A multifaceted analysis of factors influencing the number of days an animal spends in the Dallas shelter

Jiahuan Wang, Jiaxin Ye, Aulia Khoirunnia, Guanjie Wang, Dantong Wang

Introduction

- The Dallas animal shelter's dataset delves into multifaceted nature of shelter operations and animal characteristics, aiming to illuminate the factors that influence how long animals remain under care before reaching a definitive outcome.

- The primary objective of this study is to meticulously identify and analyze the determinants affecting animals' stay durations at the Dallas animal shelter, with the aim of leveraging these insights to advocate for and implement strategies that enhance the well-being of animals and optimize shelter operations.

Explanatory Data Analysis

◆Statistical Summaries

Table 1: The Description of the Dataset

Variable	Description
type	The type of animal admitted to the shelter
month	Month the animal was admitted
year	Year the animal was admitted
intake	Reason for the animal being admitted
outcome	Final outcome for the admitted animal
chip	Did the animal have a microchip with owner information
duration	Days spent at the shelter between being admitted and the final outcome

Table 1 categorical variables including animal type, intake, outcome reasons and chip status

Explanatory Data Analysis

◆Visualizations of Numerical Variables

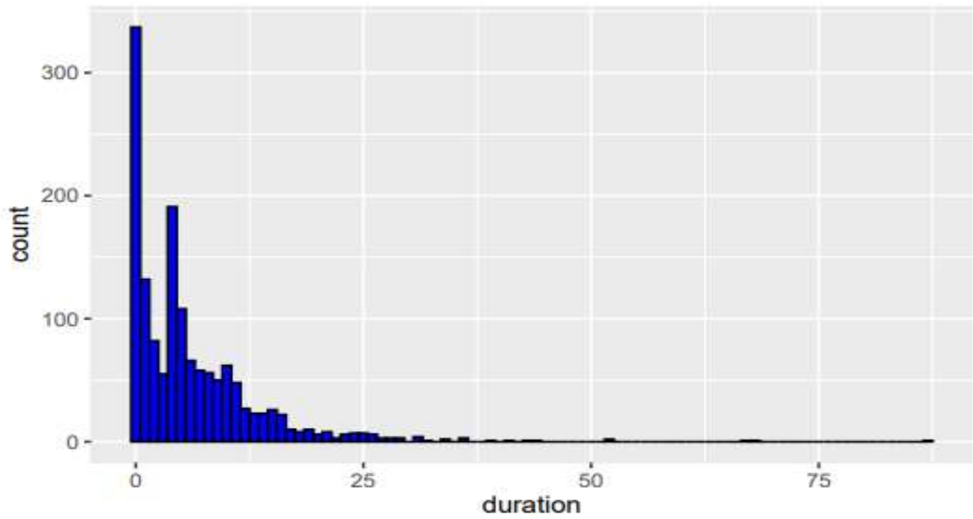

Figure 1: The Distribution of the Duration Variable

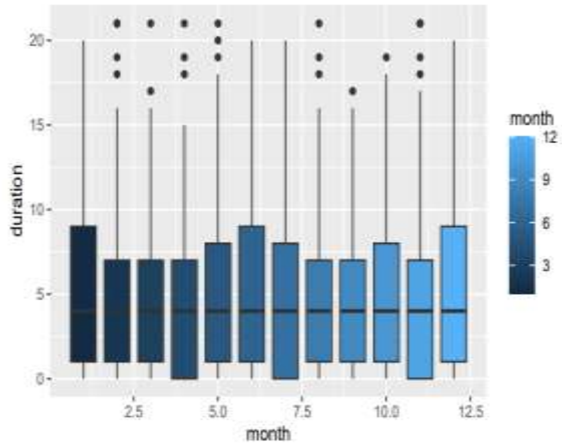

Figure 2: The Distribution of Month vs Duration Variable

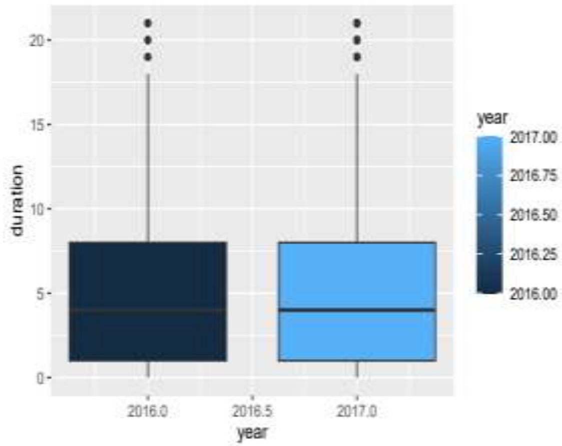

Figure 3: The Distribution of Year vs Duration Variable

Figure 1 demonstrates the distribution of how long animals stay in the shelter

Figure 2 and Figure 3 analyze the duration of shelter stays across different months and years, showing variability in the relation to these temporal factors

Explanatory Data Analysis

◆Visualizations of Categorical Variables

Table 2: The Distribution of Animal Types

Var1	Freq
BIRD	0.0020478
CAT	0.2075085
DOG	0.7822526
WILDLIFE	0.0081911

Table 2 categorizes shelter animals into types with their respective frequencies

Figure 4, Figure 5, and Figure 6 depict how intake reasons, animal outcomes, and microchip status influence shelter stay duration

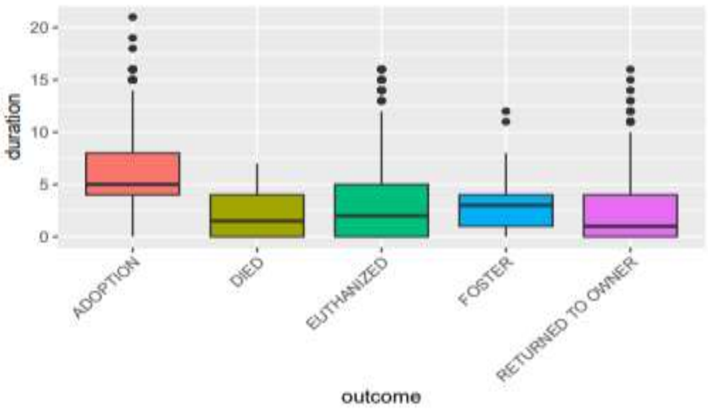

Figure 5: The Distribution of the Outcome Variable

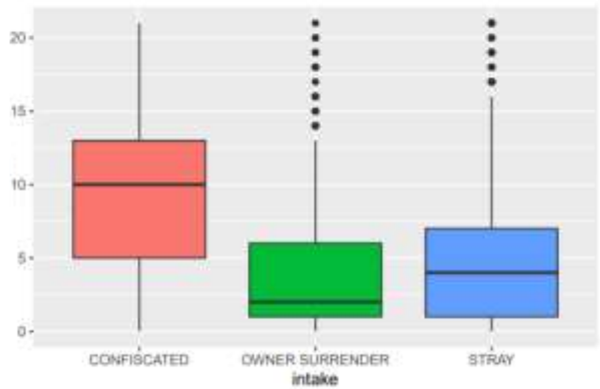

Figure 4: The Distribution of the Intake Variable

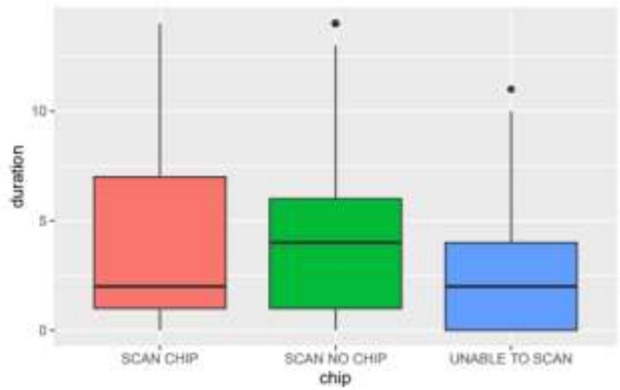

Figure 6: The Distribution of the Chip Variable

Explanatory Data Analysis

◆The Impact and Independence of Predictors on Shelter Stay Duration

Table 3: Chi-Square Test Results for Association with Duration of Shelter Stay

Variable	P_Value
type	4.4e-06
intake	0.0e+00
outcome	0.0e+00
chip	0.0e+00

Table 4: Multicollinearity Assessment of Predictor Variables

	GVIF	Df	GVIF^(1/(2*Df))
type	1.129852	1	1.062945
month	2.287867	1	1.512570
year	2.280442	1	1.510113
intake	1.211155	2	1.049059
outcome	1.357479	4	1.038943
chip	1.140194	2	1.033343

- Table 3 shows chi-square test results identifying a significant association between categorical variables and shelter stay durations.
- Table 4 assesses multicollinearity among predictors, which is crucial for evaluating their independence in influencing stay duration. GVIF values suggest excluding “month” and “year” could simplify predictive models without compromising accuracy, favoring model parsimony.

Formal Analysis

◆ Poisson Regression model

equation based on the initial analysis:

$$
\begin{aligned}
\log(\text{Expected Count of Time at Shelter}) = {} & 2.48068 \\
& + 0.18080 \times \text{TypeDog} \\
& - 1.11596 \times \text{IntakeOwnerSurrender} \\
& - 0.63153 \times \text{IntakeStray} \\
& - 0.99932 \times \text{OutcomeDied} \\
& - 0.70378 \times \text{OutcomeEuthanized} \\
& - 0.69878 \times \text{OutcomeFoster} \\
& - 1.31990 \times \text{OutcomeReturnedToOwner} \\
& - 0.14291 \times \text{ChipScanNoChip} \\
& - 0.38348 \times \text{ChipUnableToScan}
\end{aligned}
$$

Formal Analysis

Model Assumptions: mean rate of events (λ) is equal to the variance of the count of events.

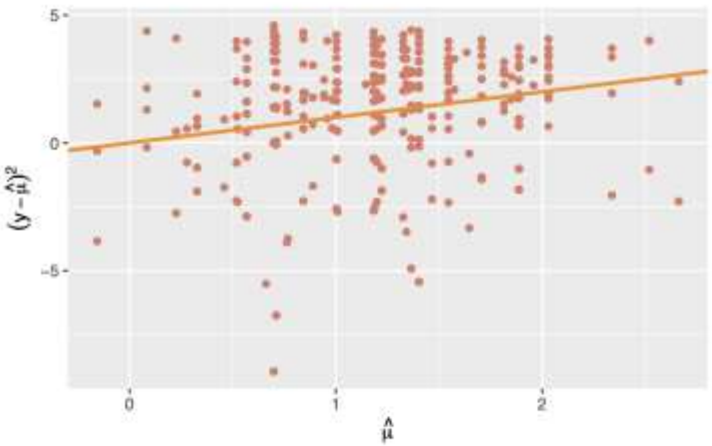

Figure 5: residual

- p-value (2.2e-16) is below the significance level, which indicates that the result is highly statistically significant

- dispersion (2.584389) is substantially greater than 1 and the variance of residuals increases with fitted values, which is a classic sign of over-dispersion.

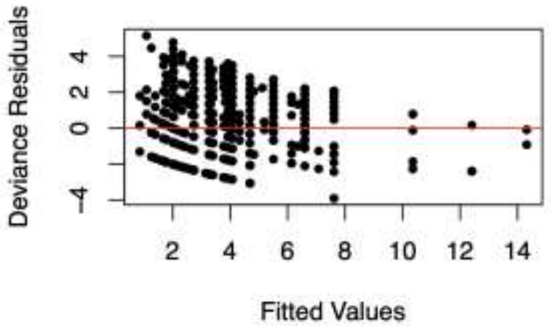

Figure 6: Residuals vs Fitted Values

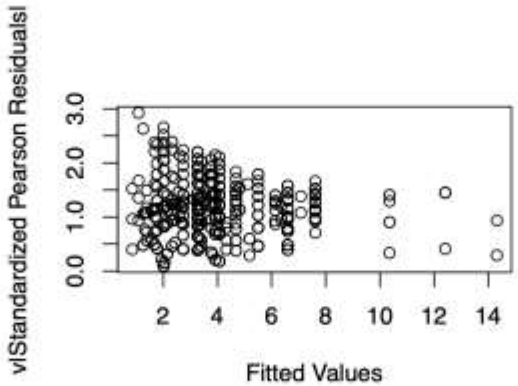

Figure 7: Scale-Location

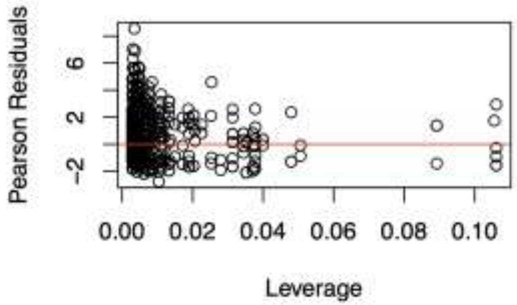

Figure 8: Residuals vs Leverage

Evidence of overdispersion suggests Poisson models is not suitable.

Formal Analysis

◆ Negative Binomial Model

equation based on the initial analysis:

$$
\begin{aligned}
\log(\text{Expected Count of Time at Shelter}) = {} & 2.58882 \\
& + 0.22863 \times \text{TypeDog} \\
& - 1.23810 \times \text{IntakeOwnerSurrender} \\
& - 0.75541 \times \text{IntakeStray} \\
& - 0.99159 \times \text{OutcomeDied} \\
& - 0.75098 \times \text{OutcomeEuthanized} \\
& - 0.67674 \times \text{OutcomeFoster} \\
& - 1.39817 \times \text{OutcomeReturnedToOwner} \\
& - 0.15081 \times \text{ChipScanNoChip} \\
& - 0.40916 \times \text{ChipUnableToScan}
\end{aligned}
$$

Formal Analysis

According to model summary:

- The parameter Theta: 2.189 and a relatively small standard error Std.Err: 0.16 indicate that the model has identified and is accounting for over-dispersion in the data.

- Residual Deviance: 1795.2 with 1440 degrees of freedom and Null Deviance: 2316.9 with 1449 degrees of freedom demonstrate that the negative binomial model is significantly improved in fit over the null model.

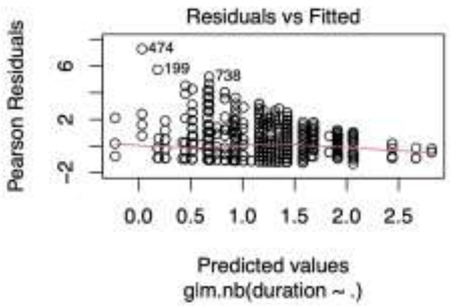

Figure 11: Residuals vs Fitted Values

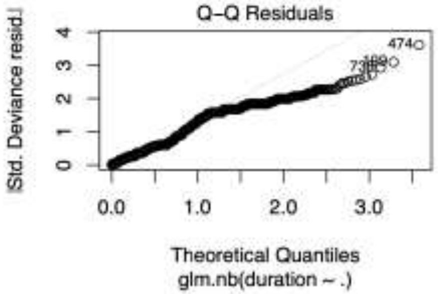

Figure 12: Normal Q-Q Plot

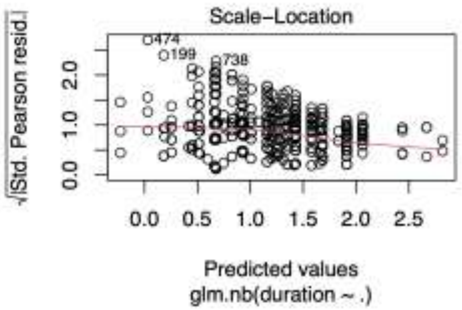

Figure 13: Scale-Location

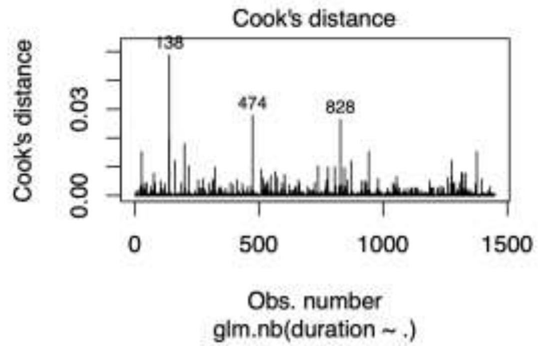

Figure 14: Residuals vs Leverage

Negative binomial model appears to perform well in fitting the central tendency and dispersion of the data.

Model Selection

The backward elimination process (guided by AIC) shows that models containing all variables performed better in fitting the data.

Negative Binimial Model:

```
Start:  AIC=6756.64
duration ~ type + intake + outcome + chip

          Df    AIC
<none>       6756.6
- chip     2 6763.7
- type     1 6769.3
- intake   2 6902.7
- outcome  4 7128.5
```

Poisson Model:

```
Start:  AIC=7440.01
duration ~ type + intake + outcome + chip

          Df Deviance    AIC
<none>         3759.4 7440.0
- chip     2   3785.3 7461.9
- type     1   3786.3 7464.8
- intake   2   4167.9 7844.5
- outcome  4   4944.4 8617.0
```

The Negative Binomial model's AIC of 6756.64 is much lower than the Poisson model's AIC of 7440.01, so the negative binomial model is a better choice for in-depth analysis.

Conclusion

Key Findings:

- The negative binomial model performed better than the Poisson regression model on handling over-dispersion in the data .
- Important predictors included animal type, reason for intake, animal outcome, and microchip status.

Implications for Practice:

- Conduct targeted outreach campaigns for specific animal types to increase animal adoption rates in shelters.
- Enhance microchip implantation practices to improve welfare outcomes.

Future Research:

- Further research is needed to enhance predictive capabilities and provide insight into the dynamics of animal shelter operations.