

Shelter Animal Outcomes in the United States: An Application of Negative Binomial Regression in R

by
Kristina Diaz-Kerkado
Advisor: Dr. Samantha Seals, PhD



An Undergraduate Proseminar
In Partial Fulfillment of the Degree of
Bachelor of Science in Mathematics
The University of West Florida
May 9, 2023

The Proseminar of Kristina Diaz-Kerkado is approved:

Dr. Samantha Seals, PhD, Proseminar Advisor

Date

Dr. Jossy Uvah, PhD, Committee Chair

Date

Accepted for the Department/Division:

Dr. Jia Liu, PhD, Chair

Date

Table of Contents

	Page
Acknowledgements	ii
Disclaimer	iii
Abstract	iv
1 Introduction	1
1.1 Statement of Problem	1
1.2 Relevance of Problem	1
1.3 Literature Review	1
2 Data and Methods	3
2.1 Data Description	3
2.2 Models and Assumptions	4
2.3 Analysis and Results	5
3 Conclusions	11
3.1 Summary of Key Findings	11
3.2 Suggestions for Further Study	11
Bibliography	12

Acknowledgements

This paper would not have been possible if it wasn't for the following people/organizations:

Dr. Seals, without her guidance and wisdom, this paper would have been tear-stained and incoherent.

Shelter animal count, without their data, this paper wouldn't exist.

The UWF math and statistics department, thank you, without them this paper would be nonsense.

My parents and family, thank you for taking care of me and my daughter, without them this paper wouldn't exist.

Disclaimer

The information contained herein was derived from data supplied by Shelter Animals Count. Shelter Animals Count specifically disclaims all responsibility for any analysis, interpretations, conclusions and opinions contained in or derived from the information presented.

Abstract

Objectives: Shelter Animals Count (SAC) is a 501(c)(3) organization funded through grants, sponsorships, and donations. SAC has established a national database for shelters to submit data to in order to build programs to protect animals. This project aims to model shelter and animal outcomes longitudinally in hopes of helping SAC and other organizations see what trends, if any, exist over time.

Methods: In this analysis, negative binomial regression was used to model the animal outcomes as a function of state characteristics. US Census and presidential election data was incorporated with SAC data at the state level. Wald's Z was constructed to determine statistical significance. Statistical significance was defined *a priori* as $p < 0.05$.

Results: When modeling for intake counts the only significant predictors were year, state population, and income. When modeling for happy (live) outcome counts, democratic affiliation point was the only predictor that was not significant. All the predictors were significant when modeling sad (other) outcome counts.

Conclusions: Sad outcomes is significantly dependent on income, population, and democratic affiliation. Happy outcomes is significantly dependent on population and income. Intakes is significantly dependent on population and income.

Chapter 1

Introduction

1.1 Statement of Problem

The purpose of this research is to assess what socioeconomic factors impact animal shelter counts. More specifically, we are checking if median income, state population, political party, are factors in adoption and euthanasia counts per state. If we show that as income goes down and euthanasia goes up, is it really appropriate to put these limitations on adoption. With analysis we can determine if time, income, population, and democratic percentage point have an effect on intake counts, adoption counts, and euthanasia counts.

1.2 Relevance of Problem

Whether it be a cat, dog, parrot, or turtle, it is known that owning pets can decrease stress and depression [1]. Due to our love for animals, humans have opened up shelters for cats and dogs since the 1800's [2]. Even though we have these shelters, due to shelter over population, euthanasia is still ongoing in the United States [3]. Knowing the animal intakes and outcomes counts per shelter, shelters can plan for any changes that may need to be made in order to avoid a high euthanasia count. As humans, we have an ethic to uphold humanity in our pets. If euthanasia is something that can be avoided, it should. Employees have been known to get burned out due to putting animals down [4]. With this study, maybe future research could find a way to end unnecessary euthanasia once and for all.

1.3 Literature Review

According to the American Veterinary Medical Association, the percentage of households owning dogs has increased by 1.9% while the percentage of households owning cats has decreased by 5% from 2012 to 2018 [5, 6]. Shelter Animal Count (SAC) has established a national database for shelters to contribute data in order to improve animals' lives [7]. Sean Andrews [8] from Case Western Reserve University researched socioeconomic factors affecting live release rates in pet population in shelters using mixed-methods. He concluded that poverty rate has a negative impact on live release rate, private shelter types have the largest positive effect on live release rates, and pets getting transferred in from other shelters has a positive effect on live release rates. Jeffrey Rodriguez, Jon Davis, Samantha Hill, Peter Wolf, Sloane Hawes, and Kevin Morris researched the impact of the first year of the COVID-19 pandemic on animal shelter intakes and outcomes[9]. Using linear regression analysis, they concluded that total intakes and euthanasia for pets

decreased over the five year pandemic, and that live release rates showed an increase over the five year span. R.J. Kilgour and D.T.T. Flockhart did a study on cat outcomes[10]. They wanted to test if cat outcomes depends on experiences prior to outcome, physical attributes and health status, human influences prior to intake, and the location and timing of the intake. Using structural equation modeling, they concluded that physical attributes, location of intake, date of intake, and length of stay prior to outcome all were significant predictor of type of outcome: adoption, euthanasia or death, or return to field. William Brown and Kelsey Morgan, using generalized linear mixed models, concluded that age, coat color and pattern, and breed designation all influenced the duration of stay for cats in a shelter in New York [11]. William Brown, Janelle Davidson and Marion Zuefle concluded that length of stay for dogs is influenced by age, size, and breed category in two no-kill shelters in New York [12]. Sloane Hawe, Josephine Kerrigan, Tess Hupe, and Kevin Morris, using descriptive statistics, found that cats were more likely to be returned after adoption for personal reasons while dogs were returned for behavioral issues [13].

Chapter 2

Data and Methods

2.1 Data Description

The original data set had 69 variables and over 400,000 observations. It listed all the organizations that contributed over 10 years of data for both cats and dogs. As illustrated in Table 2.1, the variables extracted that make up the happy outcomes, live outcomes, included: adoption count(age youth, adult, unknown), returned to field count(youth, adult, unknown), returned to owner count(youth, adult, unknown), other live outcome count(youth, adult, unknown), transferred out to agency in state(youth, adult, unknown), transferred out to agency out of state(youth, adult, unknown), and transferred out to agency internationally(youth, adult, unknown). The variables extracted to make up the sad outcomes, other than live outcomes, included: died in care count(youth, adult, unknown), lost in care count(youth, adult, unknown), and shelter euthanasia count(youth, adult, unknown). The variables that made up the intakes included: stray at large count(youth, adult, unknown), relinquished by owner count(youth, adult, unknown), impounded or seized count(youth, adult, unknown), other intakes count(youth, adult, unknown), transferred in from agency in state count(youth, adult, unknown), transferred in from agency out of state count(youth, adult, unknown), and transferred in from agency internationally count(youth, adult, unknown).

Intakes	Live Outcomes	Other Outcomes
Relinquished by owner	Returned to owner	Lost in care
Stray/at large	Adoption	Died in care
Impound/Seizure	Returned to field	Shelter Euthanasia
Transferred in(in state)	Transferred out(in state)	
Transferred in(out of state)	Transferred out(out of state)	
Transferred in(international)	Transferred out(international)	
Other intakes	Other live outcomes	

Table 2.1: Response Variables

Table 2.2 summarizes the predictor variable. In order to model the outcomes in terms of socioeconomic factors, we pulled data from the US Census: population in millions[14, 15] and income in \$5,000's[16–25]. For democratic percentage point, data was taken from the American Presidency Project[26–28].

Variable	Scale	Values
Year	Discrete	2012-2021
Population(Millions)	Continuous	0.58 - 39.51
Income(\$5,000's)	Continuous	7.44 - 18.39
Species	Categorical	Dogs, Cats
Democratic(% point)	Continuous	21.90 - 92.15

Table 2.2: Predictor Variables

2.2 Models and Assumptions

When the response variable is continuous, ordinary linear regression can be used.

Linear Regression Model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (2.1)$$

In R, this can be model like so:

$$lm(outcome \sim predictor1 + predictor2 + \dots, data = dataset) \quad (2.2)$$

Since the response variables are discrete and count data, we can use Poisson regression to model outcomes.

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k, \quad (2.3)$$

where

Y is the outcome dependent of the predictors,

β_0 is the intercept,

β_i are coefficients of the predictors, and

X_i are the predictor terms.

Poisson regression can be modeled in R like so:

$$glm(outcome \sim predictor1 + predictor2 + \dots, data = dataset, family = "poisson") \quad (2.4)$$

When interpreting Poisson models, we convert coefficients β_i into incident rate ratios (IRR).

$$IRR_i = \exp \{ \beta_i \} \quad (2.5)$$

If $IRR > 1$, then that indicates an increase in our expected count. If $IRR < 1$, then that indicates a decrease in our expected count. This is called the multiplicative effect.

When interpreting in terms of a continuous predictor we can say: For a 1 [unit of predictor] increase in [predictor name], the expected count of [outcome] are [increased or decreased] by $[100(e^{\beta_i} - 1)\% \text{ or } 100(1 - e^{\beta_i})\%]$. When interpreting in terms of a categorical predictor we can say: On average, [category predictor] expected [outcome] count is $[100(e^{\beta_i} - 1)\% \text{ or } 100(1 - e^{\beta_i})\%]$ [higher or lower] than [reference category] expected [outcome] count.

In order to test if our predictors have a significant affect on our outcomes, we can use Wald's Z test. The hypotheses are

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Our test statistic is as follows:

$$Z_0 = \frac{\hat{\beta}_i}{SE_{\hat{\beta}_i}} \quad (2.6)$$

We reject H_0 if $p < \alpha$, where $p = 2P[Z \geq |Z_0|]$

When modeling using the Poisson distribution we need to assess the assumption, that the means equals the variance. If this assumption is broken and $\text{mean} > \text{variance}$, the data is over dispersed, which will underestimate the standard errors, the test statistics will become larger, the p -values are then smaller than they should be, potentially causing a Type I error, a false rejection on the null hypothesis. To avoid this, we can use negative binomial regression instead:

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (2.7)$$

The negative binomial regression includes a parameter which accounts for an overdispersion of the data. The way to model it in R would be as follows:

$$glm.nb(outcome \sim predictor1 + predictor2 + \dots, data = dataset) \quad (2.8)$$

2.3 Analysis and Results

When taking a look at the data we determined that our response variables were skewed to the right.

As shown in Figures 2.1, 2.2, and 2.3, we see more number of observation in the lower count region of the histograms and fewer observations in higher count region of the histograms. This is because we have large

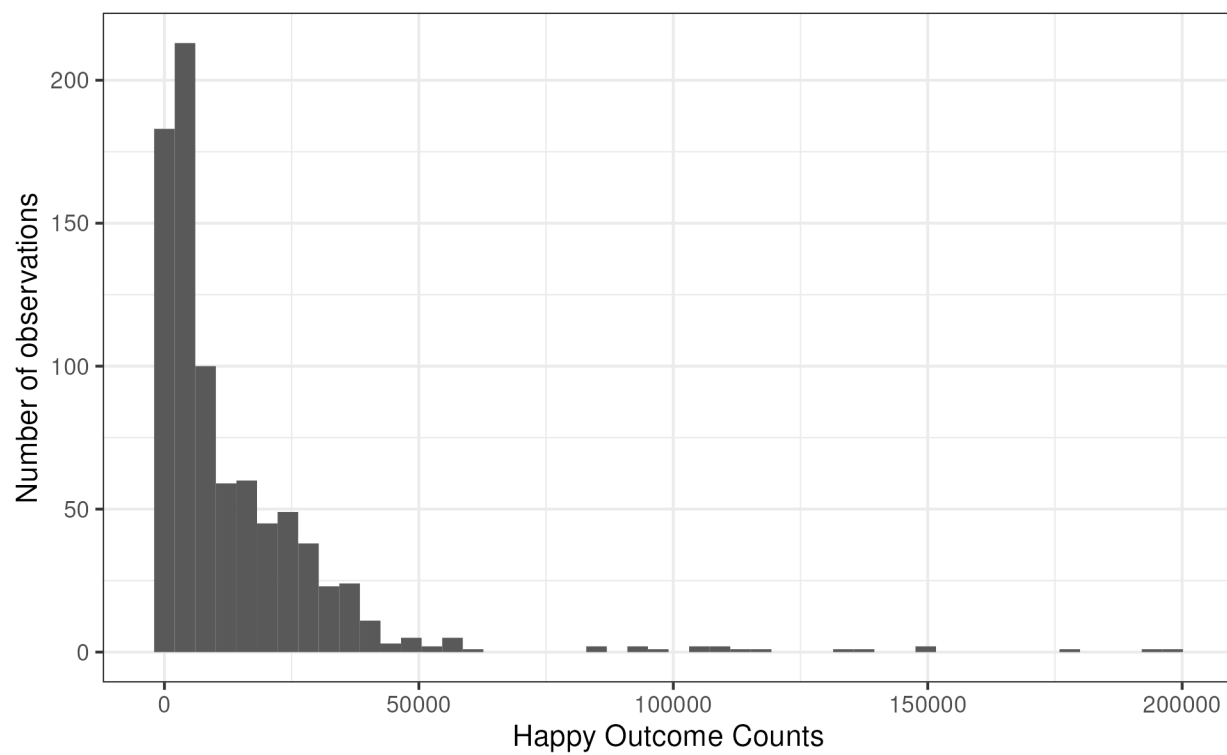


Figure 2.1: Intakes Histogram

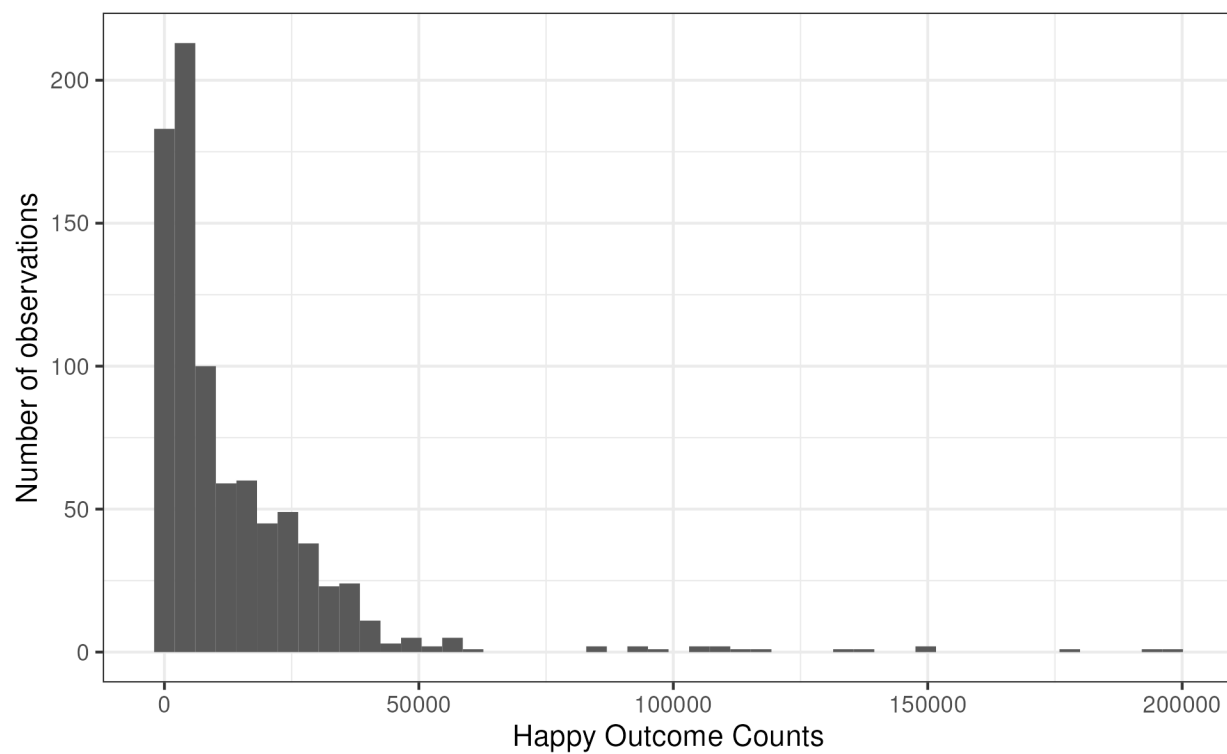


Figure 2.2: Happy Outcomes Histogram

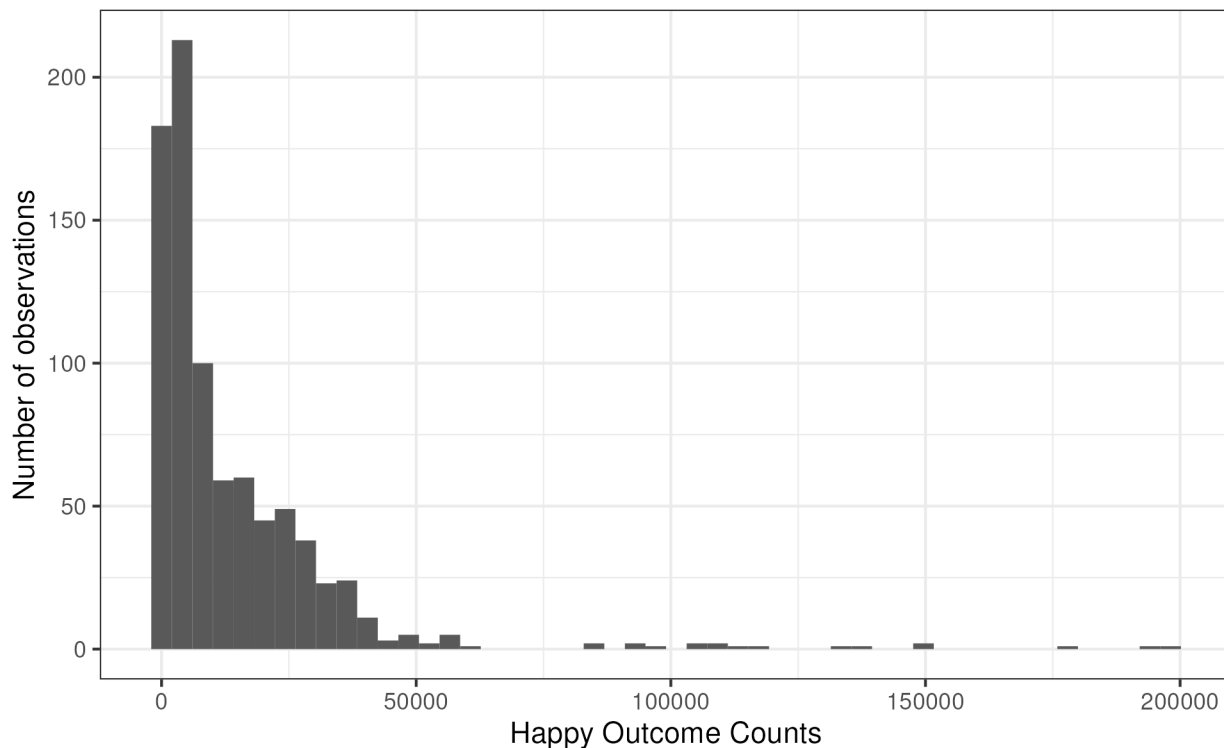


Figure 2.3: Sad Outcomes Histogram

states, like Texas, Florida, and California, with bigger populations than other states, causing their intakes and outcomes to be more than the rest. This is one indicator that our data is overdispersed. Mentioned earlier, we can also check to see if our means and variances are equal. Table 2.3 illustrates the overdispersion.

	Intakes	Live Outcomes	Other Outcomes
Mean	16412.49	13972.02	2838.94
Variance	676569396	437922843	25967030

Table 2.3: Overdispersion Test

The variances are much larger than the means, so our Poisson distribution assumption is broken, and we will use negative binomial regression to account for the overdispersion present in the data.

Table 2.4 shows the summary statistics for the data. Briefly we can see the population observations for states ranges from around 580,000 people to 39,510,000 people. With 3/4 of those observations being less than around 6,910,000. The median income observations for states ranges from around \$37,200 to around \$91,950. The democratic percentage point observations ranges from 21.9% to 92.15%. With 3/4 of the observations being less than 54.3%. The intakes count observations ranges from 27 to 266706, with 3/4 of the observations are less than 3525.5. The happy outcome counts ranges from 18 to 198149. The sad

Characteristic	Min	P25	Median	P75	Max
Population(Millions)	0.58	1.79	4.41	6.91	39.51
Income(5ks)	7.44	10.14	11.51	13.36	18.39
Democratic(% point)	21.90	38.10	46.80	54.30	92.15
Intakes	27	2539	8313	21859	266706
Happy Outcomes	18	2576	6702	19496	198149
Sad Outcomes	1	249	1228	3525.5	74599

Table 2.4: Descriptive Summary

outcome counts ranges from 1 to 74599.

The reason small counts are present is due to Alaska, where it is too cold for many pets to survive in the wild, coupled with the population is considerably much smaller than the rest of the United States. There was one year where the total happy outcomes count was 198,149 which is more than double another year where the sad outcomes count was 74,599. We can see that happy outcomes are more likely to happen than sad outcomes.

Table 2.5 summarizes the coefficients of all the models.

	Intakes	Happy Outcomes	Sad Outcomes
<i>Predictors</i>	<i>Log-Mean</i>	<i>Log-Mean</i>	<i>Log-Mean</i>
(Intercept)	-598.52	-659.83	-364.66
Year	0.30	0.33	0.19
Population(Millions)	0.10	0.10	0.10
Income(5k)	-0.11	-0.04	-0.20
Species[Dogs]	-0.07	0.15	-0.66
Democratic(% point)	-0.0005	-0.01	0.01

Table 2.5: Estimates

<i>Predictors</i>	<i>IRR</i>	<i>95% CI</i>	<i>p</i>
(Intercept)	0.00	0.00 - 0.00	< 0.001
Year	1.35	1.31 - 1.40	< 0.001
Population(Millions)	1.11	1.09 - 1.12	< 0.001
Income(5k)	0.90	0.86 - 0.94	< 0.001
Species[Dogs]	0.93	0.82 - 1.06	0.261
Democratic(% point)	1.00	0.99 - 1.01	0.885

Table 2.6: Intakes

$$\begin{aligned}
\ln(\text{Intakes}) = & -598.5239 + 0.3018(\text{Year}) + 0.1001(\text{Population}) - 0.1067(\text{Income}) \\
& - 0.0731(\text{Species}_{\text{Dogs}}) - 0.0005(\text{Democratic})
\end{aligned} \tag{2.9}$$

Table 2.6 illustrates the IRR, 95% confidence interval, and p -value when modeling the number of intakes.

Equation 2.09 is the model for sad outcomes using the values from Table 2.5.

- For a 1-year increase in time, expected intake count increases by 35%.
- For a 1-million increase in population, expected intake count increases by 11%.
- For a \$5,000 increase in income, expected intake count decreases by 10%.
- For a 1-percentage point increase in democratic affiliation, expected intake count decreases by 0.05%.
- On average, dogs expected intake count is 7% lower than cats expected intake counts.

<i>Predictors</i>	<i>IRR</i>	<i>95% CI</i>	<i>p</i>
(Intercept)	0.00	0.00 - 0.00	< 0.001
Year	1.39	1.35 - 1.44	< 0.001
Population(Millions)	1.11	1.09 - 1.12	< 0.001
Income(5k)	0.96	0.93 - 1.00	0.043
Species[Dogs]	1.16	1.03 - 1.30	0.015
Democratic(% point)	.99	0.99 - 1.00	0.058

Table 2.7: Happy Outcomes

$$\begin{aligned} \ln(\text{LiveOutcomes}) = & -659.8284 + 0.3318(\text{Year}) + 0.1028(\text{Population}) - 0.0381(\text{Income}) \\ & + 0.1468(\text{Species}_{\text{Dogs}}) - 0.0061(\text{Democratic}) \end{aligned} \quad (2.10)$$

Table 2.7 illustrates the IRR, 95% confidence interval, and p -value when modeling the number of happy outcomes. Equation 2.10 is the model for sad outcomes using the values from Table 2.5.

- For a 1-year increase in time, expected happy outcome count increases by 39%.
- For a 1-million increase in population, expected happy outcome count increases by 11%.
- For a \$5,000 increase in income, expected happy outcome count decreases by 4%.
- For a 1-percentage point increase in democratic affiliation, expected happy outcome count decreases by 1%.
- On average, dogs expected happy outcome count is 16% higher than cats expected happy outcome count.

<i>Predictors</i>	<i>IRR</i>	<i>95% CI</i>	<i>p</i>
(Intercept)	0.00	0.00 - 0.00	< 0.001
Year	1.20	1.15 - 1.26	< 0.001
Population(Millions)	1.10	1.08 - 1.12	< 0.001
Income(5k)	0.82	0.78 - 0.87	< 0.001
Species[Dogs]	0.51	0.44 - 0.61	< 0.001
Democratic(% point)	1.01	1.00 - 1.02	0.002

Table 2.8: Sad Outcomes

$$\begin{aligned} \ln(\text{OtherOutcomes}) = & -364.6637 + 0.1853(\text{Year}) + 0.0956(\text{Population}) - 0.1965(\text{Income}) \\ & - 0.6646(\text{Species}_{\text{Dogs}}) + 0.0139(\text{Democratic}) \end{aligned} \quad (2.11)$$

Table 2.8 illustrates the IRR, 95% confidence interval, and p -value When modeling the number of sad outcomes. Equation 2.11 is the model for sad outcomes using the values from Table 2.5.

- For a 1-year increase in time, expected sad outcome count increases by 20%.
- For a 1-million increase in population, expected sad outcome count increases by 10%.
- For a \$5,000 increase in income, expected sad outcome count decreases by 18%
- For a 1-percentage point increase in democratic affiliation, expected sad outcome count increases by 1%,
- On average, dogs expected sad outcome count is 49% lower than cats expected sad outcome count.

Checking our models for multicollinearity, Table 2.9 Shows our models' variance inflation factor.

<i>Predictors</i>	<i>Intakes</i>	<i>Happy</i>	<i>Sad</i>
Year	1.56	1.56	1.56
Population(Millions)	1.05	1.05	1.05
Income(5k)	2.08	2.08	2.08
Species[Dogs]	1.00	1.00	1.00
Democratic(% point)	1.62	1.62	1.62

Table 2.9: VIF: Multicollinearity Check

Since all VIF <10, multicollinearity is not present in any of the models.

Chapter 3

Conclusions

3.1 Summary of Key Findings

Table 3.1 compares the models side by side,

<i>Predictors</i>	Intakes		Happy		Sad	
	<i>IRR</i>	<i>p</i>	<i>IRR</i>	<i>p</i>	<i>IRR</i>	<i>p</i>
Year	1.35	< 0.001	1.39	< 0.001	1.20	< 0.001
Population(Millions)	1.11	< 0.001	1.11	< 0.001	1.10	< 0.001
Income(5k)	0.90	< 0.001	0.96	0.043	0.82	< 0.001
Species[Dogs]	0.93	0.261	1.16	0.015	0.51	< 0.001
Democratic(% point)	1.00	0.885	0.99	0.058	1.01	0.002

Table 3.1: All Models

Year, population, and income were significant predictors of all outcomes. Species did not significantly predict intakes, but did significantly predict both happy and sad outcomes. Political leanings of the state was only a significant predictor of sad outcomes.

3.2 Suggestions for Further Study

This analysis could be extended by using months instead of years as a time factor. We can also compare state with one another if density population was included as a predictor instead of raw population. Further, the data could be expanded to analyze at the zip code level, rather than the state level. We could use this to determine what regions need more shelters than others.

Bibliography

- [1] “The power of pets,” *News in Health*, February 2018.
- [2] K. Horecka and S. Neal, “Critical problems for research in animal sheltering, a conceptual analysis,” *Frontiers in Veterinary Science*.
- [3] M. Ramirez, “Nationwide crisis: Facing lower adoption rates, some shelters reconsidering euthanasia policies,” *USA Today*.
- [4] K. A. Anderson, J. C. Brandt, L. K. Lord, and E. A. Miles, “Euthanasia in animal shelters: Management’s perspective on staff reactions and support programs,” *Anthrozoös*, (2013).
- [5] “U.s.pet ownership statistics.” <https://www.avma.org/resources-tools/reports-statistics/us-pet-ownership-statistics>, 2018.
- [6] “Market research statistics- u.s. pet ownership 2012.” <https://www.avma.org/resources/reports-statistics/market-research-statistics-us-pet-ownership-2012>, 2012.
- [7] “Intake and outcome database (iod).” <https://www.shelteranimalscount.org/intake-and-outcome-database-iod/>.
- [8] S. P. Andrews, “Mixed methods study examining organizational and socioeconomic factors affecting management of pet populations in shelters,” 2021.
- [9] J. Rodriguez, J. Davis, S. Hill, P. Wolf, S. Hawes, and K. Morris, “Trends in intake and outcome data from u.s. animal shelters from 2016 to 2020.,” *Front. Vet. Sci.* 9:863990, 2022.
- [10] R. Kilgour and F. DTT.(2022), “Direct and indirect factors influencing cat outcomes at an animal shelter,” *Front. Vet. Sci.*
- [11] W. Brown and K. Morgan, “Age, breed designation, coat color, and coat pattern influenced the length of stay of cats at a no-kill shelter.,” *J Appl Anim Welf Sci*.
- [12] W. Brown, J. Davidson, and M. Zuefle, “Effects of phenotypic characteristics on the length of stay of dogs at two no kill animal shelters.,” *J Appl Anim Welf Sci*.
- [13] S. Hawes, J. Kerrigan, T. Hupe, and K. Morris, “Factors informing the return of adopted dogs and cats to an animal shelter.,” *Animals (Basel)*.
- [14] “Table 1. annual estimates of the resident population for the united states, regions, states, and puerto rico: April 1, 2010 to july 1, 2019,” *U.S. Census Bureau*.

- [15] “State population totals and components of change: 2020-2022,” *U.S. Census Bureau*.
- [16] “2012 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [17] “2013 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [18] “2014 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [19] “2015 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [20] “2016 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [21] “2017 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [22] “2018 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [23] “2019 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [24] “2020 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [25] “2021 poverty and median household income estimates - states and national,” *U.S. Census Bureau*.
- [26] “The american presidency project: Election statistics 2012,” *The American Presidency Project*.
- [27] “The american presidency project: Election statistics 2016,” *The American Presidency Project*.
- [28] “The american presidency project: Election statistics 2020,” *The American Presidency Project*.