Final Report

Isaac Slagel and Jack Welsh

1. Introduction A few paragraphs that contain background information, motivation for your research, and a statement of your research goals. Be sure to incorporate your supporting references into the text. The purpose of the background is to place your work in the greater context of the literature in the area you are investigating. Then you should explicitly identify a hypothesis that you will investigate with your analysis.

include some more general stuff about the effect fo

Our project looks into animal shelters in the Dallas area using a dataset provided by the ciry of Dallas through dallasopendata.com. This data includes information about every animal that has been brought into an animal shelter in the Dallas area. There are 61634 individual animals in this dataset with with information about when and where the animal was found, the health of the animal on arrival, and whether the animal was adopted, euthinized or transfered. We will be investigating the factors that influence the death of an animal and the ammount of time the animal will spend in the shelter before dying, being adopted, or being transfered. We hyppothesize that pitbulls will be adopted at lower rates and euthinized at higher rates. In addition, we hypothesize that dogs without chips are less likely to survive their time at the animal shelter than dogs with chips. We believe that this work is important to help better understand what factors influence the death of animals in animal shelters and improve the outcomes of for animals in animal shelters.

Talk about the literature and recent movements to get rid of pitbulls.

In recent years there has been a movement to regulate pit bulls, many people believe that pitbulls are inherintly not safe and a movement to make them illegal. # add some citations and stuff.

Also find some papers about chip status.

2. Materials and Methods Three to five paragraphs (or fewer) that...

- Briefly describe your data, where it came from (source), definitions of important variables, and how it was collected

We accessed our data from the Dallas Open Data website. This website contains many public datasets, including animal shelter records. Anyone can request an access key to this dataset and then download it. We downloaded the data, selected the variables of interest to us, and created a .csv file in the CreateDataset.Rmd file. Our dataset is saved as adoptions.csv. The initial variables from this dataset are described below in Table 1.

- Indicate any modifications made to the data, recoding, or decisions about missing data

We used the imported data described in table 1 to create a fleet of indicator variables: `adopted`, `chip_status`, `summer`, `treatable_intake`, `adopted`, `dead`. These indicator variables were created to streamline our modleing process. They allow us to increase the interpretability of variables with multiple categories (like outcome_type).

In order to fit a binomial regression, we had to summarise our dataset with respect to certain variables we controlled for in our model. One interesting issue that we encountered was the significance of the dataset we fit on SE estimates. If we summarized our data for each specific model, parameter estimates had increased SEs and more insignificant p-values. However, if we fit all of our models on the same, more extensive, summary table, SEs were small and parameters were more likely to be significant. We decided to proceed with the latter option, as it allows us to carry out nested F tests and would be more like a modeling process for true grouped data. This could mean that our SEs are artifically low.

- Briefly but thoroughly describe the statistical inference methods used to quantify the association between your outcome and predictor variables (at least one method must have been introduced in Stat 316). What summary statistics were calculated? What statistical tests were performed?

Table 1: Description of Variables

Variable Name	Variable Role	Variable Type	Range of Values	Units
animal_breed	explanatory	categorical	296 unique breeds	NA
animal_origin	explanatory	categorical	4 sources of shelter animals	NA
animal_type	explanatory	categorical	5 species of animal	NA
chip_status	explanatory	bianary	(0,1)	NA
intake_type	potential confounder	categorial	how animal came to be at the shelter	NA
outcome_type	reponse	categorial	how animals was removed from shelter	NA
intake_condition	potential confounder	categorial	keyword description of animal status at intake	NA
outcome_condition	potential confounder	categorial	keyword description of animal status at outcome	NA
intake_date	response	date	(2017-10-01, 2019-04-03)	y-m-d
outcome_date	response	date	(2017-10-01, 2019-04-03)	y-m-d

We used binomial regression to compare how the outcomes of pitbulls differ from other breeds of dog. In fitting this model we had to control for a plethora of cofounding variables including season of outcome, chip status, and intake condition of the animal.

We used a Cox Proportional hazard test to compare how the outcomes of . . .

- Specify strategies employed when building your models

We used findings from our exploratory data analysis to guide decisions on variable selection. Additionally we refered to models fit in the literature to guide decisions about model inclusion.

- Do not report results in the Materials and Methods section!

Note: If you are using a method not covered in Stat 316, you may choose to expand Materials and Methods a bit to describe your statistical method.

3. Results The meat of your report, which should include. . .

- A general description of your data (completed via your exploratory data analysis)

EDA Results: Pitbull

To look into how pit bulls are handled within animal shelters, we decided to explore how the outcomes of pit bulls may differ from non-pitbulls. Table 1 describes the differences in outcome rates for non-pitbulls compared to pitbulls.

Table 2: Pitbull Outcomes

Outcome	Pitbull (%)	Non-Pitbull (%)
Adoption	32.03	35.73
Euthanized	28.42	10.25
Returned to owner	22.25	30.96
Transfer	10.92	17.56
Foster	2.25	1.64
Other	1.88	1.59
Dead on arrival	0.93	0.82
Treatment	0.77	1.01
Died	0.51	0.39
Missing	0.04	0.06

We see that pit bulls, while only adopted at a slightly lower rate (~5%), are euthanized at well over double the rate of other dogs. Further, we see that other dogs have a much higher chance of being transferred to another facility or returned to their owner. To get a better understanding of this relationship, but we likely need to control for things like chip status, intake condition, and season of outcome in a binomial regression.

- A description of the results from your analyses, including interpretations of parameter estimates, tests, and confidence intervals in context.

Table 3: Modeling Dog Outcomes in Dallas Animal Shelters

	Dependent variable:		
	Proportion of dogs who died		
	(1)	(2)	(3)
Intercept	0.116***	0.111***	0.552***
	(0.069,0.186)	(0.069,0.170)	(0.454,0.669)
Pitbull	3.440***	3.424***	3.489***
	(1.795,6.557)	(1.905,6.130)	(3.022,4.027)
Scannable Chip	0.789	0.799	0.781***
	(0.377,1.566)	(0.412,1.483)	(0.667,0.911)
Summer Outcome	1.461	1.447	1.478***
	(0.725,2.852)	(0.771,2.649)	(1.271,1.718)
Contagious		7.286**	3.975***
		(1.324,44.137)	(2.568,6.168)
Treatable At Intake			0.161***
			(0.133,0.196)
Overdisperson Parameter	139.72	111.46	6.27
Nested F Test		F : 5.1142*	F : 313.62***
Note:			*p<0.1; **p<0.05; ***p<0.01

- Tables that summarize results and figures that illustrate results. These tables and figures should be well-labeled, numbered (e.g. Figure 1), and have a good, descriptive caption. Each report should have a

minimum of two plots; rarely are residual plots part of the main body of the report unless they are an integral part of the story.

- While you should interpret tests, confidence intervals, and coefficients in this section, you should not editorialize here! Save that for the Discussion.

4. Discussion A few paragraphs that:

- Begin with an accurate summary statement; describe how the results help answer your research questions and what was most interesting from your analysis. In fact, the first paragraph of the Discussion is very important – in professional journals, it is often the first and sometimes the only paragraph that is read in a paper. After the first sentence highlights primary results, the remainder of the first paragraph might compare your results to others in the literature or include interesting secondary results.
- Discuss possible implications of the results in the context of the research question.
- Make a statement regarding potential confounding variables in your study
- Make a statement about the generalizability of your results. Don't give generic statements of possible causation and generalizability, but thoughtfully discuss relevant issues – confounding variables, representativeness of the sample, etc.
- Identify any limitations of your study. Discuss the potential impact of such limitations on the conclusions.
- Identify strengths and weaknesses of your analysis.
- Make suggestions for future research. Identify important next steps that a researcher could take to build on your work.
- Do not include test statistics or p-values in this section.

5. Annotated Appendix

- Tables and figures that are informative but were not referenced specifically in the main report. Include a short annotation – one or two sentences on what they show.
- R scripts and output (annotated) so that I can trace how you constructed your final data set, what models you ran to produce the results quoted in your report, and what intermediate models you also considered.
- Description of statistical modeling steps that were not included in the main body of your report. Possible entries here include: How you handled missing data. Evaluation of assumptions. Outlier analysis and how you decided to deal with any outliers along with rationale for your decision. Describe hypotheses testing you performed during model building and how you decided on the explanatory variables you ultimately included in your final model. Assessment of the final model.
- How you went from the model output in R to interpretations in your report (e.g. exponentiate coefficients, then take inverse)
- Anticipate questions someone might have after reading your report, and make sure those questions can be answered with information in the appendix.
- A citation for each reference article (in APA format or something similar) you included in your proposal. Also include a link, if appropriate. Remember that you must have the entire paper and not just an abstract, and at least two must be from peer-reviewed journals.