Analysis of relation between dog food and diarrhea

Karsten Binz
Matrikelnummer 4148379
karsten.binz@student.uni-tuebingen.de

Michelle Schlicher
Matrikelnummer 5889471
michelle.schlicher@student.uni-tuebingen.de

Abstract

In the context of this project, we collected data about the nutrition and excrements of our dog. Since he suffers from random diarrhea, we wanted to analyze if there is a relation between his different kind of meals and whether he gets diarrhea. We analyzed the consistency and color of the excrements related to the different meals visually. Using the Fisher's Exact Test, we did not find a significant relation between diarrhea and the individual meals. Nevertheless, grouping the meals into two, based on ingredients that were present in the three meals with the highest diarrhea rate, we discovered a significant positive relation between meals with beef lung, heart and larynx and diarrhea within the next day.

1 Motivation

Our dog Buddy, a four year old Collie, suffers from recurring diarrhea. After examination by a veterinarian, a severe disease is unlikely to be the cause. Rather, an intolerance of some ingredient in his meals is suspected. In this project we wanted to analyze if the meals Buddy gets, have a relation to whether he gets diarrhea or not.

2 Data Collection

Four weeks prior to data collection the dogs nutrition plan was changed according to the recommendation of the veterinarian. It meant removing dry-food from the diet. Therefore, his diet consisted out of 900 gram of wet food and 100 gram of puffed rice per day. All food comes from Petfit. The same kind of wet food was given to the dog for two days in a row in most of the cases.
We collected data about the nutrition and the dogs excrements from the 24th of October, 2022 until the 26th of January, 2023. At the beginning of the data collection phase, Buddy was diagnosed with a parasite that causes diarrhea, therefore we needed to exclude the whole data during his medication. After the medication phase he was examined and declared as healthy again. Therefore, we only included data from the 10th of November, 2022. Because of severe sickness of us, Buddy was in external care and we could not record any data from the 22th of December, 2022 until the 5th of January. We decided to discard the data we recorded in January, because of the break in between and because we changed his diet to not having the same kind of food for two days in a row anymore. This only left us with data from 42 days.
Regarding the excrements, we recorded the number, consistency and color. Additionally to the kind of wet food, we recorded snacks, the duration of the walks, stressors and further relevant observations of his well-being. For every day these values have been recorded mornings, midday and evenings.

The possible values of the collected factors are listed in Table 1. The data can be found online in the following Google Sheet.

Table 1: Captured variables and possible values

Variables	Values
Daytime	Morning, Midday, Evening
Number of Excrements	*Integer*
Number of Diarrhea	*Integer*
Walk Duration	*Integer*
Meal	Chicken-Beef, Lamm, Duck, Deer, Rumen, Winter, Beef, Special, Power, Vital
Excrement Color	Dark-brown, Light-Brown, Yellow, Green, Red
Excrement Consistency	Good, Too soft, Fluid
Snacks	Beef Strips, Beef Lung, Potato-Deer, Beef Rumen, Veal Bone, Wild Rabbit, Dried food
Stressors	Car ride, Bathing, Other Dogs
Other	Throw-up, Flatulence

3 Data Processing

After removing the excluded time frames from the data set, we corrected for consistent capitalization and factor naming, as well as casted the variables to the correct data types. We manually checked for further inconsistencies. Invalid or missing entries were replaced with an empty string or a zero, depending on the data type. We decided to not exclude days with missing data points, since the data was sparse already.
After consulting a domain expert on how long the digestion of a dog takes, we assumed that a meal can have an effect on the excrements from 24 to 36 hours afterwards. Since Buddy gets half of his daily portion mornings and the other half evenings, there is no food information for midday. Therefore, it would not make sense to directly map the effect 24 hours into the future, for example midday meal affects the next midday excrements. Accordingly, we decided to considered whole days as variables, aggregated over them and pushed nutrition-related variables (meal and snacks) one day into the future. We discarded the remaining non-nutrition related variables from the first data point, leaving us with 41 days. This way every data point is associated with one meal and a binary variable stating whether there was diarrhea the next day, meaning the meal is associated with diarrhea.
We derived the binary diarrhea variable from the column containing the number of diarrhea per day. We also generated variables stating how often the different colors and the different consistencies occurred per day.
Moreover, it did not seem reasonable to summarize the meals into groups from the beginning on, since they all mainly represented one kind of animal meat (e.g. beef or chicken). Also, the differentiation of for example beef and beef rumen was relevant to consider, since a specific intolerance (to rumen) is also possible. Nevertheless, the meals do not only contain one kind of animal meat, but a mixture of different animals and animal parts, which also leads to an overlap in the ingredients in the different meals.
Lastly, it is important to mention that not every kind of wet food was fed the same number of times. The ten different meals were fed from one to nine times. Therefore, the data set is unbalanced. The meal *special* was only fed once, since it was a leftover can. The quantification can be seen in plot 1c.

4 Analysis

4.1 Descriptive Statistics

Because of the low number of data points, we decided to not further analyze the factors snacks, walk duration, stressors and other saliencies. The 41 data points contained 110 excrements, from which thirteen were diarrhea. On average we had 2,68 excrements per day. Plotting the diarrhea over time,

did not show any accumulations. Therefore, making it unlikely, but not impossible, that a temporal sickness lead to the diarrhea. We continued assuming that some nutrition leads to the diarrhea.

4.2 Visual Analysis

First, we visualized the consistency of all excrements for the different meals. This can be seen in Plot 1a. The colors illustrate whether the consistency was good, too soft or diarrhea. The numbers behind the meals in the x-axis show how often they were fed. Also, the y-axis in subplot c and d, show the corresponding quantification. *Beef*, *chicken-beef* and *deer* showed to be the three meals with the highest diarrhea rate. Except for *special*, which only had good, all meals were evenly associated with good and soft excrements. Thus, allowing no further insight on which meals might be especially well digestible. The visualization, also, did not clearly show meals to be strongly related to diarrhea.
To further investigate whether the meals influence the digestion, we plotted the color of all excrements per meal. In a healthy case, excrements should be dark brown. Light brown is not optimal, but must not necessarily be a bad sign, since the color depends on the received nutrition. Red, green and yellow can be a sign of a sickness, but can also mean the consumption of something unusual with high pigmentation, for example grass. In plot 1b, it can be seen, that there is no clear relation between the meals and the color. The colors illustrate the true exctrement colors. Dark and light brown seem evenly distributed. Since there are so few data points for the colors red, green and yellow, no sufficient statement can be made about these colors. Noticeable, rumen has more other colors than just dark and light brown. This could be, because rumen are known to be harder to digest for dogs. Overall, the visual analysis did not give further insight on whether the different meals lead to diarrhea.

4.3 Statistical Analysis

For the statistical analysis we decided to not consider all individual excrements, but only whether there was at least one case of diarrhea on the next day. Considering the individuals could strongly bias the result. If for example a meal leads to a higher number of excrements, it would be over represented.
To check if there is a statistical significant association between the two variables meal and diarrhea on the next day, we used the Fisher's Exact Test. It is an alternative to the Chi-Squared-Test that takes two categorical variables, while allowing the confusion matrix to have cells with values lower than five. Running the test on the binary variable diarrhea (true or false) and the variable meal, with all ten possible values, we did not find a significant association ($p = .2334$).
Out of intuition, we decided to see whether dividing the meals into the two groups containing and not containing chicken, would lead to a significant association, but none could be found ($p = 1$).
Following the idea of only analyzing the association of two binary variables, we wanted to group the meals into containing suspicious ingredients and not containing them. Therefore, we checked the three meals with the highest diarrhea rate (deer, beef, chicken-beef) for same ingredients and found beef lung, heart and larynx. Afterwards, we identified that *power*, *special* and *winter* also contained these three ingredients and thus added them to the suspicious ingredients group. Running the Fisher's Exact Test indeed lead to a significant association ($p = .0285$). The counts of diarrhea for the two groups can be seen in Table 2 on the left hand-side. It can also be seen that the two groups are unbalanced (25 and 16 meals).
Since it is generally very hard to say how long the effect of a meal reaches into the future, we decided to also consider an effect of one days meal on the excrements of the next two and not only one following day. Meaning, in this case the binary variable diarrhea indicated whether there was diarrhea within the next two days. Running a Fisher's Exact Test on this diarrhea variable and the two suspicious ingredients group lead to an even lower p-value of .0093. The increase in associated numbers of diarrhea can be seen when comparing plot 1c, which illustrates the 24 hour relation, and plot 1d, which considers 48 hours. The summarized diarrhea counts for the two groups can be seen in Table 2, comparing the one day and the two day relation.

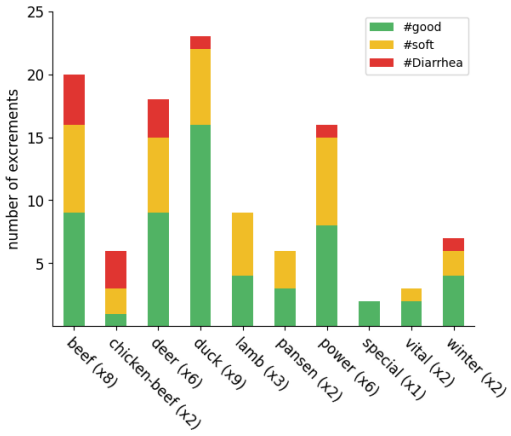
5 Discussion and Outlook

After analyzing the association of the different wet meals with the occurrence of diarrhea over the next 24 hours, we did not find the meals themselves, but the ingredients beef lung, heart and larynx

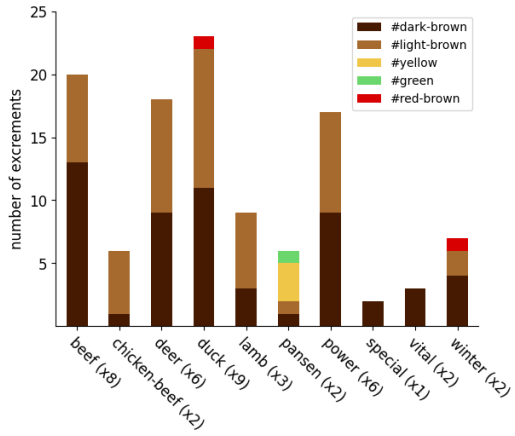
Table 2: Confusion matrix of diarrhea under meals with and without suspected beef parts

	1 day relation		2 day relation	
	sus. beef	¬ sus. beef	sus. beef	¬ sus. beef
Number of diarrhea	10	1	16	3
Number of no diarrhea	15	15	9	13

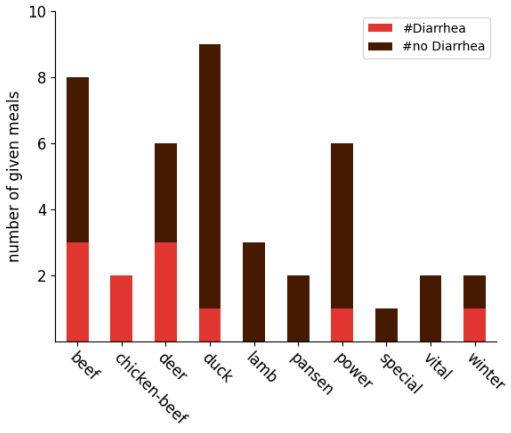
to be positively associated with diarrhea. The association was even stronger when considering an influence over two days.
These findings did not match our intuition that has built up over the last years. Also, the data set was quite small and unbalanced. During data acquisition, it was sometimes hard to classify the consistency of the excrements. Therefore, and because it is very hard to tell how long a dogs digestion really takes, we recommend taking the results with caution.
In order to confirm the findings, we are planning on adding the collected data from January (which was excluded because of a change in the diet) and repeat our analysis. If the data allows it, we would use the Chi-Square-Test rather than the Fischer's Exact Test for comparison. If we still find a high positive relation, we are planning on reducing the meals with the suspected ingredients. As long as we did not consult a veterinarian, we will not fully exclude them, to keep a divers diet. We will continue tracking whether the described change in diet, leads to less diarrhea.



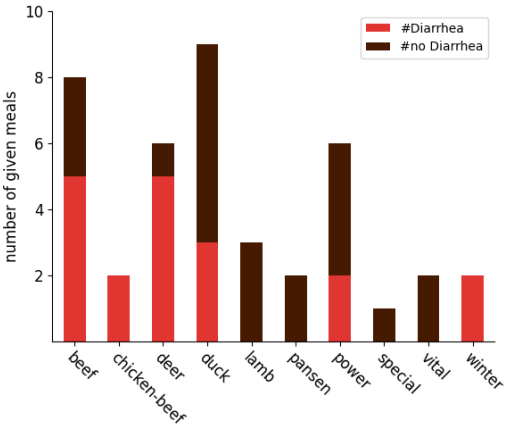
(a) Consistency proportion



(b) Color proportion



(c) 24h relation



(d) 48h relation