



Expert Tutorial

The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression

Tom Loeys^{1*}, Beatrijs Moerkerke¹, Olivia De Smet²
 and Ann Buysse²

¹Department of Data Analysis, Ghent University, Belgium

²Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

Infrequent count data in psychological research are commonly modelled using zero-inflated Poisson regression. This model can be viewed as a latent mixture of an “always-zero” component and a Poisson component. Hurdle models are an alternative class of two-component models that are seldom used in psychological research, but clearly separate the zero counts and the non-zero counts by using a left-truncated count model for the latter. In this tutorial we revisit both classes of models, and discuss model comparisons and the interpretation of their parameters. As illustrated with an example from relational psychology, both types of models can easily be fitted using the R-package *pscl*.

I. Introduction

It is not uncommon in psychological research that the outcome of interest is counting the occurrence of a behavioural event. In relational psychology, for example, the number of steps taken towards separation and divorce, or the number of post-breakup unwanted pursuit behaviours in former partners may be measurements of interest (e.g., Atkins & Gallop, 2007; De Smet, Buysse, & Brondeel, 2011). In addictive behaviour research, participants may report over a prespecified period the number of times they used hard drugs, the number of times they drove a car when they felt at least drunk, etc. (Lewis, Clayton, Geisner, Lee, Kilmer, & Atkins, 2010). Paediatric psychologists may focus on the number of unintentional injuries that occurred during a certain age period (e.g., Karazsia & van Dulmen, 2008). Such count data are typically very skewed and exhibit a lot of zero count observations. It is well known by now that analysing these data using classical linear models is mostly inappropriate, even after transformation of outcome variables (Vives,

*Correspondence should be addressed to Tom Loeys, Department of Data Analysis, Ghent University, Dunantlaan 1, 9000 Ghent, Belgium (e-mail: tom.loeys@ugent.be).

Losilla, & Rodrigo, 2006). Standard statistical textbooks (e.g., Agresti, 2002) present Poisson regression as a possible tool for the analysis of count data. Unfortunately, the underlying Poisson distribution has several implications that are often neglected. First, it assumes that the mean and the variance are equal. However, count data often exhibit larger variance than predicted by the mean (overdispersion). Alternatives overcoming this violation include Poisson regression with an overdispersion parameter or the negative binomial distribution (Hilbe, 2011). A second issue with Poisson regression might be that the observed number of zero counts exceeds the predicted number of zero counts. Zero-inflated Poisson (Lambert, 1992) or zero-inflated negative binomial models offer a solution here. These models are mixture models in which the complete distribution of the outcome is represented by two separate components, a first part modelling the probability of excess zeros and a second part accounting for the non-excess zeros and non-zero counts. Hurdle models (Mullahy, 1986) are an alternative class of mixture models more commonly used in other domains such as econometrics (Cameron & Trivedi, 1998). In contrast to the zero-inflated model, the zero and non-zero counts are separated in the hurdle model.

In this tutorial we formally introduce the zero-inflated Poisson model and the Poisson logit hurdle model, as well as their negative binomial extensions. Real-world data, with the outcome of interest counting the number of unwanted pursuit behaviours in ex-partners, are used to illustrate model fit and interpretation of parameters.

2. Motivating example: Modelling unwanted pursuit behaviour

Our motivating example is a subsample of the Interdisciplinary Project for the Optimization of Separation (IPOS) trajectories conducted in Flanders (www.scheidingsonderzoek.be) that aims to gain insight into separation trajectories. More specifically, we focus on a sample of 387 participants who responded to an adapted version of a Relational Pursuit-Pursuer Short Form (RP-PSF; Cupach & Spitzberg, 2004) used to assess the extent of unwanted pursuit behaviour (UPB) perpetrations displayed since the time the couple broke up. The total of 28 items (ranging from 'leaving unwanted gifts' to 'threatening to hurt yourself'), each measured on a five-point Likert scale (from 0 = never to 4 = over five times), was used as an overall index of perpetration (with higher scores indicating higher levels of perpetrations). A participant who answered 'never' to all these 28 UPB items will have an UPB count equal to 0, while a participant who answered 'over five times' to 'leaving unwanted gifts' and 'never' to all other items will, for example, have an UPB count equal to 4.

In this example we will explore the impact of two predictors on this outcome: a binary indicator for 'education level' (0 = lower than bachelor's degree, or 1 = at least bachelor's degree), and a continuous measurement for the level of 'anxious attachment' in the former partner relationship. The latter was measured using a total of five anxious attachment items (e.g., 'My desire to be very close sometimes scared my ex-partner away') from an adapted Experience in Close Relationships Scale-Short Form (ECR-S; Wei, Russell, Mallinckrodt & Vogel, 2007) with results normalized to a *z*-score.

Since the primary focus of this expert tutorial is to explain in detail the interpretation of zero-inflated regressions and hurdle regressions and to highlight existing misconceptions of these models, the example is mainly used to illustrate these conceptual issues and – for simplicity – limited to these 2 predictors. A more in-depth psychological review of the UPB data is described in De Smet *et al.* (2011).

All analyses described below are performed in R, version 2.12.1 (R Development Core Team, 2011). The complete R code needed to perform these analyses as well as

the data and further supplementary material are available on Wiley Online Library. Other examples of well-illustrated R code for the analysis of count data can also be found in Atkins and Gallop (2007) (though these authors do not consider the hurdle model), and in Zeileis, Kleiber, and Jackman (2008) (for a detailed description of the *pscl* package).

Figure 1 shows the left-skewed distribution of the number of UPB perpetrations, with a preponderance of zeros. The maximum number of UPBs observed was 34, but almost 95% of the observations fell below or were equal to 10, and so for clarity we grouped together all observations larger than 10 in a single category in the histogram. After a formal introduction of the (zero-inflated) Poisson and negative binomial model and hurdle model in the next section, we assess how well these models predict this observed UPB perpetration distribution. Next, we will contrast the interpretation of the estimated parameters from the zero-inflated model with those from the hurdle model in the simplest setting of a single binary predictor – education level in this particular example. The section thereafter discusses a multiple testing issue arising from the use of mixture models and illustrates how to test whether education level impacts the UPB count distribution. We conclude by showing how to interpret and present effects from the most complete model, allowing for main effects of education level and anxiety and their interaction.

3. Zero-inflated count distributions

Zero-inflated Poisson (ZIP) models for handling zero-inflated count data were first introduced by Lambert (1992). Consider a sample of n independent observations of counts Y_i . In ZIP regression, the counts Y_i equal 0 with probability p_i and follow a Poisson distribution with mean μ_i with probability $1 - p_i$. The ZIP model can thus be seen as a mixture of two component distributions. It can be derived that

$$\Pr(Y_i = 0) = p_i + (1 - p_i) \exp(-\mu_i), \quad (1)$$

$$\Pr(Y_i = k) = (1 - p_i) \exp(-\mu_i) \mu_i^k / k!, \quad k = 1, 2, \dots \quad (2)$$

Following (1), it is important to acknowledge that zero observations arise from both the zero-component distribution and the Poisson distribution. The zero-component distribution is therefore related to model the ‘excess’ or ‘inflated’ zeros that are observed in addition to the zeros that are expected to be observed under the assumed Poisson distribution.

Hurdle models were originally proposed by Mullahy (1986) in the econometrics literature. Like the ZIP model, the Poisson logit hurdle (PLH) model is a two-component model: a hurdle component models the zero versus the non-zero counts, and a truncated Poisson count component is employed for the non-zero counts:

$$\Pr(Y_i = 0) = p_i^*, \quad (3)$$

$$\Pr(Y_i = k) = (1 - p_i^*) \frac{\exp(-\mu_i^*) (\mu_i^*)^k / k!}{1 - \exp(-\mu_i^*)}, \quad k = 1, 2, \dots \quad (4)$$

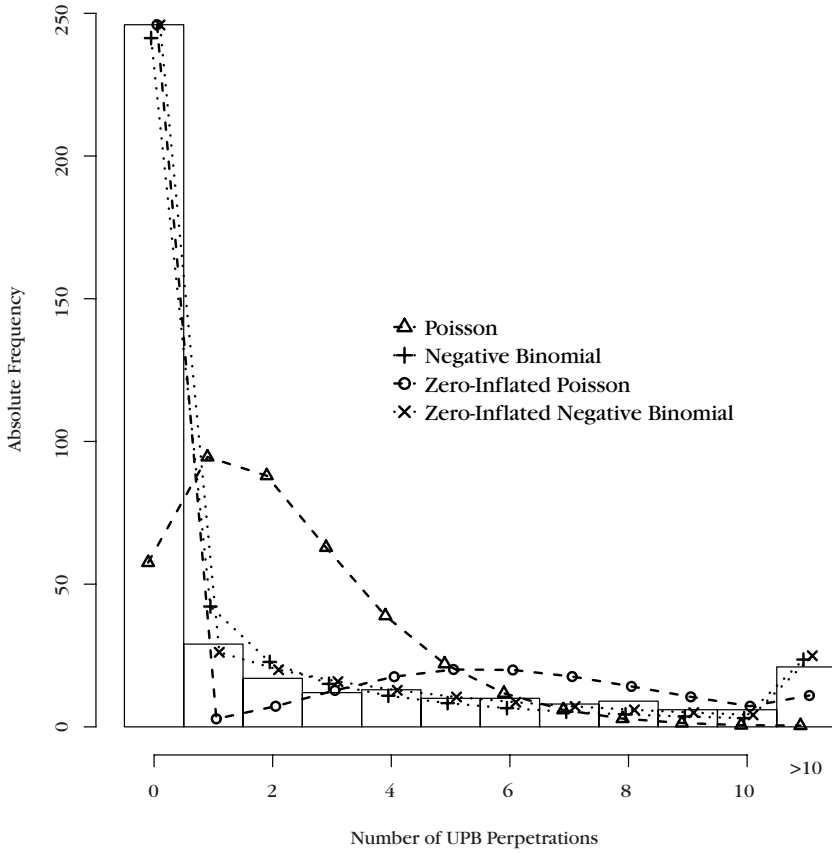


Figure 1. The observed distribution of number of UPB perpetrations. The predicted distributions shown are based on a Poisson model, a negative binomial model, a zero-inflated Poisson model and zero-inflated negative binomial model. The Poisson logit hurdle model and negative binomial logit hurdle yielded fits nearly identical to those of their zero-inflated counterparts.

Unlike p_i in the zero-inflated model (1), p_i^* in (3) does not model the *excess* zeros, but *all* zeros.

To assess the impact of covariates on the count distribution in a ZIP model, p_i and μ_i can be explicitly expressed as a function of covariates. The most natural choice to model the probability of excess zeros is to use a logistic regression model,

$$\text{logit}(p_i) = x_i^T \beta, \quad (5)$$

where x_i represents a vector of covariates and β a vector of parameters. Other link functions (probit, complementary log-log, cauchit or log link) can be used as well, but will not be considered here. The impact of covariates on count data excluding the excess zeros can be modelled through Poisson regression:

$$\log(\mu_i) = x_i^T \gamma. \quad (6)$$

Similarly, for the PLH model, the most natural choice to model the probability of zeros is to use a logistic regression model,

$$\text{logit}(p_i^*) = x_i^T \beta^*, \quad (7)$$

while the impact of covariates x_i on strictly positive (i.e., censored) count data can be modelled through Poisson regression,

$$\log(\mu_i^*) = x_i^T \gamma^*. \quad (8)$$

Estimates for β and γ (β^* and γ^*) in zero-inflated models (5) and (6) (or in hurdle models (7) and (8)) can be obtained by maximum likelihood estimation.

As mentioned in the introduction, count data often exhibit more variability than predicted by the mean of a Poisson distribution, even after accounting for excess zeros. A way of modelling overdispersed zero-inflated count data is to assume a zero-inflated negative binomial (ZINB) distribution for Y_i :

$$\Pr(Y_i = 0) = p_i + (1 - p_i) \frac{\theta^\theta}{(\mu_i + \theta)^\theta}, \quad (9)$$

$$\Pr(Y_i = k) = (1 - p_i) \frac{\Gamma(k + \theta)}{\Gamma(\theta) k!} \times \frac{\mu_i^k \theta^\theta}{(\mu_i + \theta)^{k+\theta}}, \quad k = 1, 2, \dots, \quad (10)$$

with mean μ_i and shape parameter θ ; Γ is the gamma function. The generalized negative binomial (Cameron & Trivedi, 1998), which additionally allows the shape parameter θ to depend on i , will not be considered here. When $\theta = 1$, the negative binomial distribution reduces to the geometric distribution.

Similarly, for the hurdle models, the negative binomial distribution (NBLH) can be used instead of the Poisson distribution in case of overdispersion:

$$\Pr(Y_i = 0) = p_i^*, \quad (11)$$

$$\Pr(Y_i = k) = (1 - p_i^*) \frac{\Gamma(k + \theta)}{\Gamma(\theta) k!} \times \frac{(\mu_i^*)^k \theta^\theta}{(\mu_i^* + \theta)^{k+\theta}} \times \frac{1}{1 - \theta^\theta / (\mu_i^* + \theta)^\theta}, \quad k = 1, 2, \dots, \quad (12)$$

where, for ease of notation, we use the same scale parameter θ as in the zero-inflated model. While both the zero-inflated and hurdle model need distributional assumptions for their count component, it is worth noting that both classes differ with respect to their dependencies of the estimation of the parameters of the 'zero' component on these assumptions. Indeed, unlike the zero-inflated model, the estimation of the parameters β^* related to p_i^* in the hurdle model is not dependent on the estimation of the parameters γ^* related to μ_i^* (Dalrymple, Hudson, & Ford, 2003). Hence, if assumptions about the (truncated) Poisson/negative binomial model are violated (e.g., due to extreme outlying observations), the hurdle model will – in contrast to the zero-inflated model – still yield consistent estimators for the parameters in the logit part of the model (if correctly specified).

4. Model comparison and goodness of fit

How well do the (zero-inflated) Poisson or negative binomial distributions introduced in the previous section capture the observed UPB count distributions in Figure 1? We fitted a Poisson model (P), a negative binomial model (NB), a zero-inflated Poisson model (ZIP), a zero-inflated negative binomial model (ZINB), a Poisson logit hurdle model (PLH) and a negative Binomial logit hurdle model (NBLH) with main effects for the two predictors, education level and anxious attachment level. To fit these models we used the functions `glm.nb()` from the MASS package (Venables and Ripley, 2002) and `zeroinfl()` and `hurdle()` from the `pscl` package (see supplementary R code).

The choice between nested models (e.g., P versus NB) can be made using a likelihood ratio test (LRT). However, the usual asymptotics do not apply; since the null hypothesis of a Poisson model corresponds to a parameter value for the shape parameter on the boundary of the parameter space, the appropriate asymptotic distribution for this LRT statistic under the null hypothesis should use a probability mass of 0.5 at 0 and $0.5\chi^2_{(1)}$ distribution above 0. Alternatively, the LRT statistic as $\chi^2_{(1)}$ gives a conservative test. Either way, we have overwhelming evidence ($p < .001$) for overdispersion from the LRT in our example (see Table 1, column P versus column NB).

To assess the ZINB model versus the NB model (or ZIP versus P), the restricted model requires p_i to vanish. Note that setting the parameters β to zero does not produce the restricted model; it produces $p_i = .5$. Rather, this requires some elements of β in the zero-inflated model to explode, and so a simple LRT cannot be used here either. A possible way to deal with this is to use Vuong's test (Vuong, 1989). In our example, Vuong's test ($p = .005$) prefers zero-inflated negative binomial above the negative binomial.

To choose between non-nested mixture models (e.g., NBLH versus ZINB) one could rely on the Akaike's information criterion (AIC), preferring the model with the smaller AIC. However, it turns out that in practice there is no or little difference in AIC between the hurdle model and zero-inflated model (Cameron & Trivedi, 1998). Indeed, it is interesting to note that – for a single binary predictor – the ZINB model can be seen as reparametrization of the NBLH model, and vice versa. The parameters of the count part γ^* are then equal to γ (including the scale parameter), while β_0^* and β_1^* are given by

$$\exp(\beta_0^*) = \frac{1 + \exp(\beta_0)}{1 - \theta^\theta / (\exp(\gamma_0) + \theta)^\theta} - 1, \quad (13)$$

$$\exp(\beta_1^*) = \frac{\{1 + \exp(\beta_0 + \beta_1)\} / \{1 - \theta^\theta / (\exp(\gamma_0 + \gamma_1) + \theta)^\theta\} - 1}{\exp(\beta_0^*)}. \quad (14)$$

A fortiori, the ZINB and NBLH models are equivalent for any saturated model. As a consequence, the likelihoods and fitted values for corresponding ZINB and NBLH models are identical for saturated models, and typically very close for non-saturated models.

On comparing the AICs of the six fitted models (Table 1, final row), we find the AIC of the corresponding zero-inflated models and hurdle logit models to be very close. The lowest AIC is observed with the ZINB- and NBLH models. These findings are further corroborated with the predicted frequencies from each of these models shown in Figure 1 (hurdle not shown as almost identical to zero-inflated): clearly the simple Poisson model cannot account for the large percentage of zero observations; the zero-inflated Poisson

Table 1. Estimated parameters (with standard error), log-likelihood value and AIC for Poisson model (P), negative binomial model (NB), zero-inflated Poisson model (ZIP), zero-inflated negative binomial model (ZINB), Poisson logit hurdle model (PLH) and negative binomial logit hurdle model (NBLH), with education level and anxious attachment level as predictors.

	P	NB	ZIP	ZINB	PLH	NBLH
COUNT COMPONENT						
(intercept)	0.817 (0.044)	0.855 (0.155)	1.921 (0.044)	1.723 (0.150)	1.921 (0.044)	1.725 (0.148)
education	-0.216 (0.070)	-0.353 (0.250)	-0.350 (0.071)	-0.490 (0.206)	-0.350 (0.071)	-0.487 (0.206)
anxiety	0.422 (0.033)	0.486 (0.122)	0.133 (0.034)	0.205 (0.108)	0.133 (0.034)	0.207 (0.107)
ZERO COMPONENT						
(intercept)			0.673 (0.142)	0.340 (0.210)	-0.675 (0.142)	-0.675 (0.142)
education			-0.232 (0.222)	-0.459 (0.297)	0.220 (0.221)	0.220 (0.221)
anxiety			-0.483 (0.111)	-0.520 (0.147)	0.486 (0.111)	0.486 (0.111)
# parameters	3	4	6	7	6	7
log <i>L</i>	-1388.20	-638.96	-802.5	-626.1	-802.5	-626.3
AIC	2782.4	1285.9	1616.9	1266.3	1616.9	1266.5

model can address this lack of fit for the zero observations, but fails to capture the non-zero frequencies correctly; while the zero-inflated negative binomial distribution fits the data best.

5. Modelling and interpreting main effects

In this section we start with a very simple model exploring the effect of education level on the number of UPB perpetrations and fit a ZINB and an NBLH model. As we do not have any *a priori* idea about which component (the zero component or the count component) might be affected by this predictor, we assume an effect on both parts.

The most important output from the fitted ZINB model is as follows:

Count model coefficients (negbin with log link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.7593	0.1525	11.534	<2e-16***
EDUCATION High	-0.4183	0.2046	-2.044	0.0409*
Log(theta)	-0.2470	0.2789	-0.886	0.3758

Zero-inflation model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3237	0.2070	1.564	0.118
EDUCATION High	-0.4647	0.2826	-1.644	0.100

For the fitted NLBH model we obtain the following:

Count model coefficients (truncated negbin with log link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.7593	0.1525	11.534	<2e-16***
EDUCATION High	-0.4183	0.2046	-2.044	0.0409*
Log(theta)	-0.2470	0.2789	-0.886	0.3758

Zero hurdle model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.6614	0.1377	-4.804	1.55e-06***
EDUCATION High	0.2614	0.2153	1.214	0.225

By default, the output shows estimated coefficients, standard errors, values for the Wald test and associated p -values, but no confidence intervals. Several observations can be made here. First, both models yield identical results for the count part (Table 2, columns 2 and 5). This observation naturally follows from the equality of count parameters for the zero-inflated and hurdle models derived in the previous section. Second, in the zero parts not only are the estimated parameters different in magnitude (with expressions (13) and (14) linking the estimated zero-component parameters from both models), but also their signs seem to have reversed. The difference in signs does not come as a surprise either as the hurdle() function in pscl is modelling the probability of a non-zero count instead of the probability of a zero count. More importantly, the interpretation for the zero-count parameters differs between zero-inflated and hurdle models, but is often neglected in the literature (e.g., Atkins & Gallop, 2007; Ravert, Schwartz, Zamboanga, Kim, Weisskirch & Bersamin, 2009; Lewis et al., 2010). From the zero-inflated model we can derive that the estimated odds of observing an *excess zero* in highly educated people is $\exp(-0.46) = 0.61$ times (with 95% confidence interval from 0.45 to 1.35) the odds in less educated people (a marginally significant effect, $p = .10$). This may lead to the erroneous interpretation that the odds of observing no UPB perpetration is (marginally) significantly smaller in highly than in less educated people. The latter odds ratio can only be derived from the hurdle model, which clearly separates the zeros and non-zeros, equals $\exp(-0.26) = 0.77$ (with 95% confidence interval from 0.51 to 1.17, $p = .23$) and can be directly linked to the observed probabilities of showing no UPB perpetration by education level, presented in the left-hand panel of Figure 2. In contrast to the zero component from the hurdle model, it is more difficult to give an intuitive interpretation to the notion of 'excess' zeros in the zero-inflated model. Following Karazsia and van Dulmen (2008), we could say that the odds of membership in the 'always zero' group (i.e., the group of ex-partners who would never show any UPB perpetration) is estimated to be 0.61 lower in highly educated people than in less educated people. As this latent group of ex-partners who would never show such behaviour is not identifiable, this statement cannot be verified directly on the data. While this example already illustrates well the potential confusion in interpreting the parameters from the zero part of the zero-inflated models, an even more extreme artificial example can be found on Wiley Online Library, revealing a highly significant effect of a binary predictor on the excess zeros but no effect on all zeros. If statements about the latter provide a more insightful answer to the research question of interest as in our example, the hurdle model should be preferred above the zero-inflated model.

Table 2. Estimated parameters (with standard error), log-likelihood value and AIC for zero-inflated negative binomial model (ZINB) and negative binomial logit hurdle model (NBLH) with (i) education level as main effect, (ii) education level and anxious attachment level as main effects, and (iii) education level and anxious attachment level as main effects and their interaction.

	ZINB			NBLH		
	education	education + anxiety	education + anxiety + interaction	education	education + anxiety	education + anxiety + interaction
COUNT COMPONENT						
(intercept)	1.759 (0.153)	1.723 (0.150)	1.782 (0.143)	1.759 (0.153)	1.725 (0.148)	1.780 (0.144)
education	-0.418 (0.205)	-0.490 (0.206)	-0.747 (0.221)	-0.418 (0.205)	-0.487 (0.206)	-0.719 (0.217)
anxiety		0.205 (0.108)	0.055 (0.118)		0.207 (0.107)	0.050 (0.120)
educ*anx			0.639 (0.254)			0.583 (0.238)
log(shape)	-0.247 (0.279)	-0.198 (0.275)	-0.096 (0.261)	-0.247 (0.279)	-0.187 (0.273)	-0.104 (0.264)
ZERO COMPONENT						
(intercept)	0.324 (0.207)	0.340 (0.210)	0.389 (0.194)	-0.661 (0.138)	-0.675 (0.142)	-0.676 (0.142)
education	-0.465 (0.283)	-0.459 (0.297)	-0.592 (0.328)	0.261 (0.215)	0.220 (0.221)	0.222 (0.223)
anxiety		-0.520 (0.147)	-0.528 (0.164)		0.486 (0.111)	0.491 (0.142)
educ*anx			0.308 (0.345)			-0.012 (0.227)
# parameters	5	7	9	5	7	9
log <i>L</i>	-638.2	-626.1	-622.9	-638.2	-626.3	-623.2
AIC	1286.5	1266.3	1263.9	1286.5	1266.5	1264.4

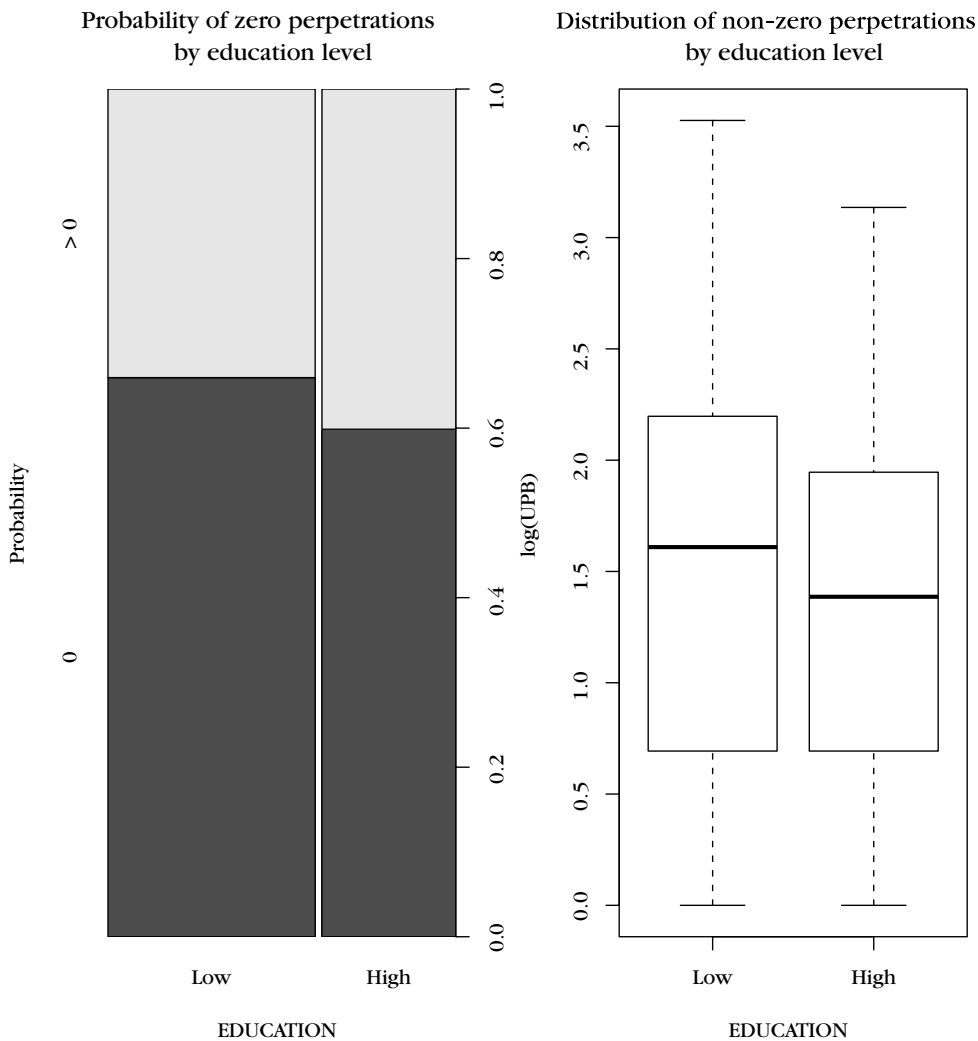


Figure 2. The effect of education level on the probability on no UPB perpetrations (left) and on the number of UPB perpetrations given that at least one perpetration is displayed (right).

Next, we interpret the parameters from the count part. In both the zero-inflated and hurdle model, we find that $\mu = \exp(1.76) = 5.81$ in the less educated group and $\exp(1.76 - 0.42) = 3.86$ in the highly educated group, which may lead to the erroneous interpretation that higher education leads to a $1 - \exp(-0.42) = 34\%$ reduction in the mean number of UPB perpetrations. Following the latent class interpretation in zero-inflated models (Karazsia & Van Dulmen, 2008), the latter interpretation is only correct in the ‘not always zero’ group (i.e., the ex-partners who are at risk of showing UPB perpetrations). Indeed, the caveat here is that the mean of *all* Y_i under the ZINB model is given by $(1 - p_i)\mu_i$ and not by μ_i . It thus follows that the expected number of UPBs is equal to $\exp(1.76)/(1 + \exp(0.32)) = 2.43$ for less educated people and $\exp(1.76 - 0.42)/(1 + \exp(0.32 - 0.46)) = 2.05$ for highly educated people, leading to only a 16% decrease.

The interpretation of the parameters γ^* in the NBLH model is not straightforward either. Indeed, the mean of *all* the Y_i under (7) and (8) is given by

$$\frac{(1 - p_i^*) \mu_i^*}{1 - \theta^\theta / (\mu_i^* + \theta)^\theta}$$

(i.e., the calculation of the mean number of UPBs under the hurdle model is further complicated due to the left censoring). The `predict()` function in R turns out to be very useful for doing the latter calculations and leads – given the equivalence between the zero-inflated and hurdle model in case of single binary predictor – to identical estimated overall means. As the hurdle model separates the zeros and non-zeros, it has the additional advantage that the estimated (functional) relationship between the predictor and the non-zero outcomes can be directly linked to the observed data (right-hand panel of Figure 2).

Finally, it is worth mentioning that the Wald test for the logarithm of the shape parameter θ being equal to zero is a test for the geometric distribution against the alternative of a negative binomial distribution, and not a test for overdispersion. Rather, an exploded value of θ would be an indication that the Poisson distribution holds (i.e., equidispersion). As noted before, an LRT (with a probability mass of 0.5 at 0 and $0.5\chi^2_{(1)}$ distribution above 0) comparing the ZIP model with the corresponding ZINB model could be used to test for overdispersion.

6. Multiple testing issue

In terms of significance testing, the question can be raised whether we can conclude from the fitted NBLH model with education as predictor in both components that education level impacts the number of UPB perpetrations or not (an argument similar to that below can be made for the fitted zero-inflated model). In fact, we have two opportunities here as we can look at the effects on both the zero and the non-zero component. If we perform both Wald tests at, for example, the 5% level (i.e., a Wald test for the impact on the zero component and a Wald test for the impact on the count component), the overall Type I error (of falsely rejecting the null hypothesis of no effect of education level on UPB) might increase up to 9.75% (if both tests are independent). One may apply a Bonferroni correction on the separate Wald tests, but such an approach is typically conservative. Considering the fitted hurdle model, we find a significant effect of education level on the count component at the 5% level without correction, but no longer after applying the Bonferroni correction. Alternatively, one can perform an LRT by comparing the likelihoods of the model with and without the predictor in the two components (Tse, Chow, Lu, & Cosmatos, 2009). This can easily be implemented in R using the `lrtest()` function from the `lmtest` package, revealing a marginally significant effect of education ($p = .064$) on the number of UPBs. If the LRT rejects the null hypothesis, one can look at the individual Wald tests for each component (such a step-down approach preserves the overall Type I error).

Yet another alternative to overcome the multiplicity issue due to a predictor being present in both components of these models is the simultaneous inference framework as proposed by Hothorn, Bretz, & Westfall (2008) in the context of general parametric models. Based on the asymptotic multivariate distribution of the estimated parameter vector and a consistent estimate of its covariance matrix, these authors show how, for

any set of linear combination of model parameters, the corresponding null hypotheses can simultaneously be tested while controlling the overall Type I error. The multcomp package in R with its `glht()` function can perform these tests, but due to the two components in this context (which contain the same predictor in our example), the contrast matrix must be constructed by the user here. An example illustrating the simultaneous testing for the effect of education in the zero and count component in the hurdle model is presented in the supplementary R-code and does not reveal a significant effect of education level on either component at the 5% level.

While the LRT approach yields a single p -value to assess whether or not a predictor significantly impacts the outcome of interest (i.e., in either component), the simultaneous inference framework presented results directly in a p -value for each component of the specific predictor separately. Simulations (results not shown) indicate that the LRT approach turns out to be slightly more powerful for testing the joint hypothesis that effects in both components are equal to zero. However, a major advantage of the approach of Hothorn *et al.* (2008) is that it can easily deal with the multiplicity issue that arises from looking at multiple predictors simultaneously.

7. Presenting and interpreting interactions in mixture models

As a final illustration, we also fitted a ZINB model with main effects for education level and anxiety, and their interaction, resulting in the following output:

Count model coefficients (negbin with log link):				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.78163	0.14328	12.434	< 2e-16***
EDUCATION High	-0.74695	0.22057	-3.386	0.000708**
ANXIETY	0.05539	0.11827	0.468	0.639555
EDUCATION High:ANXIETY	0.63949	0.25446	2.513	0.011966*
Log(theta)	-0.09611	0.26062	-0.369	0.712303
Zero-inflation model coefficients (binomial with logit link):				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3891	0.1942	2.004	0.04508*
EDUCATION High	-0.5920	0.3277	-1.807	0.07084
ANXIETY	-0.5282	0.1639	-3.222	0.00127**
EDUCATION High:ANXIETY	0.3083	0.3452	0.893	0.37186

The LRT reveals a significant interaction ($p = .041$) between education level and anxiety on the number of UPB perpetrations (Table 2, twice the difference in log L in columns 3 and 4), but its interpretation is more tedious. The left-hand panel of Figure 3 shows the effect on the count component in ‘the not always zero group’: while the effect of anxiety is relatively small for the less educated, a large increasing effect of anxiety is observed for the highly educated. Although not statistically significant, the middle panel of Figure 3 shows a large difference in the probability of not belonging to the ‘always zero’ group between high and low education level at low anxiety level, but almost no difference at high anxiety levels. While the left-hand and middle panels of Figure 3 reflect the (presence or absence of) interaction effects on the count and zero component of the zero-inflated model, its practical interpretation in these latent groups is again not that useful here. Moreover, the interaction seen on the count component of the zero-inflated model does not necessarily imply an interaction on the (observed)

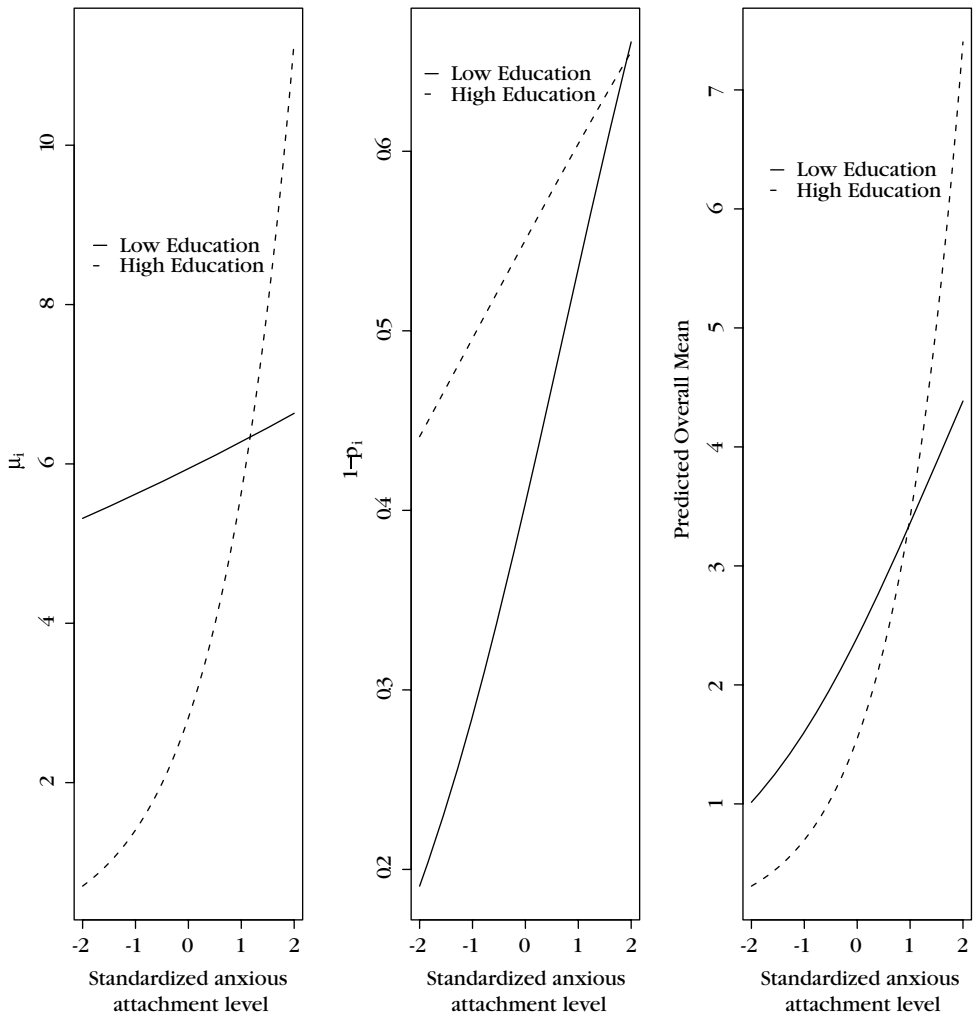


Figure 3. The effect of anxious attachment level by education level on the count and zero component and the mean of the ZINB model.

overall mean outcome. Indeed, the expected number of UPBs shown in the right-hand panel of Figure 3 (which is the product of the components shown in the left and middle panels), does not show such huge differences between education levels, but a clearly increasing number of UPBs with higher anxiety levels for both education levels. On Wiley Online Library, another more extreme artificial example is presented showing how looking at the overall mean may not reveal significant interaction effects present in one or both components of the zero-inflated (or hurdle) model.

The left-hand panel of Figure 4 shows an alternative presentation of the interaction model as suggested by Böhning, Dietz, Schlattmann, Mendonca, & Kirchner (1999) and presents the two components and the mean in a single figure at specific levels of anxious attachment (to illustrate the effect of a continuous predictor like anxiety estimated values were shown at particular values, for example, 1 standard deviation below or above the

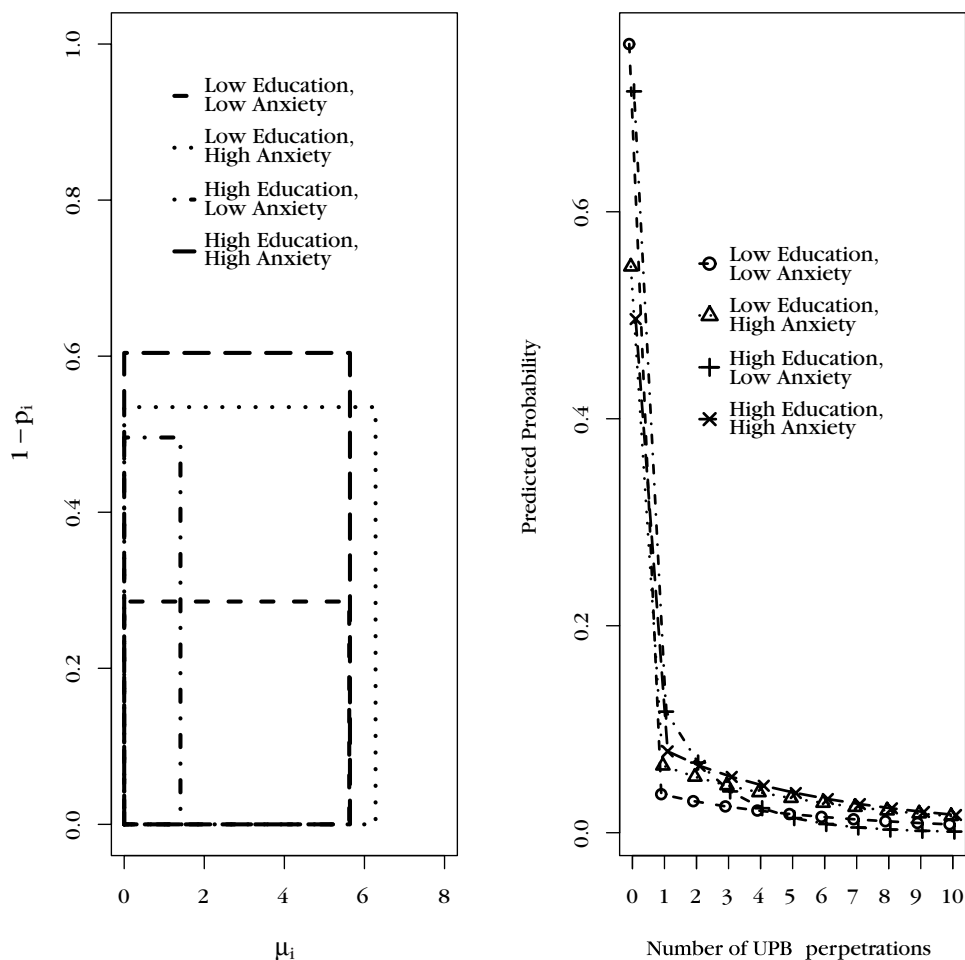


Figure 4. Alternative graphical presentations of interaction between anxious attachment level and education level of the ZINB model on the number of UPB perpetrations.

mean). The area of the rectangle formed by the estimated count component, μ_i (on the X -axis), and the estimated complement of the zero component, $1 - p_i$ (on the Y -axis), corresponds to the overall mean of Y_i . In such a presentation, interaction effects on the count and zero component of the ZIP model and on the overall mean can readily be captured by looking at the shifts on the X -axis and on the Y -axis and by comparing the areas of the rectangles, respectively. Rather than focusing on the components separately or on the mean, the right-hand panel of Figure 4 presents the model-predicted distribution of the number of UPBs at different levels of the predictor. It clearly illustrates the effect of predictors on the entire count distribution and allows one to judge which probabilities of belonging to a certain count category are impacted, and to what extent.

Finally, we also fitted a similar NBLH model with main effects for education level and anxiety, and their interaction. The output is as follows:

Count model coefficients (truncated negbin with log link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.77997	0.14426	12.338	< 2e-16 ***
EDUCATION High	-0.71927	0.2169	-3.316	0.000914***
ANXIETY	0.05029	0.11999	0.419	0.675129
EDUCATION High:ANXIETY	0.58292	0.23776	2.452	0.014217*
Log(theta)	-0.10428	0.26418	-0.395	0.693050

Zero hurdle model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.67555	0.14215	-4.752	2.01e-06***
EDUCATION High	0.22192	0.22303	0.995	0.31971
ANXIETY	0.49106	0.14192	3.460	0.00054***
EDUCATION High:ANXIETY	-0.01233	0.22749	-0.054	0.95678

While the corresponding zero-inflated and hurdle models yield almost identical fitted values, the more convenient interpretation and the direct link with the observed data of the hurdle model are an important asset for the interpretation of interactions too. The upper left-hand and right-hand panels of Figure 5 reveal a similar trend of decreasing probabilities of no UPB perpetrations with increasing anxious attachment levels in people with low and high education, and hence the absence of any interaction between education and anxiety in the zero component of the hurdle model is not surprising. On the other hand, once the hurdle is crossed and at least one perpetration is shown, the number of UPB perpetrations is increasing with anxious attachment levels in the high education group, but not in the low education group. This differential association is clearly reflected by the significant interaction in the count component of the hurdle model.

8. Conclusion

In this tutorial, we have presented in detail the distinction between zero-inflated models and hurdle models. While the latter are seldom used in psychological research, we hope in the future – having introduced both models in parallel – to limit misleading interpretations of parameters from the zero-inflated models that have frequently been reported hitherto. The choice between the hurdle and zero-inflated models should be based on the aim and endpoints of the study (Rose, Martin, Wannemuehler and Plikaythis, 2006). If the goal is prediction, it is not important which modelling framework is used, because predictions are (almost) identical. However, if the goal is inference, model choice is related to the study goal. Suppose we were to ask a random group of women “how many times did you forget to take the contraceptive pill last month?”. Some women will never have used the pill, while others will not have forgotten to take it in the last month. The zero component from the zero-inflated model may then be attributed to this latent group of non-users. Limiting the sample to participants who use the contraceptive pill, on the other hand, may more naturally lead to using the hurdle modelling framework for drawing inference.

In summary, ease of implementation in the *pscl* package in R, in addition to the more straightforward interpretation of the components and its direct link with the observed data, make the hurdle model definitely a valuable alternative for researchers analysing zero-inflated count data.

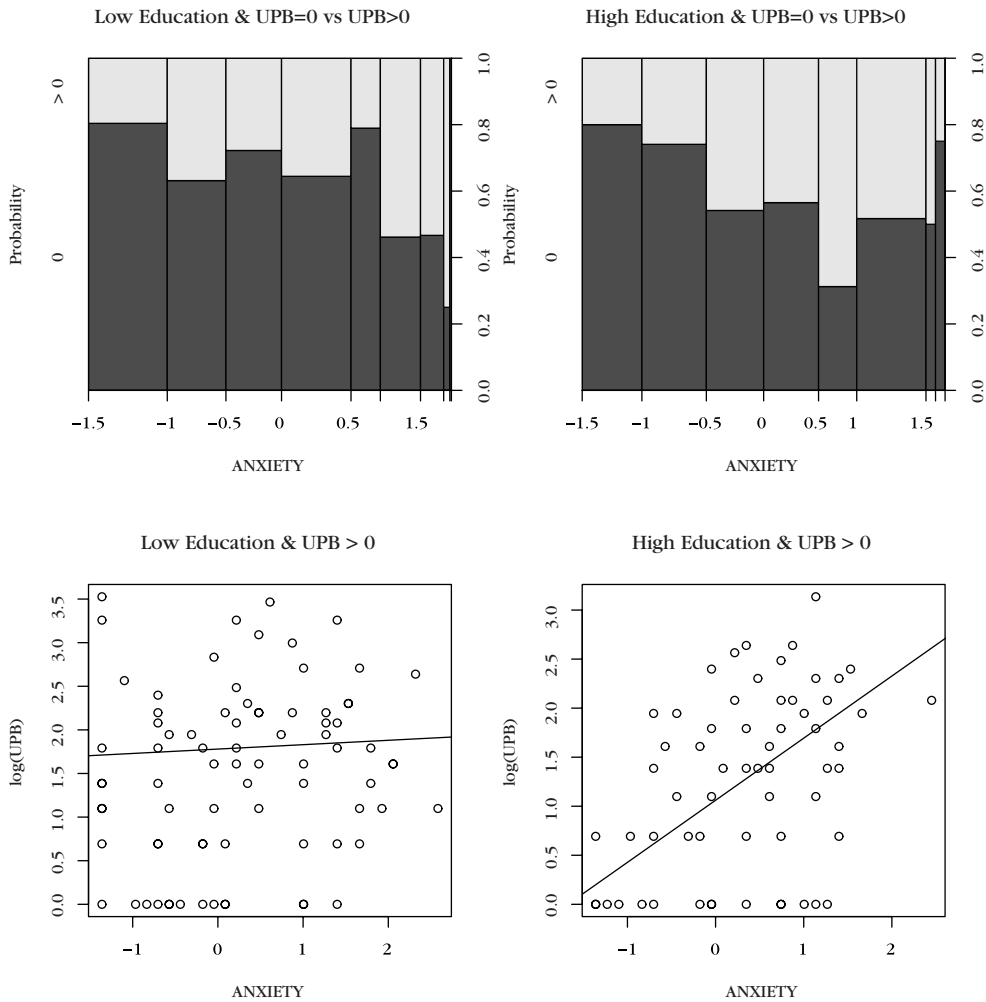


Figure 5. Direct link between the observed number of UPB perpetrations by education and anxious attachment level and the estimated effects from the NBLH model with main effects for education and anxious attachment and its interaction.

Acknowledgements

This work benefited greatly from the *pscl* package (Zeileis *et al.*, 2008) available for R. The authors would also like to thank the editor along with the anonymous reviewers for their review of the manuscript, and Isabelle Saelens for administrative support.

References

- Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.
- Atkins, D., & Gallop, R. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 21(4), 726–735.
- Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., & Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society Series A*, 162, 195–209.

- Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge: Cambridge University Press.
- Cupach, W. R., & Spitzberg, B. H. (2004). *The dark side of relationship pursuit. From attraction to obsession and stalking*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Dalrymple, M. L., Hudson, I. L., & Ford, R. P. K. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics & Data Analysis*, 41, 491–504.
- De Smet, O., Buysse, A., & Brondeel, R. (2011). Effect of the breakup context on unwanted pursuit behavior perpetration between former partners. *Journal of Forensic Sciences*, 56(4), 934–941.
- Hilbe, J. (2011). *Negative binomial regression*. Cambridge: Cambridge University Press.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 346–363.
- Karazsia, B. T., & van Dulmen, M. (2008). Regression models for count data: Illustrations using longitudinal predictors of childhood injury. *Journal of Pediatric Psychology*, 33, 1074–1084.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–17.
- Lewis, M. A., Clayton, N., Geisner, I. M., Lee, C. M., Kilmer, J. R., & Atkins, D. (2010). Examining the associations among severity of injunctive drinking, norms, alcohol consumption and alcohol-related negative consequences: The moderating roles of alcohol consumption and identity. *Psychology of Addictive Behaviors*, 24(2), 177–189.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341–365.
- Ravert, R. D., Schwartz, S., Zamboanga, B. L., Kim, S. Y., Weisskirch, R. S. & Bersamin, M. (2009). Sensation seeking and danger invulnerability: Paths to college student risk-taking. *Personality and Individual Differences*, 47, 763–768.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modelling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics*, 16, 463–481.
- Tse, S. K., Chow, S. C., Lu, Q., & Cosmatos, D. (2009) Testing homogeneity of two zero-inflated Poisson populations. *Biometrical Journal*, 1, 159–170.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
- Vives, J., Losilla, J. M., & Rodrigo, M. F. (2006). Count data in psychological applied research. *Psychological Reports*, 98(3), 821–835.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.
- Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The experiences in close relationship scale (ECR)-short form: Reliability, validity, and factor structure. *Journal of Personality Assessment*, 88, 187–204.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27(8), 1–25.

Received 18 October 2010; revised version received 5 July 2011

Supporting Material

The following supporting information may be found in the online edition of this article:

analysis UPB.R: Accompanying R-code

couple.txt: Illustrating Data

ARTIFICIAL EXAMPLE1.PDF: Artificial Example 1 - Interpretation

ARTIFICIAL EXAMPLE2.PDF: Artificial Example 2 - Interpretation

artificial example1.R: Artificial Example 1 - R-code

artificial example2.R: Artificial Example 2 - R-code

The .R files can be opened as .txt files for example with NOTEPAD.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.