21.7 of *The Sleuth*

- Extra-binomial variation/over-dispersion
- Inferential procedures which accounts for over-dispersion
- Interpreting logistic regression models with interactions
- Binomial logistic regression modeling workflow
- Multinomial regression