International Journal of Mathematics and Statistics

ISSN (Online): 2208-2212

Volume 09 Issue 03 September 2023

DOI: https://doi.org/10.69980/kwpx3275

MODELING OVERDISPERSED COUNT DATA: A COMPARATIVE STATISTICAL ANALYSIS OF PERCEIVED CARNIVORE ABUNDANCE

Dr. Elena M. Hartwell[1]*, Prof. Rafael J. Mendoza[2], Dr. Hiroshi K. Tanaka[3]

[1]Department of Ecological Statistics and Wildlife Analytics, Northbridge Institute of Environmental Research, Edinburgh, United Kingdom*

[2]School of Wildlife Ecology and Conservation Science, Universidad de Sierra Verde, Mérida, Mexico

[3]Center for Applied Biostatistics and Ecological Modeling, Pacific Institute of Science and Technology, Sapporo, Japan

Abstract:

In ecological and perception-based research, counts often do not follow the Poisson distribution of equidispersion, which may impact the fit of the model and statistical inference. This study aimed to characterize dispersion patterns in perceived bear, leopard, and wolf abundance, estimate associated predictors, and compare Poisson and Negative Binomial regression performance. A common predictor structure was used to analyze the species-specific count outcomes. Descriptive statistics, dispersion indices, and Poisson and Negative Binomial regression were analyzed. Model performance was assessed by AIC, BIC, Pearson dispersion, and residual diagnostics, and adjusted effects were presented as incidence rate ratios (IRRs) with 95% confidence intervals (CIs). There was strong overdispersion for all species, with dispersion indices of 268.949 for bear, 269.515 for leopard, and 529.372 for wolf. The fit of the models was significantly enhanced using Negative Binomial regression, with the Pearson dispersion values of 1.585, 2.016, and 0.960 obtained for the three models, respectively. The perceived bear abundance was positively correlated with danger score. For leopards, age and family kill-belief were significantly positively correlated, and happiness/pride was significantly negatively correlated, while none of these factors was statistically significant for wolves. Negative Binomial regression was a better fit for the observed count structure, highlighting the need to consider the dispersion of counts and species-specific inference when modeling heterogeneous count outcomes.

Keywords: *overdispersion, count regression, Negative Binomial regression, perceived abundance, carnivore modeling*

1. Introduction:

Count data is encountered in applied research in all contexts where the outcome is a count of events (occurrences), individuals, or reported observations within a specified context. Counts are non-negative integers and are often skewed, have unequal variability, and have concentrations at specific values, unlike continuous responses. The distributional properties demand a statistical model that does not make the usual Gaussian assumptions but treats the response as discrete. Poisson regression is one of the basic models for the relationship between a count response and predictor variables in terms of the conditional mean and the log link function (Tang et al., 2023). Categorical and count-data methods are therefore of particular value as a link between the probability models and practical applications where both response and predictor are categorical, and where the response is count-based (Chen & Anderson, 2024). One of the main restrictions of the Poisson model is that of equidispersion, that is, the conditional variance of the count response equals the conditional mean response. The assumption is often not true in empirical observations due to unobserved heterogeneity, clustering, differences in exposure or processes of underlying events between sampling units. When the variance is greater than the mean, the model is said to have overdispersion and can significantly influence the goodness of fit of a Poisson specification. To overcome the limitation of the standard Poisson variance structure, generalized count models have been developed to suit situations where this variance structure is too restrictive (Saputro et al., 2021). One of the key reasons for examining the mean–variance relationship is that extreme variation may signify the presence of fundamental problems in the Poisson distribution for the actual count data (Bektashi et al., 2022).

The extent of overdispersion also impacts model selection and statistical inference. Shrinking or excess variability not explained can lead to inappropriate estimates of uncertainty and to weaken confidence in inferential statements that may be made from a Poisson model. The empirical assessment is thus required before deciding if a more flexible distribution of counts is needed (Payne et al., 2018). An often-used alternative is Negative Binomial regression, which adds another dispersion parameter that allows the conditional variance to be greater than the conditional mean. This formulation keeps the same multiplicative mean structure and interpretability of the count regression, while providing much more variation in the response (Cummings, 2019). A lot of attention has been paid to the statistical estimation under the Negative Binomial regression, as the estimation of the regression parameters is very crucial when counts exhibit heterogeneity not accounted for by the Poisson regression. When using a Negative Binomial variance structure, improvements in parameter estimation can increase confidence in fitted relationships (Reangsephet et al., 2018). Tailor choice of model to the dispersion, goodness-of-fit, diagnostic evidence, and not only because the response is an integer. This is particularly important if several related outcomes are considered, and if the variation in the outcomes can vary.

The statistical principles described above are useful when applied to the abundance of carnivores since abundance-related counts may fluctuate significantly among species, locations and observation contexts. Estimates of carnivore community structure have demonstrated the importance of quantitative methods that are able to separate variation in a community among species and that take into account the nature of ecological observations (Jiménez et al., 2017). Additionally, abundance can vary greatly across space and time for carnivores, leading to highly variable patterns which may not, and often are not, represented by a simple abundance model (Broekhuis et al., 2021). In addition to the ecological diversity, there is also a perceived abundance diversity in that the reported counts depend not just on the ecological conditions, but also on the experience, demographic, belief, and perception differences among the respondents. Such variation should be considered in the context of the environment. Species–habitat relationships can be used as a foundation to better understand how spatial and environmental features could be related to the observed ecological responses and their statistical representation (Matthiopoulos et al., 2023). For ecological data analysis, models that link observed variation to explanatory variables while at the same time being coherent with the response variable's probability structure are needed (Clark, 2020). Model evaluation is also critical, as ecological data often have several sources of variation and inferences may vary across models that make different assumptions and/or specifications about that variation (Harrison et al., 2018). Although there has been much methodological work on count regression, relatively little work has been devoted to the key statistical problem of overdispersion when modeling species-specific abundance of carnivores under a shared predictor structure. Although studies on carnivore abundance and ecological variation have been conducted that provide substantial bases, the abundance data also have well-established alternatives to Poisson regression, and count-data research has well-established alternatives to Poisson regression that offer a route for a direct comparative assessment that combines distributional characteristics with dispersion measurement, information-criterion-based model comparison, and residual diagnostics. One of the advantages of a cross-species approach is that the degree of overdispersion and the modified relationship between perceived abundance and environmental, demographic, socioeconomic and perception-related variables do not have to be consistent among carnivores.

Therefore, the present study seeks to describe the distribution and dispersion of perceived bear, leopard, and wolf abundance, and to test the different statistical models (Poisson and Negative Binomial) using count-regression models to determine which model best describes each carnivore species in predicting the effect of environmental, demographic, socioeconomic, and perception-related predictors. In this framework, the mean–variance structure of the count response is central, and it allows for a common comparison of the effect of predictors and model performance between species.

2. Materials and Methods

2.1 Research Design

A quantitative applied-statistical design was employed to model the perceived abundance as an overdispersed count response, and to compare count-regression alternatives. Data for bear, leopard and wolf were analyzed separately, while employing the same modeling framework to ensure consistency across species. These analytical steps included data screening, descriptive and dispersion analysis, predictor specification, Poisson regression analysis, Negative Binomial

regression analysis, model comparison and residual diagnostics. The primary statistical goal was to assess if the explicit modeling of extra-Poisson variation is helpful for improving fit and inference for the three abundance outcomes.

2.2 Data Source

Data were from Ebrahimi Monfared et al. (2025) and included bear, leopard and wolf abundance data from the respondent and site-level data. Demographic and socioeconomic data, habitat and elevation data, livestock-loss experience, and levels of perceptions and attitudes toward carnivores were included in the records. Separate species-specific files were used for the comparative statistical analysis.

2.3 Data Preparation and Predictor Specification

All species data were checked for blank fields, duplicates, missing data and variable structure. The fields that were identified and the completely blank columns were removed. Final samples of 154 bear, 202 leopard and 168 wolf observations were obtained using complete case analysis. An abundance response of non-negative integers was described. A set of common predictor variables was used, including habitat, gender, elevation, age, literacy, income, livestock-loss experience, fear, perceived danger, family kill-belief, and happiness/pride, so that the fitted coefficients could be compared across species.

2.4 Variable Transformation and Multicollinearity Assessment

Predictor distributions and category frequencies were examined before model estimation. Household income was natural log transformed due to high positive skewness. Age and elevation were rescaled to 10-year and 100-m units, respectively, to allow for the interpretation of regression coefficients without changing the relationships between these and the response. Ordinal perception measures were coded as numbers. Variance inflation factors were used to screen the candidate predictors for multicollinearity. Other variables were removed due to redundancy with environmental variables: ethnicity, and due to very few or strongly imbalanced categories: occupation, and conservation attitude.

2.5 Distribution and Dispersion Analysis

Summary of species-specific abundance distributions was summarized by use of the mean, standard deviation, median, minimum, maximum, variance and frequency of zero counts. The dispersion index was used to quantify the extra-Poisson variation,

$$D=\frac{s^2}{\bar{y}},$$

where s^2 denotes the sample variance and $\bar{y}$ the sample mean. Numbers much higher than one meant that the variance is larger than the mean. The abundance was presented graphically using a log-transformed count scale because the abundance data were very right-skewed, but regression models were fitted to the original count abundance data

2.6 Count Regression Models

For each species, a Poisson regression with a log link was fitted first. For observation i, the conditional mean was modeled as

$$\log(\mu_i)=\beta_0+\sum_{j=1}^{p}\beta_j X_{ij},$$

where μ_i represents expected perceived abundance and X_{ij} represents the specified predictors. Pearson dispersion was then explored to evaluate the Poisson mean–variance assumption. A variance structure was used to fit Negative Binomial regression to account for excess variation.

$$\mathrm{Var}(Y_i)=\mu_i+\alpha\mu_i^2,$$

where α is the estimated dispersion parameter. Incidence rate ratios and 95% confidence intervals were calculated by exponentiation of model coefficients, and p-values were considered statistically significant at $p < 0.05$.

2.7 Model Comparison and Diagnostic Assessment

Log-likelihood, Akaike Information Criterion, Bayesian Information Criterion and Pearson dispersion were used to compare the Poisson and Negative Binomial specifications. These values were lower for the comparatively better fit, and remaining variance misspecification was evaluated using Pearson dispersion. Selected models were fitted, and the following diagnostics were used: Pearson residual standard deviation, proportions of absolute Pearson residuals greater than 2 and 3, quantile residual summaries, and Spearman correlations between the residuals and the fitted values. Additionally, the observed and fitted mean counts were compared between species.

3. Results

3.1 Distribution of Perceived Carnivore Abundance

There were 154 observations for bear, 202 observations for leopard, and 168 observations for wolf. The perceived abundance varies greatly between the three species and is evident in Table 1. Mean abundance was 117.773 for bear, 87.525 for leopard, and 352.506 for wolf, while the corresponding medians were 50, 20, and 175. Wolf had the largest variability (SD = 431.980), followed by bear (SD = 177.974) and leopard (SD = 153.588). Variance was found to be significantly above the mean for all species, and the dispersion index ranged from 268.949 to 529.372. Only 7.143% of the wolf observations had zero, and less than that for bear and leopard.

Table 1. Distributional Characteristics of Perceived Carnivore Abundance

Species	N	Mean	SD	Median	Minimum	Maximum	Variance	Dispersion Index	Zero n	Zero (%)
Bear	154	117.773	177.974	50.0	0	1,000	31,674.870	268.949	1	0.649
Leopard	202	87.525	153.588	20.0	0	800	23,589.256	269.515	3	1.485
Wolf	168	352.506	431.980	175.0	0	2,000	186,606.695	529.372	12	7.143

The distributions of perceived abundance (log10(count + 1)) are presented in Figure 1. The median (50.0) is below the mean (117.8), as shown in Figure 1A. The number of leopards is shown in Figure 1B, with 20.0 as the median and 87.5 as the mean. Wolf numbers have the highest median (175.0) and mean (352.5) of the three species, as shown in Figure 1C. The median is shown as a dashed line and the mean as a dotted line.

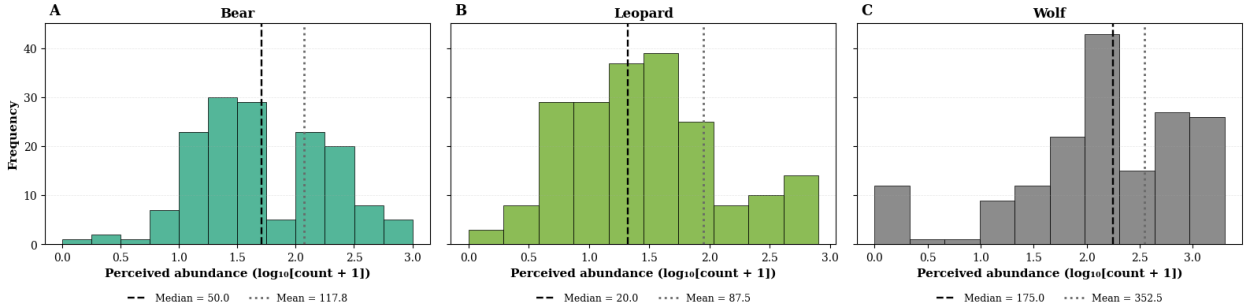

Figure 1. Distribution of Perceived Carnivore Abundance. (A) Bear abundance. (B) Leopard abundance. (C) Wolf abundance.

3.2 Poisson and Negative Binomial Model Comparison

There are significant differences with regard to model fit between Poisson and Negative Binomial regression (Table 2). For bear, AIC decreased from 21,669.244 to 1,739.288 and BIC from 21,705.687 to 1,778.768. For leopard, AIC decreased from 25,589.937 to 2,107.561 and BIC from 25,629.637 to 2,150.568. Wolf produced the largest reductions, with AIC decreasing from 67,023.759 to 2,263.475 and BIC from 67,061.247 to 2,304.086. The values of Poisson-Pearson dispersion were 216.679, 222.158 and 525.999 for bear, leopard and wolf, respectively. These values were lowered to 1.585, 2.016 and 0.960 in a model of Negative Binomial regression. The Negative Binomial was thus chosen as the model for all three species.

Table 2. Comparison of Poisson and Negative Binomial Regression Models

Species	Poisson AIC	NB AIC	ΔAIC	Poisson BIC	NB BIC	ΔBIC	α	Poisson Pearson Dispersion	NB Pearson Dispersion	Preferred Model
Bear	21,669.244	1,739.288	19,929.956	21,705.687	1,778.768	19,926.919	1.174	216.679	1.585	Negative Binomial
Leopard	25,589.937	2,107.561	23,482.377	25,629.637	2,150.568	23,479.069	1.477	222.158	2.016	Negative Binomial
Wolf	67,023.759	2,263.475	64,760.285	67,061.247	2,304.086	64,757.161	1.786	525.999	0.960	Negative Binomial

Model changes are illustrated in Figure 2 by the change in information criteria and Pearson dispersion between the two model specifications. Figure 2A shows ΔAIC/ΔBIC values of approximately 19,930/19,927 for bear, 23,482/23,479 for leopard, and 64,760/64,757 for wolf. Figure 2B shows the reduction in Pearson dispersion from 216.68 to 1.59 for bear, 222.16 to 2.02 for leopard, and 526.00 to 0.96 for wolf.

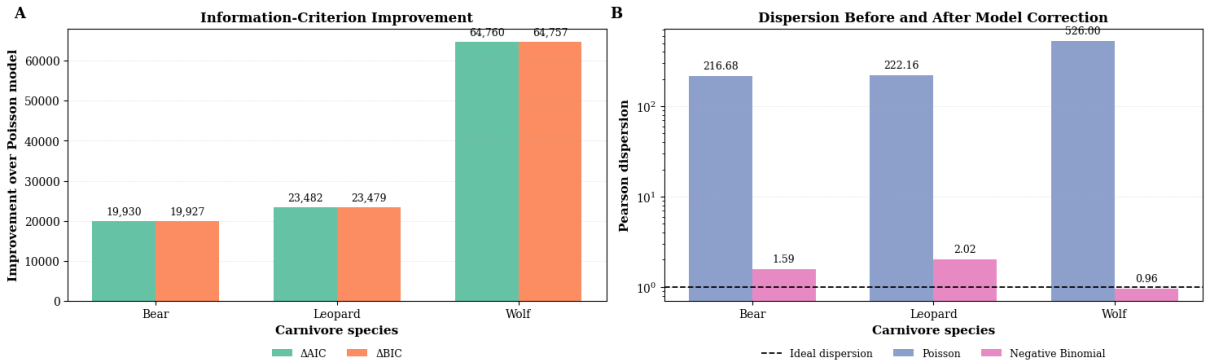

Figure 2. Comparison of Poisson and Negative Binomial Model Performance. (A) Information-criterion improvement. (B) Pearson dispersion before and after model correction.

3.3 Negative Binomial Model Diagnostics

The observed mean abundances of bear, leopard, and wolf were 117.773, 87.525, and 352.506, respectively, and the predicted abundances were 119.563, 90.845, and 356.551, respectively, as indicated in Table 3 (Table 3: Residual statistics of the selected models). The SDs of the Pearson Residuals were 1.213 for bear, 1.380 for leopard and 0.947 for wolf. The absolute Pearson residuals were >2 for 7.143%, 5.941% and 4.167% of the observations, respectively. Quantile residual means ranged from −0.028 to 0.020, with SDs between 0.971 and 1.003. Residual–fitted Spearman coefficients were −0.039 for bear, −0.008 for leopard, and 0.062 for wolf, with p-values of 0.633, 0.905, and 0.426.

Table 3. Residual Diagnostics for the Negative Binomial Models

Species	Observed Mean	Predicted Mean	Pearson Residual SD	|Pearson| > 2 (%)	|Pearson| > 3 (%)	Quantile Residual Mean	Quantile Residual SD	Residual–Fitted Spearman	p-value
Bear	117.773	119.563	1.213	7.143	3.247	−0.015	0.988	−0.039	0.633
Leopard	87.525	90.845	1.380	5.941	3.465	−0.028	0.971	−0.008	0.905
Wolf	352.506	356.551	0.947	4.167	1.786	0.020	1.003	0.062	0.426

3.4 Negative Binomial Regression Estimates

Adjusted Negative Binomial regression estimates are shown in Table 4. In the bear model, danger score had an IRR of 1.347 (95% CI: 1.177–1.541, p < 0.001). All other bear predictors had a p-value greater than 0.05. In the leopard model, age per 10 years had an IRR of 1.195 (95% CI: 1.019–1.401, p = 0.028), family kill-belief score had an IRR of 1.193 (95% CI: 1.049–1.355, p = 0.007), and happiness/pride score had an IRR of 0.728 (95% CI: 0.609–0.871, p < 0.001). None of the examined predictors in the wolf model reached statistical significance at p < 0.05 .

Table 4. Negative Binomial Regression Estimates for Perceived Carnivore Abundance

Species	Predictor	Coefficient	SE	IRR (95% CI)	p-value
Bear	Intercept	1.788	2.136	5.979 (0.091–393.689)	0.403
Bear	Habitat: Steppe vs Forest	0.110	0.466	1.117 (0.448–2.784)	0.813
Bear	Gender: Male vs Female	0.465	0.301	1.593 (0.883–2.872)	0.122
Bear	Elevation (per 100 m)	−0.105	0.060	0.900 (0.800–1.013)	0.080
Bear	Age (per 10 years)	0.122	0.077	1.130 (0.971–1.315)	0.114
Bear	Literacy score	0.180	0.128	1.197 (0.931–1.538)	0.160
Bear	Log-transformed income	0.065	0.105	1.067 (0.868–1.313)	0.536
Bear	Loss experience	−0.298	0.263	0.742 (0.443–1.243)	0.257
Bear	Fear score	0.124	0.064	1.132 (0.999–1.283)	0.052
Bear	Danger score	0.298	0.069	1.347 (1.177–1.541)	<0.001
Bear	Family kill-belief score	0.111	0.058	1.117 (0.996–1.253)	0.058
Bear	Happiness/pride score	−0.137	0.070	0.872 (0.760–1.001)	0.052

Leopard	Intercept	5.355	2.262	211.687 (2.513–17,833.509)	0.018
Leopard	Habitat: Steppe vs Forest	0.185	0.421	1.204 (0.528–2.745)	0.660
Leopard	Gender: Male vs Female	0.578	0.328	1.782 (0.938–3.386)	0.078
Leopard	Elevation (per 100 m)	−0.014	0.047	0.986 (0.899–1.082)	0.773
Leopard	Age (per 10 years)	0.178	0.081	1.195 (1.019–1.401)	0.028
Leopard	Literacy score	0.144	0.124	1.155 (0.906–1.472)	0.243
Leopard	Log-transformed income	−0.117	0.117	0.890 (0.707–1.118)	0.316
Leopard	Loss experience	−0.394	0.240	0.674 (0.421–1.080)	0.101
Leopard	Fear score	0.064	0.074	1.066 (0.923–1.232)	0.385
Leopard	Danger score	0.135	0.078	1.145 (0.982–1.334)	0.083
Leopard	Family kill-belief score	0.176	0.065	1.193 (1.049–1.355)	0.007
Leopard	Happiness/pride score	−0.317	0.091	0.728 (0.609–0.871)	<0.001
Wolf	Intercept	2.043	2.866	7.714 (0.028–2,123.994)	0.476
Wolf	Habitat: Steppe vs Forest	−0.513	0.567	0.599 (0.197–1.820)	0.366
Wolf	Gender: Male vs Female	0.006	0.315	1.006 (0.543–1.864)	0.985
Wolf	Elevation (per 100 m)	0.032	0.066	1.032 (0.906–1.176)	0.631
Wolf	Age (per 10 years)	−0.127	0.099	0.881 (0.725–1.070)	0.202
Wolf	Literacy score	−0.099	0.149	0.906 (0.677–1.212)	0.507
Wolf	Log-transformed income	0.183	0.135	1.201 (0.922–1.565)	0.175
Wolf	Loss experience	−0.237	0.278	0.789 (0.458–1.359)	0.393
Wolf	Fear score	−0.002	0.083	0.998 (0.849–1.173)	0.978
Wolf	Danger score	0.169	0.102	1.184 (0.968–1.447)	0.100
Wolf	Family kill-belief score	0.125	0.072	1.133 (0.983–1.306)	0.084
Wolf	Happiness/pride score	−0.114	0.078	0.892 (0.766–1.039)	0.143

Note: IRR = incidence rate ratio; CI = confidence interval.

The three Negative Binomial models gave adjusted IRRs with 95% confidence intervals that are presented in Figure 3. The 95% CI for bear danger score was higher than IRR = 1 (Figure 3A), and the intervals for all other bear predictors overlapped 1. Confidence intervals above 1 for age of leopard and family kill-belief score, and below 1 for happiness/pride score are seen in Figure 3B. The confidence intervals for all the wolf predictors overlapped IRR = 1 (Figure 3C). Points represent adjusted incidence rate ratios (IRRs) and horizontal bars represent 95% confidence intervals (CI); the dashed vertical line represents IRR = 1.

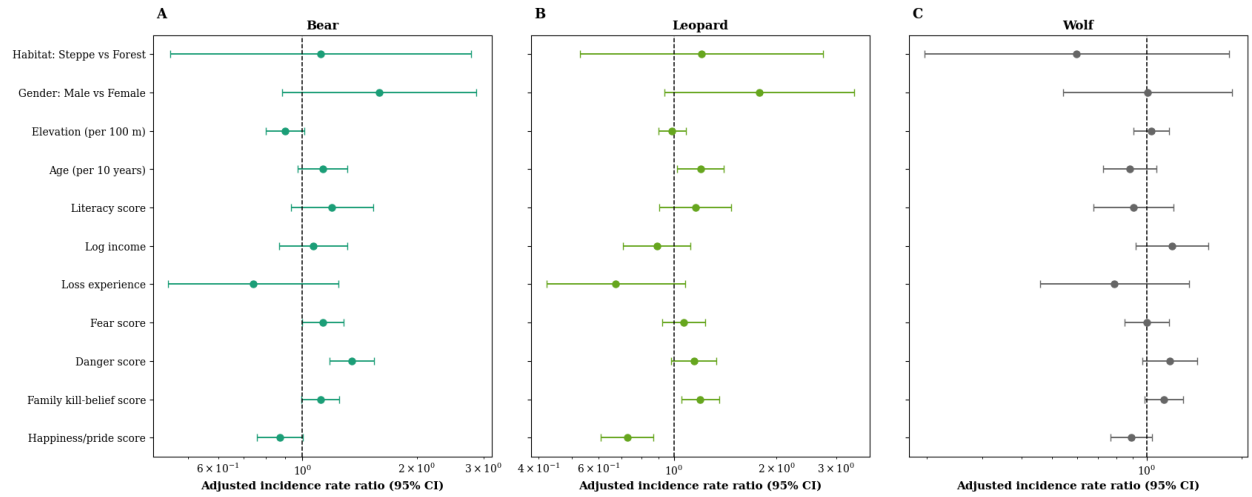

Figure 3. Adjusted Predictor Effects on Perceived Carnivore Abundance. (A) Bear model. (B) Leopard model. (C) Wolf model.

4. Discussion

The results show significant deviations from equidispersion at each of the three perceived carnivore abundance outcomes. The dispersion indices for bear, leopard, and wolf were 268.949, 269.515, and 529.372, respectively, indicating substantial variance relative to the corresponding mean counts. There were also large differences in abundance means and medians, which was another reflection of the skewed structure of the counts, in this case with wolves. Consistent with these distributional patterns, the Poisson models also produced very high Pearson dispersion values across all three species. A Poisson mean–variance structure was therefore not a good representation of the count-generating variability. The statistical fit for each species was significantly improved using Negative Binomial regression. Negative Binomial regression substantially improved model fit for all three species, as indicated by large reductions in AIC and BIC. Pearson dispersion decreased from 216.679 to 1.585 for bear, 222.158 to 2.016 for leopard, and 525.999 to 0.960 for wolf. The estimated dispersion parameters were also positive in all the models. The selected specification was supported by the residual diagnostics: observed and predicted mean counts were similar; quantile residual means were close to zero and residual–fitted correlations were weak and nonsignificant. The leopard dispersion was slightly larger than the other dispersions (bear and wolf) but was significantly smaller than the Poisson dispersion. There was also a clear difference in species-specificity of the adjusted models. The only statistically significant predictor of perceived bear abundance was danger score, which had an IRR of 1.347, indicating an increase in expected bear abundance of 34.7% for each increase of one unit in danger score while other variables were held constant. Age, family kill-belief, and happiness/pride were significant factors in leopard abundance. A 10-year increase in age was associated with a 19.5% increase in expected perceived leopard abundance, and a unit increase in family kill-belief was associated with an increase in the expected count of 19.3%. The abundance of leopards was inversely related to happiness/pride with an IRR of 0.728. Although wolves were the most abundant and most variable of the species examined, none of the examined predictors were statistically significant.

The species differences seen here are consistent with previous work showing that interspecific differences in abundance and density of carnivores within the same ecological community may be significant. Forsyth et al. (2019) showed that the impact of interspecific structure, abundance and density on carnivore communities needs to be considered when assessing them. Patterns of abundance may also be influenced by environmental heterogeneity, with habitat features demonstrating the relationship between carrying capacity and population recovery of carnivores (Tinker et al., 2021). These results can be used to give a more general sense of the unique distributions in count data that were found for bear, leopard, and wolf. When carnivore abundance is estimated by heterogeneous observations, reliable statistical specification is also important. Elliot and Gopalaswamy (2017) noted the need for proper and precise analytical tools in estimating carnivores' density. In line with comparative evidence that suggests that some count data are overdispersed and therefore need methods to capture the overlying variability, which extends beyond the Poisson structure (Kruppa & Hothorn, 2021), the strong preference for Negative Binomial analysis in the present analysis is also statistically warranted.

When fitting more complex models, it is important to assess the distribution of ecological count data, as it may have excess zeros and overdispersion. Sidumo et al. (2024) showed that the methods of count regression specifically tailored for zero-inflated and overdispersed ecological outcomes are valuable. The abundance data contained relatively few zeros, particularly for bear and leopard, indicating that overdispersion rather than excess zero frequency was the dominant distributional feature. This distinction is relevant because, when the observed count process is not a good fit, the model selection and inference may change as a result of assuming unnecessary zero inflation (Campbell, 2021). If the ecological count structure is more complex, then models that allow for dispersion, autocorrelation and zero inflation may be required, as can be seen in Carvalho et al. (2020). These current results did not suggest any further complexity in the main analysis. Rather, the big enhancement achieved by explicitly modelling the extra-Poisson variation is in line with evidence that the presence of large amounts of unexplained count variation should be included in the model structure (Payne et al., 2017). Similarly, when count responses have a hierarchical structure or are zero-inflated (ZI), flexible modeling strategies are shown to be important by Li et al. (2024).

The results have implications for applied count data analysis. Ensuring that a count outcome is interpreted within a dispersion diagnosis is important because simply having a count outcome does not mean that the Poisson regression is appropriate. These large decreases in AIC, BIC and the Pearson dispersion are all evidence of the utility of matching the variance specification to the observed data structure. Negative Binomial regression also maintained simple interpretation of IRRs and provided a significant amount of heterogeneity. Further, the cross-species results indicate that similar distributional issues do not necessarily mean similar predictor relationships. There were greater associations with danger for bear perception as well as associations with demographic and belief variables for leopard perception, and more uncertainty across predictors for wolf estimation. Thus, a common statistical specification is sufficient for comparability and does not require homogeneous effects across species. This separation enhances the importance of species-specific estimation in an integrated analytical approach.

There are a few caveats. Perceived abundance represents respondent-reported counts rather than direct measurements of carnivore population abundance. Sample sizes also varied across species, and some candidate predictors were excluded because of sparse or highly imbalanced categories. There was some small residual dispersion in the leopard model, but it was significantly reduced compared to the Poisson model. The relationships presented here could be evaluated for their reproducibility in independent samples, and repeated, spatial or temporal observations could be used if available in future research. Where future data sets have more pronounced zero inflation, clustering and/or hierarchical dependence, a comparative evaluation of generalized count, hurdle and/or multilevel Negative Binomial models may be warranted. These extensions would maintain the key focus on an appropriate specification of mean–variance but also would enable the inclusion of additional sources of count heterogeneity to be explored.

5. Conclusion

Perceived abundance of bears, leopards and wolves was found to be over-dispersed, such that the Poisson specification was inappropriate for all three count outcomes. The Negative Binomial regression was a much better fit, with significant decreases in AIC, BIC, and Pearson dispersion and satisfactory residual behavior. Species-specific effects were also apparent. Bear abundance and perceived danger were positively related, and age, family kill-belief and happiness/pride were associated with leopard abundance. When considering the adjusted relationships, more uncertainty in the wolf model existed due to the lack of statistical significance for each predictor. The results highlight the need to compare the mean–variance structure before deciding on a count-regression model to use and conclude that a standard statistical method can still show different patterns of predictors for each species. The analysis indicates that Negative Binomial regression is a useful and interpretable approach to count data that is strongly overdispersed, but maintains direct comparisons via incidence rate ratios. Further research is needed to confirm these relationships in other samples, and to explore more complex count structures as they become available with repeated, spatial, or hierarchical observations.

References

1.Bektashi, X., Rexhepi, S., & Limani–Bektashi, N. (2022). Dispersion of count data: a case study of Poisson distribution and its limitations. Asian Journal of Probability and Statistics, 19(2), 18-28.

2.Broekhuis, F., Elliot, N. B., Keiwua, K., Koinet, K., Macdonald, D. W., Mogensen, N., ... & Gopalaswamy, A. M. (2021). Resource pulses influence the spatio-temporal dynamics of a large carnivore population. Ecography, 44(3), 358-369.

3.Campbell, H. (2021). The consequences of checking for zero-inflation and overdispersion in the analysis of count data. Methods in Ecology and Evolution, 12(4), 665-680.

4.Carvalho, F. J., de Santana, D. G., & Sampaio, M. V. (2020). Modeling Overdispersion, Autocorrelation, and Zero-Inflated Count Data Via Generalized Additive Models and Bayesian Statistics in an Aphid Population Study: Carvalho et al. Neotropical entomology, 49(1), 40-51.

5.Chen, D., & Anderson, C. (2024). Categorical data analysis. arXiv preprint arXiv:2409.02942.

6.Clark, J. S. (2020). Models for ecological data: an introduction. Princeton university press.

7.Cummings, P. (2019). Negative Binomial Regression. In Analysis of Incidence Rates (pp. 271-292). Chapman and Hall/CRC.

8.Ebrahimi Monfared, Z., Mirkarimi, S. H., Mohammadi Kangarani, H., & Soofi, M. (2025). Fear and belief predict perceived carnivore abundance in Golestan National Park, Iran. https://doi.org/10.5281/ZENODO.17469318

9.Elliot, N. B., & Gopalaswamy, A. M. (2017). Toward accurate and precise estimates of lion density. Conservation Biology, 31(4), 934-943.

10.Forsyth, D. M., Ramsey, D. S., & Woodford, L. P. (2019). Estimating abundances, densities, and interspecific associations in a carnivore community. The Journal of Wildlife Management, 83(5), 1090-1102.

11.Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ, 6, e4794.

12.Jiménez, J., Nuñez-Arjona, J. C., Rueda, C., González, L. M., García-Domínguez, F., Muñoz-Igualada, J., & López-Bao, J. V. (2017). Estimating carnivore community structures. Scientific Reports, 7(1), 41036.

13.Kruppa, J., & Hothorn, L. (2021). A comparison study on modeling of clustered and overdispersed count data for multiple comparisons. Journal of Applied Statistics, 48(16), 3220-3232.

14.Li, H., Luo, W., & Baek, E. (2024). Multilevel modeling in single-case studies with zero-inflated and overdispersed count data. Behavior Research Methods, 56(4), 2765-2781.

15.Matthiopoulos, J., Fieberg, J. R., & Aarts, G. (2023). Species-Habitat Associations: Spatial data, predictive models, and ecological insights.

16.Payne, E. H., Gebregziabher, M., Hardin, J. W., Ramakrishnan, V., & Egede, L. E. (2018). An empirical approach to determine a threshold for assessing overdispersion in Poisson and negative binomial models for count data. Communications in Statistics-Simulation and Computation, 47(6), 1722-1738.
17.Payne, E. H., Hardin, J. W., Egede, L. E., Ramakrishnan, V., Selassie, A., & Gebregziabher, M. (2017). Approaches for dealing with various sources of overdispersion in modeling count data: Scale adjustment versus modeling. Statistical methods in medical research, 26(4), 1802-1823.
18.Reangsephet, O., Lisawadi, S., & Ahmed, S. E. (2018, June). Improving estimation of regression parameters in negative binomial regression model. In International Conference on Management Science and Engineering Management (pp. 265-275). Cham: Springer International Publishing.
19.Saputro, D. R. S., Susanti, A., & Pratiwi, N. B. I. (2021, February). The handling of overdispersion on Poisson regression model with the generalized Poisson regression model. In AIP Conference Proceedings (Vol. 2326, No. 1, p. 020026). AIP Publishing LLC.
20.Sidumo, B., Sonono, E., & Takaidza, I. (2024). Count regression and machine learning techniques for zero-inflated overdispersed count data: application to ecological data. Annals of Data Science, 11(3), 803-817.
21.Tang, W., He, H., & Tu, X. M. (2023). Applied categorical and count data analysis. CRC Press.
22.Tinker, M. T., Yee, J. L., Laidre, K. L., Hatfield, B. B., Harris, M. D., Tomoleoni, J. A., ... & Miles, A. K. (2021). Habitat features predict carrying capacity of a recovering marine carnivore. The Journal of Wildlife Management, 85(2), 303-323.